

CRITERIA FOR THE USE OF SIMULATION MODELS

by

JOHN S. HOFFMAN

B.A., University of Pennsylvania  
(1970)

SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF  
MASTER OF CITY PLANNING

at the

MASSACHUSETTS INSTITUTE OF

TECHNOLOGY

June, 1972

Signature of Author.....  
Department of Urban Studies, April 12, 1972

Certified by.....  
Thesis Supervisor

Accepted by.....  
Chairman, Departmental Committee on Graduate Students

Rotch



## CRITERIA FOR THE USE OF SIMULATION MODELS

by

JOHN S. HOFFMAN

Submitted to the Department of Urban Studies on April 12, 1972, in partial fulfillment of the requirements for the Degree of Master of City Planning.

### ABSTRACT

Deciding when computer simulation models provide useful information for decisionmakers requires some idea of how reliable the information the simulations provide will be. Such a decision about usefulness will depend on the type of problem, along with alternate models that are available for providing information to decisionmakers. Within this framework of applicability the validation process has meaning.

In simulations where system structure is not well or completely understood, such as large policy simulations, the researcher is presented with a likelihood that policies and experiments on his model are likely to influence the structure of his model. In order to escape the errors this could cause the researcher must develop a set of 'validation' and 'invalidation' tools which can be applied to every policy test, and which will allow the researcher to realize when the model must be changed.

Some possibilities for solving this 'validation' problem exist in the more careful delineation of boundary conditions, and the assumptions upon which the model is built. Urban Dynamics, a model by Jay Forrester, is examined in light of these 'validation' techniques, and found to fail to guard against changes in model structure which are likely as a consequence of policies suggested by Forrester.

Thesis Supervisor: John R. Harris

Title: Associate Professor of Urban Studies

### ACKNOWLEDGEMENTS

I would like to thank the following people for their assistance: Malcolm Jones, Leon White, Ross Cope, and especially my advisor, John Harris, for his helpful counsel and tolerance.

I would also like to thank my typist, Yvonne Wong, for her typing under ridiculous deadlines. Needless to say, all faults in this thesis are my own.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

ABSTRACT

	page
PART I - "The Main Idea".....	1
PART II - "Urban Dynamics".....	93
PART III - Summary.....	123

## CHAPTER 1

Computer simulation is a form of mathematical modeling. Models are artificial representations of the world as they abstract some part of reality but do not recreate reality in all aspects. As a mathematical model simulations allow one to make a set of assumptions about reality and to test the consequences of those assumptions on a computer. As such computer simulations are dynamic and numeric.<sup>1</sup>

Another form of mathematical models are analytic varieties which normally do not rely on computers for their solution but on analytic techniques of problem solving. Generally analytic models are highly abstract and often are in a generalized form which can be applied to many situations although it is possible to build analytic models for some specific situations. The difficulties encountered in solving such models often mitigate against this method, however. Simulations on the other hand are problem dependent and since they simply extend assumptions through brute force can be used in more complex situations.<sup>2</sup> Analytic models are generally used when one believes that the world can be simplified enough to meet the rigid constraints of the model. Examples are linear programming models, queuing systems and many of the operations research types of models. Simulations being generally imitative usually are used in situations where the simplifications necessary for the use of analytic models cannot be made and where analytic models will not

allow the rich policy alternatives that exist to be tested. As an experimental technique simulation allows a great deal of flexibility including more exploration of the dynamic paths to various states, although as an experimental technique finding an optimal solution is more difficult if not impossible in certain circumstances. The experimental feature of simulations allows for more realism in policy testing along with a greater range of possibilities.

The differences between analytic and simulation models has practical implications for using and validating models. For analytic models the most important component of validating use is to make sure that the world fits the model; for example in linear programming checking to see if the process has a linear function. The assumptions of analytic models are usually clear and exact. For simulations the problem is more complex. Since the simulation model has been constructed to look like the world and probably to explain historical data it is more difficult to know when the simulation is valid. The assumptions on which the computer simulation model are built are not as clear as those on which the analytic model are constructed. As a consequence it is more difficult to tell when the simulation holds and does not hold. This leads to some important differences which will be explored in greater detail later in this thesis.

An important difference between computer simulations and analytic models is that simulation is an experimental technique. Gordon

describes simulation as being "essentially an experimental problem solving method".<sup>3</sup> Before discussing the advantages and disadvantages of the experimental approach it would be worthwhile to examine why computer experiments may in some situations be advantageous to more traditional experiments.

The most important reason that computer experiments are preferable to real ones is cost. For instance, suppose one desired to test the effects of changing ordering policies of inventory for some firm which kept large warehouses. Computer simulation might give one some feeling for policies which meet one's objectives without incurring any cost except the building and running of the computer model. In the actual situation changing these policies might be extremely expensive if not disastrous. This expensiveness is likely to be particularly extreme if the policies to be tested are not intuitively clear or incremental.

Most situations in which simulations would be useful would have the same characteristics of cost for real experimentation. This is particularly true in large systems such as the national economic system. Here a second important factor becomes important, time.

Experiments on the national economy would take a great deal of time besides the unreasonable costs involved. In a problem where the number of experiments to be done is large, and the duration of the

experiments are long relative to the frame of reference for decision-making, computer simulation offers a method of doing experiments rapidly. Even in cases where a real experiment would take a short time if there are many to be performed computer simulation might offer a time saving mechanism.

Doing experiments in a real environment may also be methodologically difficult.<sup>4</sup> There exist numerous control, implementation, and design problems, along with problems of generalizing results. Of course, these problems also exist in experiments on computers; the difference lies in the solution of these problems. Suppose the experimental process will change the future response of the group being experimented on. Doing the experiment would lead to a dead end. Doing a simulation, however, might allow the impact of the experiment to be investigated including the impact of sensitivity analysis and of making structural changes in the model. In this case the use of the model might clarify some difficult methodological problems of the experimental process.

Finally, when a system does not exist computer simulations may be the only type of experiment available.<sup>5</sup> The Apollo space program comes to mind immediately, although more mundane problems such as designing computer systems have also yielded to such experiments.<sup>6</sup> Experimentation on systems before they are designed can influence the process of design and theory.

Simulation as an experimental technique serves two purposes. It allows different policies to be tested. Here a greater richness is possible than in analytic models. Secondly, it can allow, under certain circumstances, optimization to take place. This richness achievable under simulation allows policies of tremendous complexity to be tested, which means that they can also be tested under changing conditions. Because simulations do not rely on analytic techniques it is possible to make structural changes almost at will, which allows more realistic conditions to be examined, as well as broadening the types of sensitivity analysis done.

Furthermore, there exists a visibility to laymen of simulation models that does not exist for analytic varieties.<sup>7</sup> For clients the imitative aspects of simulation allow them to see themselves whereas with analytic models this is generally not so. This lack of visibility can often stand in the way of adequate use of analytic models.

Related to the ability to test policies in a more realistic fashion and to 'capture' more reality than analytic models is the capability of computer simulation for directing research.<sup>8</sup> By the use of sensitivity analysis, which is easily performed on computer simulations, the more important areas of study can be located. As a consequence research can be redirected as the needs of the computer simulation dictate. For example, in a model of the urban area many behaviors might be examined in detail. A simulation by revealing what areas

are sensitive to the model could direct limited research capability into those crucial areas, so that theory and the simulation can be improved. Use of simulations for such heuristic explorations is generally easier than for analytic models.

The disadvantages of simulations will be discussed in the rest of this thesis. Generally they surround the problems of dealing and validating such models. The gains have accompanying costs. In summary the advantages of computer simulations over analytic models though are many. Raser nicely summarizes them as:

- (1) dealing with greater complexity
- (2) having economy
- (3) having clarity to users
- (4) safety
- (5) being able to explore unknown regions

To understand the disadvantages it will be necessary to discuss the purposes and uses of simulation.

To discuss criteria for the use of simulation models it will be essential to delineate the various purposes, different users and creators of simulation models may have, since different uses may have different criteria. If one were concerned strictly with validation this point might be more difficult to accept, but since in this thesis validation will be discussed only within the framework of practical use the more complex epistemological issues surrounding that type of validity will be skirted. Using Bishop Occums razor the thesis will do away with unneeded assumptions.

One of the most common uses of simulations are for the purposes of training and education.<sup>9</sup> Examples of such simulations abound ranging from such Urban games as APEX and CLUG to the military games used to train future and present war commanders. Characteristically such games attempt to provide an environment in which the actors can take on roles of different people in a system, and attempt to make the decisions which these 'roles' would have made. For instance, in Clug each player is a developer who every turn (time period) can make a variety of decisions about how to develop a city.<sup>10</sup> The players can cooperate, ignore each other or make cartels. After each turn, a computer processes all the decisions and the consequences are accounted for in the next time period. The process repeats itself until the city is developed or the game ceases to interest the players.

In other games the players may take on different roles. In a game the author participated in given by David McClelland of the Harvard Social Relations department each person was assigned to a group which represented a community interest. Essentially, the simulation was a negotiation between militants, government officials, the businessmen and labor or the working class neighborhood. Concerned with the building of a highway the simulation showed when negotiations break down, violence can erupt, as it did in this simulation when the Mayor's daughter was kidnapped and then brutally killed by the radicals seeking to foster their ideas on the city by force. Unlike the CLUG game the rules of

action were less well developed, the accounting system was by hand.

Thus, these simulations are characterized by role playing, an accounting system, and some sets of rules as to what actions are tolerable. Depending on the design there may be an arbitrator who plays nature or the players may be left to settle disputes themselves.

Another type of training or educational simulation is slightly different. Such simulations depend on creating an environment in which a person can interact and feel that he is in a real system, at least in some respects.<sup>11</sup> Examples include flight trainers, controllers, and moon missions. In these simulations the purpose is to expose the actors to events that will replicate reality.

Even the purposes of this type of educational simulations can vary. In the case of the flight trainers the purpose is to produce as narrow a response in the actors as possible: the proper response to the proper stimuli. In some of the psychological simulations the purpose may be to make a person less sure of what response is correct. In a business game the purpose might be to give the businessman a particular empathy with some person or group with which he presently has none. The common characteristic is learning by the actors. It should come then as no surprise that at least one of the measures of validity suggested for educational simulations is the ability to learn something realistic.<sup>12</sup> Such a simple analysis clearly would not be enough to tell when learning was occurring.

It would be necessary to define what learning is desired. What is the response the designers aimed at with the simulation? How well did the game accomplish it? Was enough learned and was it accurate enough? How much 'bad learning' took place? If responses which weren't intended were learned then the simulation might not have been very useful but rather harmful.

In examining these criteria one would like to make them as explicit and objective as possible before testing the simulation. Such testing is difficult needless to say. Thus, in many cases the validity of games and simulations is a face validity - does it appear to teach what it claims to teach? The danger in this type of validation of criteria is that often the dysfunctional behavior learned in a situation will be ignored unless specific tests to examine what learning took place are set up.

Another set of simulations which exist although they are not usually classified this way are simulations designed to improve existing policies. The simulations are usually micro level simulations where all testable policies are clearly understood not to effect the simulation except through the given structure. Forrester has termed such simulations open loop although such an appellation may not totally be deserved.<sup>13</sup>

The purposes of such simulations are to improve policies regarding a highly structured operation. Examples of such simulations

include job shop simulations, packaging of various products, most queuing problems, and many of the industrial applications of simulation.<sup>14</sup>

In some sense the purpose of such simulations can be seen as improving the estimations of consequence of incremental policy changes to greater accuracy. Of course, sometimes the changes can be radical but the nature of the goal is usually similar. The value of such simulations will depend on their improving the decisionmaker's accuracy in making some type of adjustment to this process.

The nature of such simulations can vary although often they are open looped, which means that they have no feedbacks between policy choices and model structure, and that within their own structure feedbacks play a small role. As a consequence one does not have to worry about the impact of policy testing on model structure (most of simulation literature argues that this is always true in simulations; it is the main thesis of this paper that this is not true and represents a major problem to policy oriented simulators).

The criteria then to evaluate the usefulness of this class of simulations will be how accurately the simulation reproduces the real world as it exists now and in the past and will exist under policy changes in the future. A variety of tests exist which will be described later in this thesis which could be useful in this task.

It is fairly easy to see why this criteria dominates in the type of problem discussed in this area. For instance, an assembly line may be speeded up. The decisionmaker needs to know exactly how much additional strain will be put on the other parts of the line. Accuracy is extremely important because small inaccuracies can be disastrous. The building of a queue even at a slow rate could mean the collapse of the line unless one is aware that this will happen. Emphasizing accuracy of estimation then can be seen as the prime purpose of such simulations.

The justification of this as a category may seem odd, if not perverse to the reader. Because this thesis is examining the applicability of simulations to various situations by their usefulness it becomes clear that a functional category such has been defined in this case is of great use. The main remarks of this thesis will be aimed at simulations outside the realm of factory operation or where extremely accurate predictions are necessary or possible. Furthermore, the types of problems encountered in doing and validating those types of simulations are vastly different from those used for other simulations including those used for training and those used for large policymaking purposes.

The final category of models to be discussed are those used for major policy decisions. Generally, these models are characterized by large size, undefined assumptions about the environment, and at least to some degree shifting structure. Usually such models contain many behavioral assumptions, and a large number of relationships which make

them difficult to 'validate'.<sup>15</sup> Such models seek to improve decision-makers' ability to make decisions in a world of uncertainty. Using Don Schon's terminology they seek to change uncertainty to calculatable chance.<sup>16</sup>

Policy testing models seek to increase the awareness of the decisionmaker. Since a decision must be made the question is whether the decisionmaker will use the simulation model or his mental model.<sup>17</sup> He does not have the choice not to make use of some model because a decision about what to do is necessarily forthcoming even if that decision is to do nothing. Necessity forces choice and choice forces implicitly the use of some model even if that model is completely based on intuition. The question then becomes defined as whether the simulation model improves the ability of the decisionmaker to make decisions.<sup>18</sup>

Policy models deal in areas where the necessity for detailed knowledge is not usually great. More often the question is what is the general direction of change expected and in what general magnitude. For instance the Urban Dynamics model seeks to examine what the consequences of slum demolition is on in-migration of the poor. In examining whether that model would be useful one would really be most concerned with just knowing the general impact of the change; the exact number of people deterred from coming might be nice information to know,

but is not necessary except to tune the policy later. Where this accuracy is not possible the general solution is imposed and incremental changes are made to get an optimal policy.<sup>19</sup>

Lowrey's model is another example of an attempt to make a policy model. In it the concern is to explain the growth of the city and while accuracy is more important in it than in Urban Dynamics, the same types of characteristics about accuracy hold true. One does not expect exactness, but rather one seeks to know the impact of changes that will take place under various policies.

The criteria of usefulness when applied to these policy making simulations means that they will be of use to the decisionmaker in making decisions which will meet his objectives. A consequence of this is that ordinarily it is necessary for the model to include all the endogenously generated behavior relating to the issue being studied.<sup>20</sup> Furthermore, the model must be capable of testing a variety of potential policies, many of which differ significantly in the way they would effect things, the way they would be effectuated, and the way in which the implementation will affect the world.

From this point of view policy simulations must attain flexibility rather than accuracy, although of course gross errors cannot be tolerated (grossness would be defined by the cost of the error function). What the policy maker seeks is a reduction of the things that he doesn't understand. Ordinarily, he will not be committed to a rigid structure

of policy because the best method for accomplishing his goals cannot be encompassed by any rigid structures. He simply does not understand the world well enough to reduce it to a factory problem.

For example, the World Dynamics model is concerned with the ecological future of earth.<sup>21</sup> It is not concerned with the prediction of the date of the crisis, or the ultimate impact in lives lost, but only in the formation of such a crisis out of present policies. Not that the model is necessarily correct, but if it is the policymaker is more concerned with the direction of the changes the model says will occur than their exact specification.

There would even be situations where a model could not tell a decisionmaker the exact state of the world in the future but would be of use. An example might be a model to predict the influence of an electioneering ploy; to be of use it only need predict the influence a decision would have on shifts of voters. Even if the prediction of the way people will vote is incorrect or unavailable from the model the useful fact is predicting what type of voter shift would take place.<sup>22</sup>

In conclusion it is hoped that the reader is convinced that the usefulness of a computer simulation is problem dependent. Depending on the type of simulation, the purpose the simulation is being done, and the costs of errors, it would be fair to argue that computer simulation applicability is based on the particular problem at hand. All the uses

of computer simulations certainly have not been listed; the reader could think of many more such as theory building, learning, having fun and so on. The same types of arguments could be extended to those simulations in evaluating their usefulness. All that has been attempted so far is to show that different criteria can exist.

FOOTNOTES

1. Gordon, Geoffrey, System Simulation, page 7.
2. Emshoff, James, The Design and Use of Computer Simulation Models, page 53.
3. Gordon, op. cit., page 18.
4. Campbell, Experimental and Qusai Experimental Design for Research.
5. Van Horn, "Validation in Design of Computer Experiments," page 234.
6. Emshoff, op. cit., page 72.
7. Raser, Simulation and Society.
8. Gordon, op.cit., page 17.  
Orcutt, Microanalysis of Society.
9. Armstrong, The Use of Gaming/Simulation in Decisionmaking.
10. Cornell Land Use Game Manual.
11. 1971 Winter Computer Simulation Conference:  
"Applications of Computer Drawn Motion Pictures to Visualization."  
1971 Summer Computer Simulation Conference:  
"Real Time Simulation of Apollo Manned Mission for NASA Flight Controller Training."
12. Taylor, op.cit.
13. Forrester, Principles of Systems.
14. 1971 Summer Computer Simulation Conference:  
"Simulation of the Basic Oxygen Steelmaking Process,"  
Middleton.  
"Halpegewachs "Optimal Simulation of Nose Cones from Billets of Material."

15. Schechter, Mordechasi, Simulation and Games, Vol. I, No. 1.
16. Schon, Technology and Social Change.
17. Forrester, Industrial Dynamics, page 17.
18. Raser, op. cit., page 140.
19. Lindblom, The Policy Making Sciences, pages 12-30.
20. Forrester, Principles of Systems, pages 1-5.
21. Forrester, World Dynamics.
22. Poole, Candidates, Issues and Strategies, page 43.

## CHAPTER 2

Until now the reader has had to accept the definitions used in speaking of the "applicability" of models. In this chapter a clear definition of what applicability, usefulness, validation, truth, and policy oriented models will be attempted. Part of the problem in reaching definitions of these terms is the general confusion which exists in the literature and the widely varying definitions. Often the reason the same words are used differently are that the authors make different assumptions about what they are addressing. Some authors may be thinking of small models when they speak of validation; others may be thinking of large-scale policy models.<sup>1</sup> [See SCSC book for various definitions.] This section comes after a few assumptions have been made in the first chapter; it is hoped the reader will now have some idea of the assumptions of the author.

Usually in simulation literature before using models it is argued that the models must be validated.<sup>2</sup> The problem with such arguments is that they obscure the purposes of validation by not defining the contexts in which validation is being used. Thus one author can argue that it is necessary to "validate specific set of insights not necessarily the mechanism that generated the insights"<sup>3</sup>....while another author argues that not only must the model yield outcomes similar to the real world, but must use the same mechanisms to reach the outcomes.<sup>4</sup>

It is impossible to know whether the authors are really in disagreement over what they mean by validation or whether they are talking about different problems because they have not defined what they are referring to.

Generally, although not in all cases, validation is taken to mean that a model is internally consistent, and that it represents the actual world.<sup>5</sup> Validation of a model means arriving at agreement between behavior of the simulation tests and the behavior of the real system.<sup>6</sup> However, this type of definition can lead to ambiguity and confusion. What does looking like the world mean? When does something look sufficiently like the world to be valid?

Essentially the ambiguity comes from two different interpretations of what valid can mean - truth or usefulness. If one accepts the 'truth' criteria one may seek to define a set of objective criteria. The usefulness criteria would lead to many different sets of validation criteria. Later when this paper's definition of validation is put forth these ambiguities will be resolved by changing the meaning of the words.

However, presently it is not intuitively obvious to the most casual of observers which of the two possibilities should be adopted. The truth criteria has some inherently appealing characteristics. For one it would allow a standard procedure to be used in examining simulation models. It would, therefore, remove the murkier aspects of validation and the need for different validation strategies.

Furthermore, its acceptance would force common discussion of the issues of validation. Finally there exists a psychological appeal; 'is a model true or not' is the question most laymen would ask. Therefore before this criteria is done away with it will be necessary to present some cogent reasons.

The problem though is that truth is nowhere adequately defined. The whole philosophical battle between logical positivism and more traditional philosophies has shown the difficulties of defining truth.<sup>7</sup> When one says something is true what does one mean? If a model were true would it reproduce all the behavior of the system? And if it would not then when would it be true, when it was useful? Clearly if the latter is the case the truth criteria has degenerated to the other side, and if the former is true no one should bother to model since no model will be true in all respects. The problem is that in arguing between the various parties in the literature each one adopts his own definition of truth, such that various degrees of the concept of usefulness are admitted, but only implicitly, thus creating argument and confusion.

In order to erase the confusion created in the literature and which the reader is probably beginning to feel in this paper a new definition of validation will be offered. Validation will be defined as the process of executing tests based on the criteria the simulator has defined as useful. Validation will simply be determining whether the

criteria the simulator has decided to use are being met, not something separate from the simulation itself. If this seems to beg the question it does so very purposely. The important issue in the use of models is whether a model is applicable to the situation which a decisionmaker intends to use the model. This applicability will vary on the purpose of the model, and the alternative mechanisms the decisionmaker has for helping him to make decisions. Validation without reference to a particular purpose is really quite meaningless. Rather this thesis argues that for any model it will be possible to define a purpose and from that purpose to determine tests of whether the model can help meet that purpose. With a quick cut of Occum's razor the concept of truth will be eliminated; it is not necessary for the use of models.

This may seem to leave an enormous amount of ambiguity to the concept of validation since none of the concepts by which it is defined here are well understood or easy to pin down. This is probably true; however, the definition of validation of usefulness will serve to take discussion away from the realm of philosophy into the more mundane area of whether the models are meeting the goals of the whole decisionmaking process. To test if a model is applicable, one will define its purpose, tests to tell if that purpose is being achieved; from that point useful arguments can grow.

Some, although sparse, support exists in the literature for such a definition. Brewer states that "utility is not determined by truth criteria, rather it is measured with respect to a given set of well-defined purposes."<sup>8</sup> Raser argues that isomorphism (truth) is only one test of usefulness.<sup>9</sup> Van Horn argues that "There is no such thing as the appropriate validation procedure. Validation is problem dependent."<sup>10</sup>

There are problems in using usefulness as a criteria. It makes discussion of 'validation' difficult, since even in the same class of models such as policy-oriented models, different procedures and tests might be applicable. Different policies might imply different tests. For example, in using the same model of urban renewal it might be important to develop closely fitted predictions and statistical tests to improve relocation policymaking, although for community stability in the same model the possible consequences of policy need only crudely be predicted and the tests to insure this might emphasize a close fit to structure.

The construction of tests to check usefulness (to validate usefulness to use the new definition) is difficult. The tests used throughout simulation, particularly in micro simulations are intended to improve accuracy and are not of great use in policy-oriented models. Reproduction of time series (and this will be discussed more elaborately later) is not sufficient to check policy models.<sup>11</sup> The definition of

such tests and their application is one of the purposes of this thesis.

Other dangers exist in the usefulness definition. Churchman argues that "the trouble with this pragmatic philosophy is that it is correct, absolutely correct - mainly because it does not say anything at all. What it says is do what is feasible."<sup>12</sup> This is true only insofar as no attempt is made to define useful any further. Otherwise one can readily realize the wisdom of doing what is feasible; feasibility clearly is not intuitively obvious since many people devote their lives to the infeasible and much debate exists about whether that concept is correct. (The author is reminded of a defense project which is underway where workers on it have privately pointed out to him that for the system to function one would have to have movement three times the speed of light. Apparently the feasibility principle is not as readily accepted as Churchman feels it is, at least to the defence department.)

Another danger in the concept of usefulness can be more insidious and dangerous to the practitioner of simulation. Sisson and Emshoff argue that "If decisionmakers believe the model is useful and use it, the analyst has done his job."<sup>13</sup> Such an attitude is extremely dangerous, not that this is what Emshoff and Sisson intended; it would be unfair to hold them to something they might qualify if they had the chance. However, such an attitude does exist in the more charlatan elements of the profession. Clearly it is important to make sure the model does aid

the decisionmaker in making decisions that help him reach his goal. It may be easier to convince someone that you are helping them than to actually do so. Unless the purpose of the simulation is to be a placebo or laxative to decisionmakers it is necessary that the simulation meet not only the criteria of making the decisionmaker use it but helping him to his goal.

This discussion leads naturally into the place of simulation as a technique of systems analysis. Systems analysis is defined in terms of purpose.<sup>14</sup> Generally it is characterized as a goal-seeking method of problem solving or design. The conceptual approach used in this thesis is part of the systems analysis philosophy.

In defining usefulness one has problems distinguishing between folk knowledge (which may be extremely useful) and logically implied arguments (which often are close to worthless). Clearly the definition of usefulness with respect to policy models, which are the only type the thesis will deal with unless explicitly stated differently, is their ability to produce better policies. Such a statement is necessarily true, but not extremely helpful. More useful but not holding in every case are the various features of models that do yield better decisions. The characteristics of such information can be very illuminating. It has been suggested that some criteria are that

- (1) all the data used in the simulation are actually available
- (2) actual decisions will yield the same general consequences as simulated model
- (3) simulator decision process accurately reflects the proposed strategy and does not evoke knowledge unavailable at the time<sup>15</sup>

Other pieces of advice include decreasing size, or holding down complication can yield utility.<sup>16</sup> Many modelers argue that the model should generally resemble the processes which are being modeled in their entailment.<sup>17</sup> This often does add to ease of policy analysis along with utility.

However, further discussion of when models are useful will be left to Chapter 4.

FOOTNOTES

1. 1971 Summer Computer Simulation Conference.
2. Naylor, Computer Simulation Techniques.
3. Van Horn, "Validation," in Naylor, Design of Computer Simulation Techniques, page 234.
4. Crecine, Governmental Problem Solving, page 62.
5. Meier, Simulation in Business and Economics.
6. Fishman and Kiviat, Digital Computer Simulation, Statistical Considerations.
7. Russell, History of Western Philosophy.
8. Brewer, Evaluation and Innovation in Urban Research.
9. Raser, Simulation and Society, page 140.
10. Van Horn, op. cit., page 234.
11. de Neufville and Stafford, Engineering Systems Analysis, Volume III.
12. Brewer, op. cit., page 10.
13. Emshoff, The Design and Use of Computer Simulation Models.
14. de Nuefville, op. cit. pages 1-20.
15. 1971 Winter Simulation Conference, Besenfeller, "The Capital Asset Pricing Simulator."

16. Brewer, op. cit., page 3.
17. Forrester, Industrial Dynamics, pages 115-125.

### CHAPTER 3

In the interest of clarity this section will attempt to demonstrate where all this is taking the reader. In the introduction and the first two chapters the reader learned that the criteria for using simulation models are different than those of other models, that usefulness would define validation for simulation models, and that to accomplish 'validation' it would be necessary to construct tests.

The main exploratory point of the thesis is that such tests must be executed for every policy being tested on a model. Generally, in the simulation literature validation is a separate step from experimentation (policy testing) in the simulation process. The argument advanced in this thesis is that this is not true. Model structure is not independent of policy testing, experimentation, or even implementation. The same policy, say tearing down 5% of the slum housing as Forrester suggests in Urban Dynamics, can have vastly different consequences depending on the method of effectuation (the reader is warned here that this means method; everyone would agree that incompetence can change outcomes - in this case it means simply the technique used. An assumption of competence is made). In order to be sure of model usefulness different tests of validity may be necessary or more commonly different policies will require different model structures in some conditions. (That is the

tests of validity will show that a different policy will require some change in the model's structure). This is not an argument that this will occur in every situation, but rather that tests of validity (usefulness) must include mechanisms for knowing when this is necessary since it will occur under some conditions and the model will have no way of correcting itself, as a human's mental model can learn. Achieving this goal will be extremely difficult.

At this point the paper will return to its normal development and develop this thesis more later.

CHAPTER 4

For policy-oriented models the criterion of usefulness is the ability to improve the decisionmaking process. While academics may be interested in simulations of this type for other purposes, the policymaker feels that this is what the models must do.<sup>1</sup> Ordinarily policymakers desire models which allow them to do more than just predict future trends. Brewer argues that this is a crucial difference between academics and policymakers - the former are satisfied to reproduce a historical time series which the latter does not feel is of sufficient explanatory strength.<sup>2</sup> The reasons behind this difference is that the design of new policies requires at least some structural knowledge.<sup>3</sup>

For the decisionmaker to feel that a model is of use it is necessary to be able to consider all the existing alternatives within the structure of the model or possible alterations of the model's structure that available policies would necessitate. Ordinarily the simulation must have sufficient richness to also explore new alternatives which the policymaker may think of. Without the ability to explore alternatives the simulation simply would become a tautological justification of a given end. In a sense this experimental feature is demanded by the nature of the simulation process.<sup>4</sup>

In order to choose policies the decisionmaker must have greater belief that a given policy will meet his objectives than another policy. Ordinarily the decisionmaker will not really know the direction of all the policies he has at his command, particularly in policy areas in which simulations are built. Since in models of policy not very much accuracy is obtainable the purpose of many simulations in this category is to synthesize the complex forces which the policymaker is unable to handle.<sup>5</sup> There are, of course, notable exceptions to this rule such as the econometric models. Even there, however, real policy changes may not be able to be predicted; one need only look at the recent record of the models to realize that they too can break down.<sup>6</sup> An example of this type of policy problem could be taken from the area of prison reform. One might wish to know what the impact of such reform would be on crime committed. Would the deterrent that prison represents be lost? Would the number of prisoners released returning for new crimes committed go down? Without a simulation these questions could not be integrated to formulate policy. By extending these questions and even making only intelligent guesses as to the answer the decisionmaker may have the information upon which he makes his decision improved. Of course other factors than the shift in criminal activity may motivate his decision, but even in that case his information has still been improved, allowing him to make a more rational decision.

After knowing the direction of the policy change the next most important point is knowing the general magnitude. Suppose in the prison reform case there will be a shift in criminal activity upwards, but the decisionmaker because of moral reasons still would like to make the reforms. The magnitude becomes significant then in weighting the alternatives. Even then the decision may be difficult to make. The general shift may not be accurate enough if the decision is very close to going either way. In this case the decisionmaker can no longer be helped by the model unless its accuracy is improved, which beyond a certain point will be impossible to do with any confidence.

The elements which are necessary in this case, however, are knowing some of the structural elements involved in the process. What is there about prisons that deters virgins (those who have not committed crimes) from committing crimes? What reforms would lead to prisoner rehabilitation themselves? In order to construct a policy that is of any use it would be necessary to know some of the structural elements of the situation.<sup>7</sup>

Beyond the short-run consequences of most models are long-run predictions about trends. Since most policies are not designed just to meet immediate goals but have long-range implications and ends, the model must have some method of predicting changes within the range of consequence. Maintaining model validity, however, can be a difficult job.<sup>8</sup> If, for

instance, the consequences of prison reform was first a decrease in crime but later had a psychological impact that changed the attitudes of people outside of prison towards being put in prison (it lost its stigma because of the reformed prisoners' ability to readapt to society) and crime sharply increased, the decisionmaker would feel that the simulation had failed him. In order to be of use it is necessary that the model can be invalidated at some later point if structure is changed in some way the model could not anticipate. In order to do this some basis for invalidation must exist. The problem is very difficult, however, because even if one feels relatively sure that for a given time period the model is valid it is difficult to specify the way in which one would know when the model is breaking down. However, once the model is used for making the decision it will be necessary at least to some extent to monitor the decisions which have been made by monitoring the basis (the model) of those decisions. For humans models generally learn on their own, while computer simulations do not.

More generally to be of use the models must be directed to describing some particular problem or system goal.<sup>9</sup> This, in a sense, is a more general problem than just defining the usefulness criteria for simulation models, but is a general analytic imperative. Policies should have specific purposes, and models should be constructed to examine those areas. Without following this usually correct piece of 'folk-knowledge' it is difficult for the model to include all the important

factors defining a behavior from a control point of view. Policymakers are interested in control, not theoretical elegance.<sup>10</sup> Returning to the prison reform scenario, one must ask specific questions and have specific goals. To do a model of the prison system and society would be enormously difficult. Furthermore, "to the extent that a model is large for complications' sake its utility is decreased."<sup>11</sup> It becomes difficult to understand the implications of policies along with maintaining validity over time. Even a general model of prison reform may not encompass the proper questions, or allow the policies the decisionmaker wishes tested to be executed. Because of the interaction of parts of social science models it is important to ask directed questions.<sup>12</sup>

The use of models then is a difficult one. The demands of the decisionmaker can vary with his need for information, although he is always concerned with controlling the process or influencing it with the controls he has at hand. The psychological problems of the modeler in constructing validity tests and the dual, invalidity tests (for long run use), have only been touched on and will be returned to later in a separate chapter. It is hoped that the reader now has the sense of what in general models must do to be of use in decisionmaking. It will be valuable later in more technical chapters on validation methods to keep these necessities in mind.

FOOTNOTES

1. Brewer, Evaluation and Innovation in Urban Research, page 12.
2. Ibid, page 8.
3. Wright, Validation of Dynamic Models, page 34.
4. Gordon, System Simulation.
5. Forrester, "Industrial Dynamics - A Response to Ansoff and Slevin," page.602.
6. Kleindorfer, Lectures.
7. Wright, Op. cit. page 34.
8. Wright, "An Industrial Dynamics Design for Improved Profitable Growth," 1970 Summer Computer Simulation Conference.
9. Schlager, How Managers Use Industrial Dynamics.
10. Brewer, op. ci.t, page 12.

11. Ibid., page 13.
12. Van Horn, "Validation," in Naylor, Design of Computer Simulation Experiments, page 235.

CHAPTER 5

The problem then is to decide what validation criteria can meet the criteria necessary to make policy making simulations useful. In order to understand the need for change from the present methods it will be necessary to understand the present simulation process; both model construction and validation. Without such an understanding it would be impossible to realize the failures of present model construction, validation, and experimentation procedures.

The usual steps of the process are said to be

- (1) problem definition
- (2) model construction
- (3) model validation
- (4) model experimentation

The steps proceed recursively, although there exist feedbacks.<sup>2</sup> There exists a separation between experimentation and validation.<sup>3</sup> This separation means that one first validates the model, then decides what experiments should be run on the model (what policies should be tested). Naylor has defined the two processes as separate; validation is concerned with replicating time paths, experimentation with the optimization of factor and responses.<sup>4</sup> Naylor argues that the two processes are separate. Validation tells one whether one can use the model, while experimentation

or policy testing is concerned with whether the model will reach an optimal point on the objective function. It cannot be overemphasized that the majority of the literature sees the process as separate steps; even where feedbacks exist they do so only in the early parts of the process; once the experimentation phase is reached the model should be set.<sup>5</sup> This is exhibited in several major models, including Forrester's Urban Dynamics, Hamilton's Systems, and throughout the Conference books on simulation.<sup>6</sup>

The validation issues and steps are based on these assumptions of separate steps. The primary and most listed criteria for validation is the prediction of time series.<sup>7</sup> Naylor and Finger define three major steps in what they call multi-Stage verification.

- (1) Choose possible hypothesis
- (2) verify postulates statistically
- (3) test model's ability to predict behavior statistically.<sup>8</sup>

Their primary concern, however, is with step three; throughout the literature they are quoted for such.<sup>9</sup> The main emphasis in terms of method though is to see how well a model reproduces history.<sup>10</sup> Crepine suggests that two questions be asked:

- (1) Will the model yield similar outcomes as the real world?
- (2) Does it use the same mechanisms to reach the outcomes?<sup>11</sup>

Such an addition as point two attempts to make sure the model has some structural significance. Other simulation researchers put greater emphasis into structure; Martin argues that one must have validation of theory on which the model is based, and then validation of the computer program.<sup>12</sup> In the same book Hans Anderson suggests four types of validity; structural - it resembles the world; overall - has same behavior; technical - the model and the computer program are isomorphic; and subjective - it is convincing.<sup>13</sup>

Emshoff has more far-reaching suggestions. He suggests five methods for validation:

- (1) internal validity - when exogenous inputs are held constant does the variability range over a large set of values? (He argues that if yes the model is difficult to assess.)
- (2) Does the model have face validity - does it appear to be correct?
- (3) How are the variables - parameter sensitivities - are they reasonable?
- (4) Hypothesis validity - do the assumed relationships make sense?
- (5) Does the model predict well?

As with most authors Emshoff returns to the prediction test as the most important.<sup>14</sup>

Amstutz suggests almost an identical list of conditions which are called widely different names; viability, stability, consistency, duplication of historical conditions, and prediction of future states.<sup>15</sup>

Finally Poole suggests a confidence test where it is argued that the more independent variables that exist the more dependent variables must be predicted.<sup>16</sup>

Such lists of conditions can, however, distort the main emphasis of the authors - the prediction of events whether in the past or the future. Without many exceptions the main criteria is that the model fit the past data. Crecine, for instance, emphasizes the use of goodness of fit and consequently the chi-square test.<sup>17</sup> There exist notable exceptions to this emphasis such as Forrester, but their objections will be returned to later.

Given these general conditions the same authors suggest numerous tests to accomplish this goal. Naylor lists nine major tests:

- (1) the number of turning points
- (2) the timing of the turning points
- (3) direction of turning points
- (4) amplitude of fluctuation for time segments
- (5) average amplitude over the whole series
- (6) simultaneity of turning points of different variables
- (7) average values of the variables
- (8) exact value of the variables
- (9) measure of the variation around the mean.<sup>18</sup>

Most of these tests concern statistical validity. The comparison of input-output transformations is the main tool of most validation. Van Horn points out that the representation of means, ranges, and variance are the crucial points in this type of analysis, and that even graphical analysis can be useful here.<sup>19</sup> Recently the technique of spectral analysis has been suggested by Fishman and Kiviat, in order to surmount some of the difficulties encountered in ordinary statistical testing procedures (such as autocorrelation, multicollinearity, etc.)<sup>20</sup>

Such statistical tests primarily are concerned with checking whether the simulation can reproduce historical time series data. Once the model can achieve such reproduction then the next most emphasized criteria is that the model also have some of the structural elements of the real world. A test often suggested for such a purpose is the Turing test.<sup>21</sup> As originally conceived the Turing test was to be a test of artificial intelligence. Conditions were created where a person would ask a machine and a person questions, without knowing which was which. When the person could not tell the difference between the machine and the person's response the machine would be said to think.<sup>22</sup> Van Horn describes an adoption of this technique to validation where the people in the system are given data from the system and the simulation and asked if they can tell the difference.<sup>23</sup> This test would be fairly good where people in the system have a good idea of the functional

characteristics of the data; however, where they do not or uncertainty exists, such as in the types of policymaking models about which this paper is concerned, it is not likely to be the case.

Summarizing the present techniques is difficult, as one can see from the variety of methods mentioned. However, basically in the order of importance, simulation researchers look at (1) reproduction of time series - accomplished with statistical tests (2) structural validity - which is largely a question of confidence of the modeler, but also for which the Turing test can be useful. The concern in both cases is with the historical data however. Even in consideration of the structure of the simulation the Turing test considers past structure as what the model should seek to replicate. This point cannot be overemphasized; validation as posed here is concerned with the historical data and its inclusion into the present simulation.

There exists notable exceptions to the previous conditions, most well known of whom is Forrester. Forrester argues that false models could fit history but not be sufficiently rich enough to allow testing of new policies.<sup>24</sup> He argues that to be of use in making new policy, models must represent the mechanisms of reality. However due to the amount of noise which enters into models the ability of the models to predict future, or even past data is severely curtailed.<sup>25</sup> Rather models

must be correct in all parts. The validity of a model is examined in relationship to all variables; does the system boundary include all the important interacting variables? Does the model endogenously produce the behavior the simulation is examining? Is the decisionmaking stream compatible with the hypothesis of decisionmaking in the model? Forrester is interested in the creation of the symptoms of the model. Do the same dynamic problems occur? He suggests tests such as checking the behavior of the intervals between peaks, time phase relationships (general similarity of data to world), and transition characteristics of different modes. Forrester in a sense is more concerned with making the new system conform with the model than the model with history.

Wright extends Forrester's criticisms of the method of validation by prediction. "A device that reproduces or predicts need tell nothing about the process being examined."<sup>26</sup> Furthermore, he argues that data collection implies a theory of data which must be validated; using just the data to validate the simulation is begging the question. On a more practical level Wright argues that errors enter into simulations in three ways; errors in measurement, errors in specification of the simulation and errors in non-specified exogenous and non-random forces which may affect the model's process. This confounding of errors makes it impossible in any models where these factors are significant to predict data accurately, although they do not

necessarily remove the policy value of a model with them. Wright continues his attacks by arguing that the statistical tests on which models are validated, and variance is measured assume normality, statistical independence, and common variance which often only hold for simple systems.<sup>27</sup> Hidden and missing variables which often come out only over time make it difficult to determine which model is correct.<sup>28</sup>

DeNeufville also argues that "the statistical closeness of an equation to a set of observational data on a system is not sufficient test of its validity."<sup>29</sup> None of these critics address the problems of the interaction of experimentation and validation. Forrester while attacking the non-structural element of other models than his own claims in fact that his models are structural.<sup>30</sup> Wright is more frank and admits that at the core all models are transfer functions. Dynamic models, as he defines the term, he argues, show more of the structure than black box or econometric models; they show boxes within the box.<sup>31</sup> He then goes on to argue that one can always seek a deeper level of meaning; the problem is to determine when a sufficiently stable level has been reached. However, nowhere does he begin to examine this question nor face the issue of when models will be invalid. The problems to solve is what are the sources of model instability and what methods can be used to measure and check it so that it can be corrected, along with searching for methods to make models more stable.

At this point, it will be worthwhile to assess what remains to be done. The purpose of this paper is not to prove that any specific persons are doing validation wrongly. Therefore if the reader does not agree with this paper assessment of present techniques that need not keep him from accepting some of the ideas on how to improve the validation process. Essentially it will be necessary to examine the factors in models which influence validation, the problems they create, and some mechanisms for correcting the problems. Then the psychological issues along with the costs and benefits will be examined. Finally a model will be examined in detail to illustrate some of the concepts which this paper discussed.

The issues that have been raised here seek to show the weaknesses and methods of simulation as presently used. The problems which beset these techniques will be discussed more fully in the next two chapters.

These methodological problems which will be discussed should not remove interest in resolving the problems of actually effectuating present techniques. Determining goodness of fit, the traditional techniques of using only part of the data available to make the model and then using the model to 'predict' the non-used data, and other traditional methods have intricate problems associated with their implementation. The resolution of such issues is important; the thrust

of this thesis will be to raise a series of methodological problems associated with validation even when such traditional techniques are improved and to suggest some techniques for dealing with these new problems.

FOOTNOTES

1. Hamilton, System Simulation for Regional Analysis, page 99.  
Gordon, System Simulation.
2. Hamilton, op.cit., page 99.
3. Meier, Simulation in Business and Economics, page 23.
4. Naylor, Management Science, March 1970.
5. Naylor, Design of Computer Experiments, page 25.
6. Summer Computer Simulation Conference, 1971.  
Forrester, Urban Dynamics.  
Hamilton, op. cit., page 99.
7. Beshers, Substantive Issues in Models of Large Scale Systems, page 121.
8. Naylor and Finger, Management Science, October 1967.
9. Van Horn, "Validation," in Naylor, Design of Computer Experiments, page 236.
10. Evans, Simulation Using Digital Computer.
11. Crecine, Governmental Problem Solving, page 62.
12. Larsen and Landin, "Some Methodological Views of Digital Simulation," Models and Simulation, edited by Stockhaus, page 262.
13. Anderson, in Stockhaus, op.cit., page 32.
14. Emshoff, The Design and Use of Computer Simulation Models, page 205.
15. Amstutz, A Management Oriented Behavioral Theory of Interactions within Consumer Products Markets, page 60.
16. Poole, Candidates, Issues and Strategies.

17. Crecine, op.cit., page 64.
18. Naylor, M.S., October, 1967.
19. Van Horn, op.cit., page 241.
20. Fishman and Kiviat, Digital Computer Simulation: Statistical Considerations.
21. Crecine, op. cit., page 62.
22. Feigenbaum , Computers and Thought.
23. Van Horn, op.cit., page 242.
24. Ibid., page 116.
25. Forrester, Industrial Dynamics.
26. Wright, Validation of Dynamic Models, page 34.
27. Ibid., page 26.
28. Ibid., page 49.
29. de Neufville and Stafford, Engineering Systems Analysis, page 6.
30. Forrester, Urban Dynamics, page 14.
31. Wright, op.cit.

CHAPTER 6

The construct which most of the literature in simulation advocates - model construction, validation, and then experimentation - ignores some of the fundamental aspects of policy-oriented simulations. Instead the simulations take on the tone of academic exercises. For policymaking simulation to be of use it is necessary that they have sufficient sophistication and power to be able to incorporate the different alternatives and policies which decisionmakers have available into the simulation. In formulating these policies simulation researchers cannot just think of functional changes which they will impose on the system, but must design real policies which are implementable. Thus every policy will be aimed at some functional characteristics in terms of the model but will not just have those effects, but also the effects of the implementation of the policy. For one can never just change a functional characteristic magically, it must be done through some implementable means. What this means to the separation of experiments and validation is that such a separation can no longer be assumed. The boundary condition of side effects of implementation of a policy will be a program with no other effect than the functional characteristic - the rest of the policy situations will have some other effect. This will make it necessary to formulate tests to tell when these other effects

will have a consequence on the model itself. Even without these 'external' effects the model might be effected by just functional experimental manipulations. For instance, the bounds of the model's description of reality may be fairly small and experiments outside that range might require changes. Another case would be a policy which required simultaneous changes of experimental factors. In some situations this two variable change would cause a change in the simulations structure. Therefore it is necessary to include in any set of validation procedures a method of deciding whether the effects of experiments and policies would change the structure of the model.

The essential point to be made here is not that it is unnecessary to validate the model, but rather that for the model to be validated it must be done for every policy tested. This is not to argue that the whole set of tests must be redone. Clearly many would still be valid; others though must exist to test whether those tests are still valid and the simulation remains structured the same and thus valid.

Returning to the issue of experimental process and validation the fact that such revalidating procedures are necessary is suggested in the statements of some of those who presently do not practise this "experimental validation." Berger argues that computer simulation controls all aspects of the environment.<sup>1</sup> Such a statement while

probably meant as a proof of the theory building power of simulations also shows their weaknesses under present conditions where all aspects of the policy experimentation and its effects on the environment and the simulation cannot be controlled. In a British paper on gaming/simulations it is argued that "gaming/simulation attempts to identify areas in which relationships between variables may change; mathematical modeling is concerned with the projection of consequences, given relationships which are known or changing in a predetermined manner." This is only true in the barest sense, for not all the relationships in simulations on computers can be characterized this way, at least if they are to be useful. The first point which must be examined in making any policy test is to see if present relationships will stay fixed; this is the crux of the argument. Later it will be pointed out that continual validation through monitoring is necessary if the simulation is to be of any use.

Perhaps the most useful arguments to examine, however, are those given by Naylor in his book Design of Computer Experiments. Naylor argues that "The effect of variations in response which these factors cause in the real world have been incorporated in the computer simulation model in the form of experimental errors or random deviations. Once we have a model, the factors are determined and it is not possible to experiment on the model to identify additional sources of variation."<sup>2</sup>

Furthermore, he goes on to say "In computer simulation experiments one never has uncontrolled or unobserved factors. The role played in the real world by uncontrolled factors is played by random character of exogenous variables in computer simulations." <sup>3</sup> Finally he states that "In a computer simulation experiment on a given model it is not possible to include more factors for error reduction purposes. The inclusion of more factors requires a change in the model. Once the model has been specified all uncontrolled factors have been irretrievably absorbed in the probabilistic specifications for the exogenous inputs." This last point is exactly the reason every simulation must validate every experimental possibility, for even the level at which a policy is tested makes it a different policy and thus may incorporate new factors. It is necessary to change a model to test different policies or even to do sensitivity experimentation because different levels of policies might create vastly different implementation consequences. It is also clear that in examining the earlier statements by Naylor that these implementation effects of experimentation may introduce non-random exogenous factors for which the model does not account nor even increase the level of random change so that the model's stability could change. In a model of more than one stable state their level of random input might be an extremely important fact. Since most policies will create additional sources of variation contrary to what the first Naylor statement argues it is necessary to identify if this variation is crucial,

random or non-random, and whether it will necessitate any additional changes in the model to take it into account.

The arguments that such considerations need not be made can only be seen in the light of assuming that the boundary conditions hold, which implies that all models are perfect structural representations of reality and that these structures do not change over time. While some social scientist may believe this and may even have achieved it at certain gross levels it is certainly not true that this is the case at the policy level at which this paper deals.

Before exploring the possibilities of controlling for this change in structure caused by experimentation it would be helpful to delve into some of the causes of the changes. Often in policy-oriented models there exists many human elements. In such systems it is necessary to understand the interaction of individuals with the environment, and how each affects each other.<sup>4</sup> This is difficult to do though. Emshoff in the same article pointed out some of the problems in the project with which he was involved. In that case a naive model predicted outcomes better than the simulation model that was designed to generate such behavior. Emshoff argued that the common use of simulations was where all the relevant factors had been identified and research interests focused on studying the effects and performance of the system as was affected by joint interaction of all factors; the environment was

completely understood. This the reader will recognize as characteristic of models,<sup>5</sup> defined earlier as seeking greater accuracy. Emshoff states that most of the present thinking fails when it is applied to simulations where the factors causing behavior are not all known.<sup>6</sup> The problems of such simulations are that policies can easily add factors; the simulations in policymaking are not very well understood.

Briton Harris in discussing urban simulations brings up a similar viewpoint. In this analysis the experimental approach consists "in searching out instances in which the environment and its informational content differ significantly from other environments or its informational content differ significantly from other environments or the decision units differ significantly from other decision units, so that the general application and fruitfulness of a theory may be examined. When however, the behavior of the decision units as understood on the basis of such analysis is to be explored experimentally under changed assumptions as to policies and technology, an entirely new situation arises. We can no longer assume that various sets of decisions are independent of each other."<sup>7</sup> Such a point argues that the policies which one tests in a simulation may not be independent of the structure of the simulation, although Harris is not clear in specifying the implications of this fact. He does argue though that "system simulation and decision analysis interact strongly with each other..."<sup>8</sup>

One can surmise from this that Harris is not completely satisfied with present procedures.

Naylor's argument that all the random variation and unexplained forces of the model are contained in the random specifications of the model clearly breaks down under the conditions specified above. Since the amount of the random influences, along with the amount and type of non-random influences will be influenced by not only the level at which a policy is tested, but also how that policy is to be executed it is necessary to account for these influences in the model. Since the model is fixed in the conception of what factors are important to behavior along with its treatment of random noise it is necessary to determine whether these factors will influence the outcome of things and thus the model. It is arrogant to assume the model will include a description of all situations such that no changes are ever necessary. When the informational content of different units in the model differ they may be affected by different levels of policies in different ways and certainly may be affected by different mechanisms of implementation of policies differently. To assume otherwise is to assume a universal reaction to the same stimuli. The whole literature on aggregation attests to the difficulty of achieving this even on the grossest level. One must not mistake one's goals for one's accomplishments.

Continuing with the discussion of why separation of experimental procedures from validation cannot exist for policy-oriented

simulations it will be worthwhile to examine the comments of Van Horn on this area. Van Horn states that management science models have four characteristics:<sup>9</sup>

- (1) the structure and the parameters of the process are determined by the environment, NOT the modeler
- (2) part, and only part of the process depends on physical phenomena
- (3) people are part of the process directly as information processor or decisionmakers or indirectly as consumers
- (4) the process depends on many parts and the behavior of the process depends on the interaction between the parts.

Most of this applies to policy-oriented simulations with a few additions. The most important is that the process depends on the interaction of the parts and one of the parts is the environment, which may be affected by the parts in the process under varying circumstances. Thus the environment of the model can be influenced by experiments of the modeler and influence the structure of the model. Forrester argues that one always includes all endogenously produced behavior and thus feels that he avoids this problem. In Urban Dynamics he even adds that environment and model are affected in linearly additive and nonstructural ways.<sup>10</sup> Nowhere does he specify how to tell if this is not true and if it is when the model must be changed.

Van Horn recognizes these problems. He sees that model construction assumes a stationary world where distributions do not change.<sup>11</sup> He gives an example for which under one system the assumptions of a model might be true but for a policy change the assumptions might not be true. In this case it is batch processing orders that may appear as a non-Poisson distribution but if switched to immediate processing they might be Poisson.<sup>12</sup> He also points out the problem of time varying parameters. In another example of the control of airplanes he points out a situation where a simulation had to be changed because a policy suggested (functionally it was speed-up work) changed the reactions of people making the control decisions (the air controllers).<sup>13</sup>

Earlier it was mentioned that the aggregation problems are similar to the ones being investigated here. Certainly this is true although the perspective is different. In economics to aggregate it is necessary that variables be functionally separable.<sup>14</sup> For example:

L	B'	B''	P
100	20	5	1000
100	18	6	1000
200	20	5	1800
2000	18	6	1700

In this three factor problem the aggregation would break down because of the interaction effect. One would not expect a P of 1700 in line four.

If one built a model it would be necessary to specify conditions under which aggregation would break down for the production function as stated in the first part of the example. Changing the level of factors might change the structure of the equation. Thus the aggregation function changed. Now it is possible to argue that, in fact, it did not change but that the description was not adequate. This may, in fact, be epistemologically correct, however, the issue is understanding when the description being used does not hold, for no model is a complete description of the universe. Clearly then the area of the aggregation problem identifies an important characteristic for the validation process; checking when a process is functionally separable.

Krasnow, in the Design of Computer Experiments, is aware of some of the problems discussed in this thesis. He says "The best situation is one in which a fixed set of problems is defined in advance to be addressed by the model and experiment."<sup>15</sup> By this he hopes to include all structural elements in the model for the issues being considered and avoid the problems with which this paper is concerned. He makes explicit recognition of the problem four pages later.

"Quite often the primary experimental factors are structural elements of the model. The comparisons of two totally dissimilar storage devices in a given computer system might require separate structures to serve as experimental factors. Comparisons of operating decision rules or planning policies might require different behavioral

descriptions to serve as factors of an experiment. Good model design will seek to minimize the difficulty of establishing different factor levels during an experiment."<sup>16</sup>

Several extremely interesting points emerge from this quote. First is the recognition of the impossibility of ever building an overall structurally correct model. Second is the goal, nevertheless, is trying to achieve this. Finally is the explicit recognition of how to handle the problem at least in one situation. However, because Krasnow is concerned with simulations aimed at design of non-existing systems he does not consider the importance of his remarks to other types of simulations, so that the type of monitoring of model validity suggested here would be necessary. Rather he argues that one must hypothesize any structural changes. This is true in design, not in policy-oriented simulations where less closure exists and one is dealing with existing systems and other steps are necessary. Furthermore, he accidentally leaves out the case of range of the same functional policy possibly affecting the structure of the design. Clearly this would be true in the example he discusses of design of a computer if the device became extraordinarily fast as to cause breakdowns in the capacity of other parts of the computer being simulated to be able to withstand the pressure.

Given these wide and varied reasons for the necessity of monitoring structural change it would be worthwhile to pursue the

theoretical reasons and structure of these arguments. Fortunately Hebert Simon provides a good guide to this area.

Simon has defined in several places including the Sciences of the Artificial and the Structure of Social Science Models the necessary conditions for understanding models. Generally one seeks to make models hierarchial.<sup>17</sup> By this it is meant that they have controlled variables given variables (the environment) not influenced by variables in the system, and endogenous variables influenced by others in the system. From this it should be clear that the policy sensitive situation discussed in this thesis is when the hierarchial nature of a model breaks down, and later how to restore it. There exists an important theorem though about such hierarchial systems which recognizes that they are difficult to describe in social science. The essence of the theorem is that when a model is nearly hierarchial then it can be treated as if it were. Essentially this argument rests on the fact that nearly decomposable (the conditions described above) systems tend in the short run not to differ very much from totally decomposable and therefore hierarchial systems.<sup>18</sup> Ando and Fisher neatly sum up their theorem. "Our theorem shows that if these feedbacks are sufficiently weak relative to the direct influences, that is, if the theoretical assumptions are sufficiently good approximations, there exists a time  $T$  greater than 0 such that before  $T$  the behavior and stability of the

economic system can be analyzed in isolation without regard for the difficulties raised by the presence of such feedbacks."<sup>19</sup> The problem then is when do the feedbacks exist so that the description being used is invalid? If one seeks only to describe the system T can be significantly different than 0; however, if one seeks to change policy this may not be true.

Given these assumptions the arguments made in this thesis gain strength. Fischer in an earlier article in the same book discusses Liu's criticisms which led to the whole recognition of the problem. These criticisms are directed to estimation problems and simultaneous equation models but can be easily extended to all forms of modeling. Liu's arguments parallel those contained in this paper.

- (1) there are more variables than in an equation
- (2) more equations than in the model
- (3) the restriction that some parameters in a structural equation are zero is incorrect.

Therefore, Liu argues that the system is underidentified.<sup>20</sup> Fischer counters that it is a matter of degree, and that when the influences are small the problem is overcome. This clearly is true. The problem, however, is to define weak, when weak will become strong, so that when the model becomes invalidated it can be fixed or discarded. And although

the identification problem is not exactly the one with which this paper is primarily concerned it does affect the specification of simulation models. Furthermore, the general discussion of hierarchial models is broadly applicable to all models. Unfortunately the issue of how to measure these areas of hierarchy are not included in the discussion nor are the discussions extended to the logical consequences for models which seek to change policy.

This theoretical framework then defines the situation which it will be necessary to overcome. The reader may still not be convinced that some policies and experiments will change the structure of models. They may argue that even if the points argued so far are correct there exist delays in every system which make it highly unlikely that such changes in structure will occur rapidly. This argument however, fails to define the necessary characteristics of 'rapidly', for its only meaningful context would be a period more than necessary for the policy to be effectuated. This demands that there exist some test whether this period is short enough which is a restatement of the types of tests suggested so far. Not only will it be necessary to evaluate when a policy must necessitate a model change, but when such a change might occur later on and under what conditions.

The problem of describing a time horizon over which the model will remain valid is a difficult methodological problem. Given the assumptions of changing model structures as a consequence of the inclusion of unknown external variables becoming important the same problem will exist for simulations which test policies over long periods of time. It will not be possible to validate the model once and for good but it will be necessary to monitor the validity of the model to see if it is still performing to suggest good policies. Therefore any policy which relies on time lags to be effectuated will necessitate the same type of procedures of validation as the model did in originally testing the policy throughout the implementation of the policy in time.

The face value of this argument becomes clear when one considers the learning process; man does not always respond to the same stimuli with the same response, he is capable of learning.<sup>21</sup> The consequence of these arguments is that the tests to be devised in validating the model, particularly with regard to particular policies should also include an ability to validate models over time. From this it would appear that the strongest tests for this task are those which will ask the question, "under what conditions will the simulation become invalid?"

The changing structure of models as a consequence of the

implementation resembles the Hawthorne effect. This notorious effect in social science was noticed when some industrial engineers decided to test what lighting conditions were most conducive to productivity. First they increased the intensity of the light, and productivity rose. Then they decreased the intensity and production continued to rise. In fact, they found that just about anything they did had a positive impact on production. From this they deduced that some other factors than the ones included in their model of production and lighting were at work. It became clear that the changes themselves had changed the cognitive assertions and the motivation of the workers. The name of this became the Hawthorne Effect for the factory in which they were working.

The causes of Hawthorne Effect are not fully understood although some research into the area has been conducted. An example is Rosenthal's study of IQ and expectations.<sup>22</sup> Essentially the Hawthorne effect is caused by changes in what people see as the environment. They have certain beliefs or operating procedures which are based on some assumptions about the environment (that is what is fixed). When the experiment or policy is done these assumptions are changed which results in the model that would describe their behavior also being changed. In the case of the Hawthorne experiment the workers' attitudes on how they were being watched changed their behavior by affecting these assumptions. When an experiment to determine the effects of lighting changes occurred they reacted to their change in the perception of being watched (and

probably self importance) and worked harder. If one could predict the direction of those changes one could still run an experiment; the problem is that too infrequently are the results of such changes known because the assumptions which are being effected are not known and no experience with the change has occurred which encompasses the necessary experience.

The implication of the Hawthorne effect, which is a subset of the types of changes described earlier is that the use of large-scale policy-oriented models may introduce even more complex problems than experimentation. The reasons for this are that the areas which many of the policy-oriented models are seeking to control are areas in which many of the actors are independent and have separate goals, some of which may conflict with the goals of the policy being tested. If the model is made public then the actors with different goals might change the behavior on which the model was predicated since it would be clear that such behavior would not longer help them to achieve the goals. For example in a housing model a policy oriented to controlling rents might lead to speculation and increased rents because the developers might radically change their behavior and even use the simulation to find a position that maximized their returns. Therefore, it is possible that models which were made public would adduce a double Hawthorne effect (similar to the double whammy).

Even where the goals of the actors in the system were in sympathy with the goals of the policy bad effects could result. In this case actors would change their behavior to try to make the policy even more effective; however, if behavior changed more than planned the results could lead to an overkill and even counterintuitive consequences. Too much could result in worsening of some conditions which have an unstable equilibrium or more than one equilibrium.

If the model is used on a gnostic basis some of the second order Hawthorne effects may be washed away although the original ones described in the first sections of the paper and the previous description of the Hawthorne effect will still exist. Ideally some simulations would even include the 'learning' curve of the actors such that structure remained solid. The problem, of course, is to be aware when this rare phenomenon occurs (the validation problem). Of course, the nature of policy-oriented simulations makes their implementation for at least democratic systems unlikely to be gnostic. The consequence is that the implementation effects will be both indirect in changing assumptions on which behavior is based and also direct in giving the various actors conscious control of behavior in many instances.

Awareness of these effects is not totally lacking from the literature. Outside simulation there exist a great many discussions of the problem; a notable recent one with applicability to simulation being

Donald Campbell's "Reforms as Experiments."<sup>23</sup> Campbell is concerned with the problems of using reforms as experiments and the problems of doing experiments in society. As a consequence, to an experimental technique like simulation he has much wisdom to impart. Forrester is also aware of this problem when he says "models affect structure of behavior."<sup>24</sup> However, he does not extend the consequences of his point to the rest of his work and, in reality, by arguing that all exogenous forces should be excluded from the model negates the value of this observation. In econometrics the importance of this point is understood. The estimation of points is predicated on the structure of behavior remaining constant.<sup>25</sup>

Given the independent nature of many of the actors and segments of policy-oriented simulations the importance of exogenous changes in the universe also becomes important; one cannot assume a linear response of all elements. The biggest proponent of assuming such a linear response is Forrester in *Urban Dynamics*. There he argues that because all forces will affect the environment and the model equally the forces will have the same effect. He does not consider that within the model are different behaviors which also might be affected differentially. However, since different subsystems depend on different assumptions and facts and different actors have differing cognitive assumptions upon which they depend, it is clear that exogenous changes

outside the model's domain may have an influence on the structure of the model. Therefore it is necessary to monitor the environment when instituting policies to be sure that the simulation remains valid. Later the Urban Dynamics model will be examined to illustrate this point clearly.

The previous discussions of model structure have led to some very important criteria in using policy-oriented simulations. Models have been shown to be semi-structural representations of reality; one seeks to achieve hierarchial structural models which describe all situations but never reaches that goal. Instead there exists a transfer function in which some structural elements may exist but from which certain important factors are of necessity left out. The model is still only a vague picture of reality. Certain factors upon which the model rests are, therefore, not included in the model but are part of the environment and may, to some degree, be completely unknown. As a consequence, in the testing of any policies and in doing any experiments it is necessary to be able to know when such factors might enter into the model and affect its structure. This is true for every test of a policy along with the testing of policies which take time to implement or may be continuing in nature. The problem then is to devise tests which will include the ability to invalidate a particular model structure

and even suggest when and how such a structure must be repaired. To ignore this problem because it is difficult to solve means that one has implicitly chosen the case when there are no effects and the model is completely structural. For policy-oriented simulations this case is extremely rare. The necessity then is to be as rigorous as possible in devising and using tests to maintain validity in the face of the cost constraints of the whole validation process. Within that factor it will be necessary to examine the separation between model construction, validation, and experimentation and make those three actions as recursive and interacting as the problem demands.

FOOTNOTES

1. Berger, Simulation and Games, "Simulation and the City."
2. Naylor, Design of Computer Experiments, page 11.
3. Ibid., page 11.
4. Emshoff, "Roles for Simulation in Behavioral Research,"  
1971 Winter Simulation Conference, page 27.
5. Ibid., page 28.
6. Ibid., page 29.
7. Harris, "The Use of Theory in the Simulation of Urban  
Problems," JAIP, page 269.
8. Ibid., page 269.
9. Van Horn, "Validation," in Naylor, Design of Computer  
Experiment, page 235.
10. Forrester, Urban Dynamics, page 20.
11. Van Horn, op.cit., page 237.
12. Ibid., page 238.
13. Ibid., pages 243-250.
14. Green, Aggregation in Economic Analysis.
15. Krasnow, "Simulation Languages," in Naylor, Design of  
Computer Simulation Experiment, page 322.
16. Ibid., page 326.
17. Simon, Sciences of the Artificial, page 17.  
Structure of Social Science Models.

18. Simon, The Science of the Artificial, page 99.
19. Simon, Structure of Social Science Models, page 101.
20. Ibid., page 33.
21. Skinner, Walden II.
22. Rosenthal, Pygmalion in the Classroom.
23. Campbell, "Reforms as Experiments," Urban Affairs, 1971.
24. Forrester, Industrial Dynamics, page 125.
25. Kleindorfer, Lectures.

## CHAPTER 7

Achieving the goals of thorough validation is not an easy task, particularly in light of the constraints of time and expense which keeps the number of runs one is able to make within limits. The validation problem has been defined as making sure the model provides some information which is useful to decisionmakers. This task must be accomplished in every test of the model and for every run, at least to some minimal degree. The problem then is to devise techniques which give the simulation researcher and the policymaker more confidence that the model is correct. In social science it is often useful to define under what conditions the models one has devised will break down. To some extent the more traditional tests of validation can provide some guide to believing in a model, although little guide when not to believe in a particular model. Defining the appropriate conditions in which a model will break down then means in a sense validating its structure for the particular question at hand.

The first technique that will be discussed is the use of boundary conditions in simulations. Clearly no modeler intends his model to represent the state of the world under all conditions unless he is concerned with either definitional models or trivial events (from the point of view of policy). However, very few researchers ever define any boundary conditions except perhaps the general one known as 'all things

being equal' cetera paribus. This, however, really begs the question since all things are never equal; the important thing to do is to define what 'all things' really means, to define what must be 'equal', in fact, for the model to hold. There are several techniques in which this may be possible. One is to define the boundaries in which the model can operate. In doing this one defines a certain region in which the model is stated not to work; where all things will not be equal enough.

There exists several methods of imposing boundary conditions on simulations all with different implications in terms of what is being controlled and what is recognized as not being controlled, although all have the same effect of defining an area in which the structure of the model is said not to hold.

The first method is to bound the domain of the input variables. Doing this says that for certain situations the movement of some variable to too great a variance leads to the emergence of a new structure. For instance, in an economic model it is possible that the movement of wages to too low a point would lead to more chaos than the model would lead one to expect, perhaps through a revolutionary mechanism, or perhaps through increased expenses on police to repress revolutionary expenditures. For the purpose of the modelers it may be necessary to bound the domain of action that the government may take in driving down wages. On the upper bound minimum wages that exceeded

the concept of the managers' of 'fair' might do more to curtail production than the model might predict because managers would begin to resign or lose their motivation. Of course, it is possible to argue that the model was not a correct description of reality and that the boundary conditions should indicate a 'flip-flop' such as a step function to a new structure would accomplish. This is an unnecessary assertion. One might not know the new structure although one could believe that the old would be totally wrong, or one might now wish to consider it. Essentially the flip-flop argument is that one should have a series of models to describe the changing structure of reality; not a bad suggestion but not very pertinent to the problem of defining such boundaries.

Determining the exact point of such boundaries is not an easy task. The most important method is to try to deduce from the theory when such boundaries should exist. More formally the bounds may be said to exist where instability of the model has been observed in the past. If the variance of output was extreme for some extreme value such that the model is sensitive to small movements in the extreme the model may not be too useful in that situation. Generally, however, the problem is that the model is not too sensitive but the real world either is, or the world becomes sensitive to new things. Thus no guide for establishing guides to bounds of inputs exists that is absolute; the best that can be said is the violation of theory cannot generally be acceptable. A

model where wages were to rise to too high a point to make the price uncompetitive would not be allowable even if the model were a good guide within very small rises. It would be necessary to bound the variable.

Likewise the range of the output of a model can be bounded. For instance, in building a weapons system it is necessary to bound the range of the output to theoretically feasible solutions, even if bounding the inputs is not necessary. One cannot blow up the world, for instance, five times. This type of bounding is usually not needed because it is not usually the case that variables are described in worlds blown up or such metaphorical terms, but occasionally, it is necessary, even in conservative subsystems (systems that just pass on physical quantities). The case of Newtonian mechanics is a clear example of this bounding of output; light at 186,000 miles per second.

Finally, it is necessary to bound the range of effects outside model structure; under what conditions do external changes cause change in system behavior and thus model structure. This is addressing the cetera paribus problem even more directly and will be returned to in a few lines.

There is no need for these boundaries to be static. Not only could the boundary conditions within each class affect the boundary conditions of other variables within that system, but also within the

the other classes of boundary conditions. Therefore one could see the value in some circumstances of having the model include a section which has dynamic interaction of boundary conditions. The movement of a certain variable would redefine the boundaries of some other variable.

Delineating the cetera paribus boundaries is also a difficult task for which there is no facile solution nor one which can be derived from the model. What is clear is that it is necessary to state these conditions even more carefully than the other types of boundary conditions since those assert when one knows the model will fail while these are made to find out when things outside the model and, therefore, outside the direct observation of the system will make the model fail.

The first question to ask is what are the mediating forces upon which the functional relationships of the model depend. It is clear that the chain of stimuli and reaction is not without implicit assumptions about organizational context, systems of belief, and psychological perceptions. When it is said that wages are a function of supply and demand of labor one implicitly includes an understanding of certain functional relationships such as the existence of a market of accomplishing these tasks. There exist mediating mechanisms through which the process occurs. Usually in considering the validity of a model one is concerned with only the highest mediating functional relationship that must exist; the underlying functional relationships need not be described because as long as the highest on the hierarchy

does describe things the model will maintain a good description of behavior. This serves to terminate a needless drive to reductionist explanations. It does pose some interesting philosophical problems but perhaps the best way to resolve them is to make the functional relationships as distinct and intuitive as possible. It may become impossible to do this without proceeding to lower steps of explanation. However since the purpose of these functional relationships are to suggest invalidation the problem becomes slightly less burdensome. When do the tests of validation become invalid is certainly an issue and will not be resolved here.

Given the need to develop functional guides to the assertions upon which the model is built there are several possible steps to do this. The first is a verbal statement of the general functional relationships. Second is, under certain conditions, the possibility of devising mathematical formulations of these functional relationships, such that some indication of exactness is possible. From both of these types of statements it is possible to develop a series of indicators to delimit the boundaries. In the same way a series of indicators where the allowable deviations might serve to indicate when some functional relationship was failing and likely to be making model structure incorrect.

Because of the visual cognition of man the making of cognitive assertion maps might also prove to be fruitful to show the

interrelationship of the various functional relationships that must be maintained for the model to retain validity. Furthermore, it is possible to associate such maps with flow charts of the model such that each major loop would have specified all the major assertions affecting it. The mediating forces would be defined, and some test of when the relationships were likely to break down would be included. Perhaps checking cognitive assertions is the most important form of validation, and cognitive maps maybe is the easiest process in accomplishing this.

Finally, the inclusion of external forces might take place by the inclusion of dummy variables which represent the counterpart to random variables. These 'dummy variables' would be added to relationships that were breaking down but would be non random influence. Such a technique is not likely to increase model confidence, but if the results are still within a usable range, then the model still could be valid in the short run.

Therefore the inclusion of boundary conditions does away with the most insidious of modeling statements; be careful with what the model describes. For the inclusion of boundary conditions at least describes the meaning of such a statement and even intuitive statements can lead to discussions and guidance. The edict to be careful is the same as the edict to be right; a nice thought that is nirvana. The definition of boundary conditions serves to improve the ability of the model user to execute validation tests. By formulating the policy to be

tested and then the implementation strategy associated it should become possible to see if any of the consequences of the change are outside the range of the model as defined by the boundary conditions. Of course there are no guarantees that this will be accomplished due to the unforeseen effects of every policy; however, it does guarantee that the possible effects which are foreseeable are accounted for and used to improve decisionmaking.

Implied in this process is careful consideration of the results of various policies. This is an important but not simple task. The complexity of the task introduces some very intuitive notions into the process and requires the use of a priori models of behavior. It should be clear that this does not make it unnecessary to accomplish the task. Considering a model of birth control it would make quite a difference to the structure of a model whether a policy of birth control was accomplished through the use of contraceptives or selective death. One could envision from an a priori model some large differences in the nature of behavior and for a model of the behavior of people under some large changes. Needless to say the differences in most policies in the way they achieve a functional manipulation of the variables is not as extreme as this but in many cases very important nevertheless.

Part of the delineation of cetera paribus concerns the second order Hawthorne effects discussed earlier. The concern with the

goal structure of the simulation's actors is of paramount importance. Very few simulations of policy, at least at the larger level, describe situations where all the actors have the same goals. Most problems are in fact, a consequence of a situation where a system of multiple payoffs does not exist and where at least some of the actors see the world as a system of zero sums.<sup>1</sup> In such a situation the implementation of a known policy will lead to different expectations about not only the future of the world but the dynamic state of the world in arriving at that future. For the simulation to still hold the situation must exist where no superior action exists for the actors whose goals are not being optimized. If it does the actor will use the simulation to find that policy, even if in previous situations he would have reacted to the same stimuli in the way the model predicts to arrive at its new state. People actions are based on their goals and their models of how the world works. Thus one can see that unless they have no better alternatives or they do not believe the results of the simulation their behavior will change. Even in the case about not believing the simulation when their actions do not for them achieve the goals they thought it would they will begin to learn and change their behavior appropriately.

One method of overcoming this problem is to use the models to manipulate policy outcomes gnostically. While this technique is not compatible with democratic political structure it is theoretically possible. Then the only effect to deal with is the learning effect.

If the simulator is able to have a gnostic model of this effect it too can be accommodated. Given the fact that gnostic use of the model is unlikely in open societies (which even the Soviet Union is, by the technical definition of knowing what policies are supposed to be doing) the user of simulations must seek different solutions.

One possibility is to use the model to analyze what must be done to achieve the equivalent of gnostic use and maintenance of boundary conditions. By building implementation submodels where the purpose of the submodel is to tell what behavior is necessary to achieve functional manipulation of policy when Hawthorne effects are known to exist the problems can be overcome. These submodels need not be simulation but can be learning systems if that is more desirable.<sup>2</sup> The problem of validating submodels is less difficult because it will be possible to incorporate changing conditions into them. In this way the additional efforts every policymaker would have to take could be calculated, although certain effects of the model and implementation problems are going to be impossible to overcome.

Going back to the birth control illustration it might be possible to consider some of the impacts of the selective death policy on the structure of the behavior and to build some submodels whose purpose it would be to tell the decisionmaker what steps might be possible to ameliorate the deteriorating situation. The problems of what effect those steps might cause can also be considered but since

the steps may not be functional steps in a model it might be unnecessary or those steps might be deemed to have a negligible impact.

Without such a process it will be necessary to just consider the total impact of the second order Hawthorne effects and to do a sensitivity analysis, both within the structure of the model, and changing the structure of the model if the actors have a capacity to introduce new factors, to see the total impact of such effects. From this it should be possible to generate a set of boundary conditions in this area too.

The use of controlled Turing tests may also give some additional degree of confidence. Can the decisionmaker tell the difference between the output of the real world, the model as it is to be validated, and some dummy or modified model which is said a priori not to be correct? Furthermore, can the validation tests in general be subjected to different models and still distinguish the valid ones? This in some sense is a validation of validation.

Finally, the question of defining how valid is valid enough becomes all pervasive. Given the importance of this criteria it should be helpful to define in a more formal way what is meant by the term good.

To determine the value of validation it is necessary to return to the use to which validation is intended to be applied; to make the simulation more useful. The problem then becomes one of

determining whether the benefits of validation outweigh the costs involved in it. To determine this it is necessary to determine the expected benefits of any policy. Then the probability of the model being correct in assessing the outcome of the policy - the degree of confidence. When the costs of increasing this degree of confidence exceed the expected value of the expected benefits times the change in the degree of confidence then too much validation is taking place. When the points are equal validation steps should stop. Before that Bayesian analysis allows the analyst to determine when his guesses are likely to improve the expected gain over cost by giving him some guide to the determination of his probability estimates and their value. Of course the whole process is an evaluative one; how much the policy is worth, how confident one is before a validation step and after and so forth. Nevertheless, it may give some guide in cases where extensive costs will be incurred in each validation steps and is almost costless. None of this is meant to gloss over the very serious practical problems involved in such efforts - determining these points and values is difficult and not very reliable. Still some guide to the whole mess may be worked out.

Maintaining validity for models which require a long-run change and ongoing policy is also a difficult task. It has been argued that boundary conditions are good methods of determining validation. For maintaining it they may be essential. Unless some assurance is able to be given about the validation of a model through time the policymaker

may not have any way to know whether a model is correct, particularly if the policy being implemented is not an intuitive policy. Therefore, it is quite necessary to develop a series of tests which will examine this problem. For the most part these can be the traditional tests and the boundary conditions mentioned in this thesis.

The problem of how to determine boundary conditions and to develop cognitive maps is not easily soluble. A practical guide is to examine when the normal conditions did not exist and see whether the model held during that period. If the model did not hold it is worth examining what conditions caused its breakdown. For instance in examining a model of migration it might be interesting to examine a non normal situation such as World War II. By examining conditions then it may be possible to ask questions that are impossible to answer presently or from other time periods. Further if some of the explanations of a simulation break down one might begin to look for the cognitive assertions that did not hold for that period that occurred in war, or it might be possible to define some boundary conditions if the inputs were rather extreme, or unusual.

Beyond this it is possible to examine the theoretical basis on which the model rests. Does the model have a firm basis on which to rest? Do any of the theories on which the model rest define special conditions which other parts of the theoretical basis of the model

contradict? These can provide some help.

Examining the assumptions to see if they are linear and additive is helpful. Do the assumptions remain true under a variety of conditions and are they essentially remaining in the same relative position to each other? Do outside influences have about the same effect and wash out of the relationships? An example would be the impact of some exogenous force changing the behavior of different groups differently or more importantly changing the theoretical basis for some behavior such as profit maximization and markets. If the distortion is not uniform then more likely than not the model will yield some big problems and be invalid.

It is also helpful to try and divide the assumptions into two or more assumptions in order to see if some interaction effects (external entry of some factor) is likely in light of changing conditions.

Not much more can be suggested by the author at the present time. The point is to emphasize the direction the work must take. By now it should be clear that there exists no easy solutions to the problems presented earlier. The necessity is to develop more if the models and simulations are to be usable.

FOOTNOTES

1. Long, Norton, The Policy, p. 230.
2. Schon, Don, Beyond the Stable State.

## CHAPTER 8

Validation is not without its costs! Since in this paper the only purpose of validation is to improve the usefulness of the simulation the problem of cost enters in the process of deciding what validation procedures should be used. It is necessary to consider the tradeoff between validation costs and the improvement in policy which will stem from the increased efforts.<sup>1</sup> The reader should, by this time, realize that the costs incurred in validation are by no means small especially when the need to validate every policy change effect on the model structure is accepted.

The costs of the validation tests are high in several respects. The computer time used up represents a direct operational cost. The more validation done the greater the costs will be, particularly if the case where the validation tests depend on extensive number of runs. The second cost is the amount of analysis which must go into the examination of the validation tests. The time consumed by such work represents not only a direct economic cost but also a draining of energy from the researcher. This cost can prove to be more important than the direct economic cost.

Therefore it is necessary to weigh the value of the additional validation efforts to the potential gains. It will be necessary to weigh the expected values of the differential costs of decisions. This

is complicated by the fact that until decisions are made for each potential simulation it is impossible to do this and until models are validated the differential decisions will be impossible to determine. Therefore it will be necessary to develop good estimates of potential differences and then apply Bayesian probability theory to figure out expected values.

The benefits to be derived from validation are not necessarily small.

However certain problems enter into constructing and executing validation tests which threaten the whole value of doing them. Usually simulations are difficult to do and take a great deal of time. Researchers involved with them aim to make them represent reality. In many senses the whole validation procedure is in the building of such models. After the time and effort put into them the researcher has a great deal of confidence that they are correct. He has become functionally fixated on the model.<sup>2</sup> Instead of being able to examine the model clearly and objectively the researcher sees it as 'his baby.' The construction of validation tests makes it necessary to specify conditions under which the model won't work; the researcher may be reluctant, or unable to do this or more dangerously may define too wide a boundary. Overcoming this problem is not trivial since the assumptions of a simulation are rarely intuitively obvious and the

researcher is in the best position to really examine these assumptions critically were he able to do so. Furthermore, there will often exist a different interpretation of the way reality works. The mechanisms which produce the behavior under study will be questioned and the researcher will find it difficult to alter his model particularly since he feels he knows the most. If he has not stated the assumptions under which the model must be valid and invalid this questioning process becomes more difficult.

This same attitude creates a climate in which the potential users of simulation models are not likely to believe the modelers and their claims. An attitude of dogmatism is created which cannot be functional to the incorporation of criticism and problems. For the decisionmaker this attitude usually means that the model is unusable.

These issues may not seem to have a proper place here but it is probably one of the core problems in using simulations. The first criteria for the use of models is that they be potentially useful. The second is that if so they be convincing. Without the proper psychological state this is not likely to be the case. Users demand more attention to validation which they see in the convincing mode than research people. It will be necessary to make this jump for the simulations to be of use.

FOOTNOTES

1. Van Horn, Richard, "Validation," in Naylor, Design of Computer Experiments, p. 264.

CHAPTER 9

SUMMARY OF PART I

The purpose of the first section has been to analyze the criteria for the use of simulations which are policy oriented. The argument that has been advanced is that any policy oriented simulation which improves the decisionmaking capacities of policymakers is probably useful. In order to achieve this goal it is necessary to validate within the goals of the decisionmaker the simulation. Thus validation is below usefulness and is defined in pragmatic terms.

Following this line of argument it has been advanced that the present techniques of validation are inadequate. Particularly one aspect of this inadequacy has been investigated; the inability of simulations to incorporate external factors. The present method of treating experiments and policies as independent of model structure and as never introducing new factors has been questioned on methodological and practical grounds and found to be failing. The need to validate a model with every policy test has been demonstrated and some preliminary explorations for methods of doing this have been stated. In no way does this thesis answer all of the difficult questions it asks.

Finally some of the problems of achieving validation have been examined, particularly from a psychological viewpoint. From here, however, the thesis will move in Part II to the examination of a

simulation and application of some of the ideas and methods suggested heretofore. Understanding the ideas in all their obscurity is difficult without a real example.

PART II  
CHAPTER 10

In this Part the Urban Dynamics model will be examined in detail and some of the ideas of the previous Part will be applied to examining the model. In order to do this it will be necessary to include a short description of the model; those who wish more detail should consult the book.<sup>1</sup>

Urban Dynamics has three sectors: housing, industry, and labor. Housing and labor are both divided into three subsections: managerial, labor, and underemployed. Industry is also divided into three sections: new, mature, and declining. The model defines a finite area which is developing into a major urban center. New housing and new industry are built attracting people. This continues until the land begins to fill up. Eventually the land fills up and the city stops growing. Then a small decline takes place and the city is at equilibrium. During this growth upper class (premium) housing is declining into labor class housing and labor into underemployed and some underemployed class housing is being destroyed. The model has each class of housing more densely settled (subdividing apartments). Industry also declines and employs less people. Essentially then as the city grows it attracts more people. However when the constraint of land stops growth the balance between housing and industry is not at equilibrium. Both decline leaving too few jobs for the too many people in the area. Thus the simulation

presents a theory on how decayed areas evolve.

Then the model examines some policies aimed at reviving the city. All of these policies including manpower training and building low income housing increase the attractiveness of the city to outside migrants, particularly the poor. As a consequence they attract more poor to the area until again an unfavorable balance is created. Even new enterprise construction attracts migrants because more poor can always crowd into the 'excess housing' which is being generated by the decline of older housing. Forrester then goes on to suggest that the key to revival of the city is to destroy slums as one builds new industry. Particularly he suggests destruction of 5% of the slum housing and an increase of 40% in new enterprise construction. This results in a healthier balance ecologically and prevents immigration of the poor because they have no place to move to. As one of Forrester's co-workers stated the city becomes an economic machine.

One gathers from the book on Urban Dynamics that the purpose of the simulation is to examine methods of reviving the city. In order to do this the process by which the city developed these problems is modeled. This serves to allow one to examine numerous alternatives of revival.

Much criticism of the model has been generated, much of it is hysterical in its denunciation of the model, and much shows a penetration and understanding of urban processes which the model glosses

over or does not understand or include. The purpose of the discussion of Urban Dynamics in this thesis is not to comment on such criticism except as it is pertinent to the discussion of the application of validation procedures discussed in this paper. Some of the criticism has been about the lack of empirical verification of the model's structure, parameters, and coefficients.<sup>2</sup> For example, the treatment of migration has been attacked on the grounds it makes housing too important.<sup>3</sup> Others have attacked the model for not including social factors in the model. And still others argue that no model that does not include the whole metropolitan region can be worth anything.

Needless to say many of these criticisms have validity. In the following analysis it will be assumed that these problems have been overcome and that the traditional validation steps including goodness of fit have been successfully applied for a particular city. The present model is a good description of the way in which this mythical city has developed. It reproduces the time path and has the same coefficients and parameters as the model represents it as having. Therefore, the criticism of Forrester and his method of construction, which has much validity, is washed away for the purposes of this discussion.

The problem then is to describe the conditions under which the simulation will break down. What are the boundaries of the model's validity? To begin to answer this question it is necessary to ask what

the model is intended to do. If the model simply seeks to describe the growth of urban areas it may be perfectly alright (in this case it has been defined as such). If, however, the simulation is to be used to aid policymakers decide what policies will revive the urban area then greater definitions of boundaries are necessary.

For instance behind all the statements of the effects of various multipliers are assertions about the way in which they work. The migration multipliers for the underemployed sector, for instance, specify a certain set of conditions and assumptions about the way in which underemployed people behave, which under various policy alternatives may prove to be false. In the simulation underemployed migration is influenced by five multipliers: underemployed arrivals, underemployed housing, public expenditure, underemployed jobs, and underemployed housing program.<sup>4</sup> It has already been stated that these tables describe the world as it has been. However, in considering the future, particularly policies in the future aimed at reviving the area, it will be necessary to consider the assumptions on which the model is true and what boundaries must be preserved for the model to continue to be an adequate description of the world.

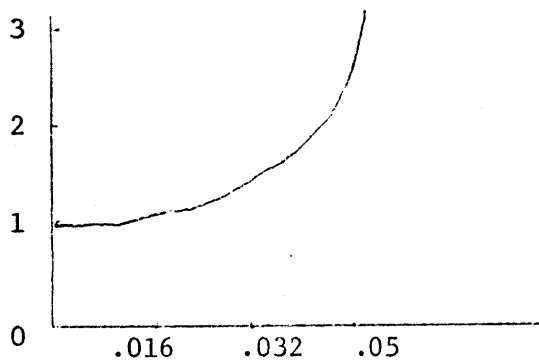
To begin the first assumption that this description of the migration of the underemployed makes is that all other forces which might effect the migration of people into the area are either included in this

equation of five multipliers, or have zero effect. Of course this zero effect could be attributed to either non-importance or to equivalent consequences everywhere in the world and a wash out of effect. Nevertheless, the structure claims that no forces can enter the migration equation except through one of the stated multipliers. Thus it is argued that the state of social conditions including schools, community, drug addiction, criminal activity and so forth has no direct effect on migration. People's decision to migrate will not, this assumption argues, be influenced by the danger of an area. Or the wash out argument would be that all areas are equally dangerous and, therefore, that fact would not influence the decision. In the same way other forces such as the degree of pollution are said to not influence the decision to migrate. Nor, for instance, is the availability of transportation allowed to influence migration.

The argument that these forces enter into the migration equation indirectly can be examined by looking at the loops in the model. Nowhere do they appear. In considering the policy changes that might be considered to revive the city it is necessary to consider whether the programs would change the validity of these assumptions. Would the policy, for instance, increase crime and in doing so make it enter into people's decisions about migration? Even though it has been agreed that in the past there was no influence of crime DIRECTLY on the model

and places development that does not mean that an increase in crime could not change this proposition. That is an issue which is central to the validation issue; can it be expected that crime will have influence? It will be necessary to be aware that the assumption of no direct influence exists to see if the policies are likely to violate that assumption.

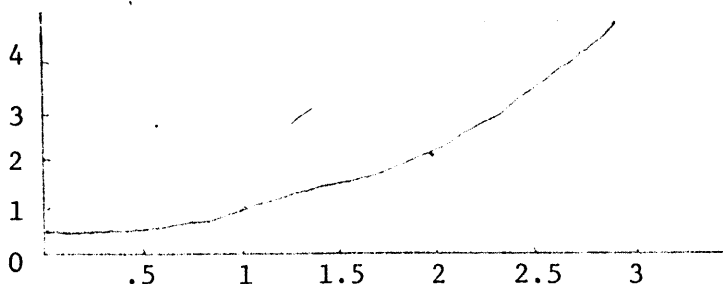
So far only a small part of the difficulties has been considered. One cannot assume that people's values will remain the same (or to more exactness the manifestation of values). For instance, the Public Expenditure Multiplier describes a certain kind of behavior or values about taxes raised and spent by the urban area. The curve looks as follows:



Essentially the shape of the curve can be said to mean that such expenditures help the poor, probably through welfare. However, the fact that such expenditures can help the poor and in the city defined earlier have done so does not mean that this is a logical necessity of the government spending policies. Mike Royko in Boss, a book on Daley's Chicago,

argues that in Chicago expenditures have risen as have taxes to the advantage of the rich and the detriment of the poor.<sup>5</sup> One could easily visualize a set of expenditures such as museums, elite schools, office services, and other benefits which would be entirely aimed at the middle class. Some expenditures such as more police for the purpose of repression could make life worse for the poor. Therefore, a boundary condition exists that the curve and expenditures must primarily benefit the poor and also be interpreted that way by them. Implicit in the table function on Public Expenditures are sets of behavior, values and perceptions which can be wrong and, therefore, tests and possibilities of invalidation must exist to protect against the tables' misapplication. Only under certain conditions will the table hold (conditions which existed in the past, yes, but which may not exist in the future particularly under the revival policies).

Another good example of the types of implicit assumptions made about migration in each of the multipliers is in the Public Housing Multiplier, which looks as follows:



This multiplier assumes that underemployed see public housing as a good attractive thing. This will depend on many things, including the quality of the housing, its reputation, and the reputation of all public housing nationally. If people believe that public housing is bad or that the new public housing will be bad then it might act as a negative force to attractiveness. For instance, in the city described so far all public housing in the past could have been very good, dispersed, low level housing which as the model argues was a factor that increased attractiveness according to the table. The assumption then holds for low level housing (that is part of the assumption). However, a new policy might necessarily have to include high rise housing which has negative attractiveness to underemployed because they know its dangers, the danger it poses to neighborhoods it is in, its disruptive social effect and so on. Thus some assumptions about the type of public housing are necessary, which any new policy is capable of violating. Beyond this an assumption of how much stigma is associated with living in such housing is necessary so that if such a stigma is changed the table can be changed. For instance, the present table could describe a situation where a lot of stigma is associated with such housing. However, a new program of let's say 25% new housing might remove that stigma, particularly if it occurs in certain geometrical and locational patterns. In that case the attractiveness of the area might be increased even more than the table function would show.

In this way it should be clear that the same types of assumptions which surround the P.E.M. also exist in this case. The table function describes only one case where many assumptions are necessary for that case to be correct, and which could be in dissonance with the policies to be tested.

In the same way there exists implicit variables which affect the ultimate value and importance of the ones described which are of a general nature. As the country gets wealthier there may be a tendency to favor the suburbs and a more rural kind of life. Any policy which raised the wealth of an area might make the urban area less attractive than the surrounding areas, particularly if the costs of not living in the area is not exceptional. Therefore the whole relative attractiveness, the concept of normal attractiveness is dependent on a set of values and facts which may not exist in the future or which could be affected by the policies to revive the city. Improving mass transportation may foster the movement out of the city by making it cheaper and practical.

Many mechanisms were discussed in Part I about methods to handle the types of problems which are being discussed here. Specifically each structural assertion, in this case each multiplier, must be examined to see if it needs to be bounded, and whether functional assumptions about the model must be assumed. Doing this can be accomplished through a variety of mechanisms previously mentioned. In addition every policy which is to be tested must be examined to see if it is consistent with

all the assumptions of the model. All the assumptions must also be consistent and if some changes are necessary in the model these must be done carefully so that other sections of the model are not invalidated.

It is important to realize that the problems which have been discussed here in regard to underemployed migration also exist for other areas of the model and similar types of analysis must also take place for those areas. For example, the sections on jobs must be examined to see what the assumptions of those areas are. A simple example would be the labor job ratio and its relationship to the types of industry existing in the area.<sup>6</sup> The boundary of the ratio might depend on the amount of export industry in the area and the relationship of the ratio to wages. If wages are competitive then a low ratio might be impossible because the high wages would be uncompetitive with the rest of the world. If the policy being tested depended on the wages to attract new people this would not happen. Therefore certain types of industry and dynamic behavior would be eliminated from the model's valid region because the assumptions of the model would not be wide enough to consider such possibilities.

What has been attempted here then is to show the types of assumptions surrounding any specification of model structure in policy-making models. In order to investigate the consequences of policies it would be necessary to formulate the types of questions used here to examine the underemployed migration area and to be able to invalidate

the model when the conditions defined by these assumptions are not going to hold. It will not be possible to explore all the areas of the Urban Dynamics model in detail; such an undertaking would be a thesis in itself, which shows some of the drawbacks of validation. Rather in the next chapter some of the basic assumptions of the model will be examined using the same types of methodology and logic described in Part I and used here.

FOOTNOTES

1. Forrester, J. Urban Dynamics.
2. Cope, Resume on Migration.
3. Ibid., page 3.
4. Forrester, op. cit., pages 20-30.
5. Royko, Boss.
6. Forrester, op. cit.

CHAPTER 11

In this chapter the fundamental assertions of the Urban Dynamics model will be examined. Since it is helpful to know for what policies the model will be scrutinized a brief description of Forrester's revival policy will put forth first.

Essentially Forrester argues that the revival of the city will increase its attractiveness to underemployed unless a counterbalancing force is found. Therefore any policy should seek to revive the city and at the same time stem the inflow of migrants. As a consequence he recommends the demolition of 5% of the slum housing and the increase of new enterprises at the rate of 40%. He does not state that these programs will cost anything nor that they must have a particular form for implementation. "Discussion of such a program does not imply that the demolition of slum housing should result from active intervention by the city. It might better be accomplished by changes in the tax laws and land zoning."<sup>1</sup> Forrester is asserting that the program or policy to implement the functional policy is not crucial to the success of the program as long as it accomplishes the immediate functional aim.

Within the Urban Dynamics model there exist some assumptions, which are very important, however, to analyzing his policies and all policies, which it would do well to state explicitly. These assertions are necessary assertions of any policy which would be successful to the

urban area if his model were to be valid.

The first assumption is that there are no costs to implementation of policies. This assumption is made clear by two facts; one that there is no feedback loops to taxes from any program but rather they are implemented for free, and two there is no tradeoff in city services if the program does cost money and no new taxes are raised. Total costs to the city could remain the same if the city stopped some other services such as street cleaning or fire protection, but this would affect the attractiveness of the city to some group and as such is not modeled.

The implications of this no cost of implementation are that there are no financial costs to implementing policies. Slum demolition is free. In Running Mailor by Flaugherty there is evidence that slum demolition even on the small scale New York runs it costs millions. Even Forrester's zoning or tax breaks have costs - one in law suits and enforcement problems, the other in lost revenue. One cannot create policies out of the air and it is a very stringent assumption that implementation will cost nothing.

No longer though are financial costs the only kind recognized. There exist social costs to most implementations. These are costs which individuals bear unfairly (the draft is an example), or costs which society also bears because the policy creates a situation which worsens life for many people. (There is a cost to society of creating drug

addicts-theft, loss of protective power, and less direct aesthetic effects which may even be more important.) Thus the Urban Dynamics model argues that there are no social costs to any policy. Neither slum demolition and ensuing crowding, nor new enterprise construction and loss of quality of life, have any effect whatsoever nor influence any of the behaviors. (The two are really a collary; generally costs are assumed to have an effect on behavior.)

As has been pointed out there is one situation where the model could be said to include costs and that is where other municipal services are curtailed to include the costs. This would lead to a conclusion that such services affect none of the behaviors of the model. Essentially the argument then becomes one that government services have no impact on people's lives.

Of course these assumptions can be true under varying circumstances. Some programs could be federally funded and have no important social impact. In defining these problems and their assumptions it is necessary to attempt to specify both what the assumptions are and what magnitudes of contradiction could be tolerated. Merely showing that some assumption is violated is not enough to prove the simulation invalid but one must also examine the importance of the contradiction.

A second major assumption of the Urban Dynamics model is that all benefits of the programs go to people living within the urban

area being described. Essentially the model describes a situation where no transportation network exists to put surrounding environments within working distance of the area. All the new jobs created by a program are taken by people within the area and all people stay within the area or leave without taking their jobs with them. (That is they do not move to a neighboring area and commute to their jobs.) As a corollary to this assumption no building takes place around the urban area being described which influences the effectiveness of the programs nor the migration of people. A zero interaction takes place between the surrounding environment and the urban area. The only distinction the model makes is that there exists an environment and an urban area. No distinctions are made between the interactions an area has with its 'suburbs' and other environment. The model makes the assumption that there is no influence that the urban area exerts on the environment, which, in turn, influences the area.<sup>2</sup> This assumption, of course, then includes the near environment.

Finally within the policy of building new enterprises at a higher rate there exists the assumption that capital is free. Increasing new enterprise construction without changing the model means that the model must assume that there is no cost for capital and no competition for it. Forrester has stated in talks with the author that he believes this to be the case and that "the world is not a zero sum game." In a sense this assumption is related to the no cost assumption because here

Forrester is also arguing that there is no cost to implementing a policy.

There are some additional assumptions which the model needs implicitly to stay a good representation of the world. One of the most important concerns the second-order Hawthorne effects discussed earlier. The model requires that there is no feedback between real estate speculators, the model, and the future beliefs of the city's conditions. Clearly if real estate people see revival as a possibility then they may build excess housing, not because it is good for the city but because they can sell it. An assumption must be made that this will not happen or that the city will take steps to prevent it. (Here is a good example of the city's ability to use the model for guidance without incurring extra costs - they could issue no new building permits.)

Some further assumptions remain to be discussed. These deal with the area of technology change. One is that the density of the city be stable in terms of buildings. If 'real land area' can be expanded then the most basic assumptions of the model may be totally incorrect. Thus the model assumes both a stable technology with regard to housing and space technology along with a constant zoning in terms of buildable space. If the city can improve its land area by higher buildings or below ground structures then the land area may be expanding in such ways as to invalidate the present policy conclusions.

One can also see that the state of transportation and communications technology are also fixed, or fixed at least so they do not influence the city's job structure and locational advantages. If the costs of such transportation and communication go down then people and industry may be able to commute in ways which would make it difficult to assess the validity of the model. An example would be centers of management where executives would not live or work in them but instead work and live from their homes. Therefore, they could vacate their city dwellings creating excess housing. On a more concrete level such developments could make cities less attractive to industry raising the cost of attracting them.

In the same way various technological changes could influence the land needs of industry so that the city either would not provide adequate space (if the changes favored horizontal industry) or would need vertical space (if changes caused the basic flows to be informational and not product-oriented).

In the same way if the technological complexity of new enterprise becomes such that the upgrading of underemployed would be too expensive for industry to undertake then other sources would have to do it for the model to remain valid, and as has been shown these would have to be sources other than the city.

These sets of assumptions are only the most important of the assumptions necessary for the model to be of value in helping policymakers decide what should happen. Certainly as has been demonstrated there exist a series of more detailed and model specific assumptions also necessary. In a model the size and complexity of the Urban Dynamics model many assumptions will exist. One could ignore few of them in deciding policy for even where they only postpone the ultimate outcome that may be enough to make the decision less than the best one. Maintaining the validity of these assumptions may seem like an impossible task; to the extent that it is they must be integrated and brought to a higher level so that they can be checked for in each policy and that the system can be monitored in the future. The assumptions stated here will serve as a guide to examining some specific revival policies in the next chapter.

FOOTNOTES

1. Forrester, Urban Dynamics, page 85.
2. Ibid., p. 17.

## CHAPTER 12

In this chapter Forrester's major policy suggestions will be examined in light of the assumptions of the model discussed in Chapter 12. The policies will be examined to see if the possible programs to implement them functionally will violate the boundary conditions and assumptions set up so far.

The first policy to be examined will be slum demolition. Essentially the impact of this policy is twofold. First it eliminates excess housing and dampens the attractiveness of the area to the underemployed class. Secondly it empties land which can be used for new enterprise creating jobs and restoring the 'ecological' balance of the area.

Several methods might exist to implement this policy. Only a few of the more obvious ones will be examined. (Creating earthquakes, fires, etc., will be excluded.) One possibility is to pattern tax incentives for the destruction of such housing. By tearing the housing down new tax shelters could be created which would make such actions profitable to real estate people. Two assumptions of the model might be violated by such a policy. First there is the problem of lost revenues, which must be made up in new taxes if the city is not to cut services. The cost might be greater than direct action. The second side effect,

a sort of second order Hawthorne effect separate from the ones listed earlier about speculation, is the possibility that real estate people might speed the deterioration of their housing. If the profits that could be gained by such a process were greater than holding the property the rate of decay which the model predicts would become wrong. Therefore any such policy would necessarily be subject to testing on this issue. To do this more information than one would like to have to invent would be necessary. However the points of testing should be clear here; some boundary condition would have to be established.

Another method might be the buying and clearing of the housing by the city directly. This would engender direct costs for both the legal problems and the destruction. Besides that it could increase hostility to the government which might influence some behaviors.

Any slum demolition policy must also be examined to see its possible impacts on social costs. Such a policy would necessarily increase the crowding and density of non-clear areas. Such an increase in density could result in a number of increased social problems. For instance, the health of the areas could suffer. This would be a case of the underemployed bearing the costs of the program as a 'health tax.' Crime might increase which could be looked on as a cost to everyone. (And possibly affecting the attractiveness of the city not only to migrants but to industry which must protect its plant.)

Another possibility that crowding could incur is the process of tipping if different ethnics are thrown together or if blacks are moved into belligerent neighborhoods. Such tipping takes place when small changes in percentage take place.<sup>1</sup> As a result there could exist more housing open to the underemployed if the labor elements are forced by their values to move to either new housing they would have constructed in the city or to the suburbs. Demolishing 5% of the housing of the underemployed could cause more than an equivalent amount of the labor housing, of a better stock, and therefore more attractive, be opened to the underemployed.

These issues are of tremendous importance. Here it has only been pointed out that the assumptions of the model might be violated. To know it would be necessary to have a specific set of facts from which boundaries could be shaped, and policy examined in detail to see if the conclusions of the model will still hold. It is clear that not only are the levels of homes destroyed important but their spatial characteristics, the way in which it is done and steps taken to alleviate these problems. What should be especially clear is that doing the policy as Forrester claims in his book is a ridiculous idea of how the process takes place and that functional manipulation of the variable may hide the demise of the model under any real policy. In order to achieve a validation of any slum demolition plan it would be necessary to have a very good idea

of the assumptions of each model and the implications of each policy, and as much as possible in a testable way. In order for the model to provide a useful guide to action it will be necessary to have the model and the policies consonant in their assumptions.

The second part of Forrester's revival policy is new enterprise construction being increased 40%. The present assumption of the model is that this will not cost the city anything. Several methods might exist for creating new enterprise. One is tax breaks. By giving firms an opportunity to expand more cheaply than normal a certain amount of additional new enterprise can be generated. However such tax breaks do violate an assumption of the model; that there be no cost to the city. First of all there is the direct revenue loss, which necessarily will extend to all new enterprise not just the additional amount, therefore costing the city money for enterprise it would have gotten for free. Besides this new enterprise consumes services which cost a great deal of money. Unless one assumes no additions to the infrastructure there will be additional costs. These must be absorbed by the city unless some outside agency will take them.

Rather than tax breaks the city can increase taxes and finance the attraction of capital out of them by boosterism, low cost loans and auxiliary services and direct investment. Here there are two losses which are related; consumption and wages. Firms which find they

must pay a competitive wage will not be able to offer the same wages as elsewhere but higher because of higher taxes. In that case there will exist no incentives for new enterprise construction and, in fact, because the bureaucracy involved is likely to be inefficient there will be a loss of competitive attractiveness. So unless the investment is financed from outside the city the new enterprise construction is likely to fail. In order to evaluate the program of course specifics are needed. Rather than argue that necessarily the new enterprise won't work the assumptions of the model should be tested on every policy and seen if that policy is lacking. Here that has been done and at the crude level the policy appears to be totally vacuous.

In the same way the costs of the capital must be examined in light of possible competition for it. Assume that a program is constructed where the costs of the capital are low enough as not to affect the validity of the model. Then the policy must be examined to see if it would lead to action in the environment, particularly in the 'local' environment which might raise the costs. For example some tax policy might work using the assumptions of the model. However if such a policy was to cause the suburbs to put a better tax policy into existence the effectiveness of the policy might be reduced so that the cost of attracting the industry was too high. Then a retaliation to raise the tax break effectiveness to the desired goal might cost an

amount that proved to be extraordinary, and would violate the cost assumptions of the model.

With the risk of too much repetition it is worth emphasizing the point that the issue here is that it is necessary to construct tests to see whether basic assumptions of the model are violated by the policies to be tested. This point is the essential one of this thesis.

In terms of the particular model it should also be clear that Urban Dynamics suffers from the non-consideration of these issues. In the author's opinion the U.D. model lacks the definition of issues so necessary to proper validation and policy testing. The program Forrester recommends is not likely to do what he says it will because he does not consider the issues which this paper brings up. Nor is it easy to do so.

In the beginning of this section the structure of the model as an explanation of the growth of the city up until the present period was accepted. This rather dubious step will now be examined.

Urban Dynamics is a complex model. In it Forrester has set up a series of new definitions, not many of which he supports empirically. For instance nowhere does he show that older, so-called mature industries actually employ less workers per acre than new industries. No where are the migration tables supported empirically, nor are his aging functions for housing with any empirical support. Rather Forrester has assumed

that his functions are correct or close enough to correct to ignore tedious estimation and empirical validation. He argues that the parameters and functions are all but in a few cases insensitive in explaining the growth of the city. To begin one wonders why there are in the model, since without them it would be easier to see what was happening. Secondly one must wonder how Forrester knows that they will not be important in the future, particularly if small structural changes in the model occur. Here is the core of Forrester's model; it assumes that it is a perfect structural model, at least for a "few hundred" years. Such a statement only need be examined in light of the changes in the last several decades to wonder whether that is true.

The validation of Forrester's model would require an effort much greater than constructing it. Just examining the specification of functions and parameters are major tasks. Up to this point in this Part II the concern has been focused on validation problems engendered in policy testing. Forrester's model should illustrate the more traditional problems and perhaps the magnitude of that task has been the cause of ignoring the no less important task discussed in this thesis.

FOOTNOTES

1. Duncan, Otis, in Hatt & Reiss, Cities and Society.

CHAPTER 13

The problems of maintaining validation may be used to the advantage of the policymaker, in designing programs for implementation. Every design of a program is faced with difficulties about what the program should contain. The very problems of including all the assumptions of the model in a policy may give guidance to important components of the program. The model can be used to detail the side effects every program is likely to generate and to separate the important side effects from those with no influence. Then steps can be taken to ameliorate the problems caused by such externalities. In the same way political problems can be included.

In the same way it would be possible to extend the use of the model to heuristic explorations. What will happen if they do that? Will the assumptions of the model be voided? By using the model as a tool of conditional prediction and to generate possible conditions facing decisionmakers it may be possible to foresee future changes likely in the nature of behavior.

Furthermore the problems which are introduced by this type of validation can lead to a stronger theory. It has been argued that operational structural models are impossible to get. However the problems of validating models can lead to greater structure and generality. In some sense the problems of validation can be used to choose

between alternative conceptions of reality. In social sciences there will often be more than one apparently good explanation of reality. By choosing the one more readily validated and maintained as validated there becomes a new type of Occum's razor for social science. From the difficulties of validation a few things can be learned about the nature of social and economic theory. Particularly important is the stress on dynamic theories which are able to include changing variables. The more a theory depends on idealized conditions the less likely it will be applicable.

PART III - SUMMARY

CHAPTER 14

The purpose of this thesis has been to evaluate the use of simulation models and to establish criteria for the use of policy oriented models. In order to do this it has been necessary to discuss when a model is valid for the purpose it was constructed. In this examination it was found that models are not adequate structural representations of reality that policies can automatically be tested in models without thought to maintaining validity. The argument has been put forth that 'experimentation' on simulation models with various policies can entail modifications of the simulation or model structure. Several possible methods of examining when models become invalid have been presented, along with a short examination of the Urban Dynamics model along this line of attack. The remaining problem is to decide whether these efforts have been successful, and in what ways they have failed, along with what questions are raised by all of this.

The problem of validation of simulation models is not one that preoccupies simulation textbook writers or researchers. An examination of the literature usually finds a quick mention of the great importance of the validation process but only a token discussion of the problems involved in it. The reasons for this are not insidious; simply very little work or thought has been done in the area that is worth noting.

The difficulties of constructing simulations are fairly large. Data usually is not available in the form that is needed. Usually large parts of the process under consideration are not really understood too well. Finally that damn world out there seems to be changing so much. This may be the core of the problem. Clearly if the arguments of this thesis are correct there is no reason to assume that the models built on recreating the past should have remained constant over time and that now suddenly they will change, bringing new forces into play. In the past this also was happening. However the models that are built usually attempt to explain all of the past in terms of the model. This is not bad since if that goal can be reached it makes it more likely that the model can be maintained more easily in the future. The problem stems that this makes building models difficult and validating them even more so. The brief discussion of the Forrester model should show some of the difficulties of the process. Given the basic difficulties of even classical validation it is not unexpected that the issues raised in this thesis are not ones that many people like to discuss; they are busy enough trying to get good historical validations.

It should also be clear from the thesis that the problems raised do not yield very easily to solutions. The author of this paper would not like to claim that he raised all the pertinent issues no less found any convincing answers to the questions raised. It would be fair to say that these problems will not yield to solutions very easily

at any time. They are complex problems and cannot simply be crunched through time consuming analysis. Given the vastness of the problems of present validation techniques and the difficulties of the problems raised here, with the tremendous increase in complexity that these problems would entail if they were dealt with, and it is not clear that they are worth pursuing at all. There may be much validity in attempting only the possible and skipping the areas of negotiable value and skill.

However if the problems brought up in this paper are as important as has been argued can they be ignored? The author believes not. Central to the use of models is their ability to describe future conditions under a variety of policy alternatives. This is more important than historical validation. Policymakers want to know whether the model will improve knowledge about tomorrow not whether the model fits better than other models of yesterday. It is the very instability of models that leads to difficulties in using them; until this problem is attacked there is little likelihood that the models and simulations will be of much use to anyone for the types of decisions that have been discussed in this thesis.

One gets the impression that people outside the simulation field understand this better than the researchers in the field. This is understandable. People involved in simulation research most often come out of engineering. As engineers they are used to dealing with

systems that are more closed and where the structure is better understood than in policy analysis. So they attempt to apply their knowledge and fail to consider a world of more fuzzy causation where external factors are almost always of some importance. Failure to consider the nature of their understanding leads to ignoring the problems of model instability under various policy decisions.

Where to go from here is a difficult question. In this thesis some questions have been posed without too many answers nor too much of an idea of how to answer them. Perhaps the answers may lie in reforming the questions in another way or seeking greater understanding of the present questions. Or perhaps the answers will never be concise but will only be found through an awareness of the problems on the part of researchers and strong examinations and debates about validity of each model, among the researchers themselves and the policymakers. To draw a conclusion would be premature.

BIBLIOGRAPHY

- Adelman, Frank and Irma, "A Simulation of Urban Politics," 1971 Winter Simulation Conference, ACM, N.Y., 1971.
- Amstutz, Arnold, A Management Oriented Behavioral Theory of Interactions Within Consumer Products Markets, PHD.1965, MIT.  
Computer Simulation of Competitive Market Response, MIT Press, Cambridge, 1967.
- Anderson, Hans, "On the Design of Goal Formation Procedures for a Sales Organization by the Use of Computer Simulation and Control Theory," in Models and Simulation by Stockhaus.
- Armstrong, R.A.R. and Hobson, Margaret, The Use of Game/Simulation in the Decisionmaking Process, DUN Interregional Seminar on the Use of Modern Management Techniques in Public Administration of Developing Countries, 1971.
- Barton, G. A Primer on Simulation and Games, Prentice Hall, Englewood 1970.
- Berger, Edward, and Barclay, Harvey, and Zisk, Betty, "Simulation and the City," Simulation and Games, Vol. I, No. 4, December 1970.
- Besenfelder, David, "The Capital Asset Pricing Simulator," 1971 Winter Simulation Conference, ACM, 1971, New York.
- Beshers, James, Substantive Issues in Models of Large-Scale Social Systems, MIT Press, Cambridge, 1965.
- Brewer, Garry, Evaluation and Innovation in Urban Research, Rand, 1971.
- Campbell, Don, Experimental and Quasi-Experimental Design for Research, Rand McNally, Chicago, 1966.  
"Reforms as Experiments," American Psychologist, Vol. 24, No. 4, April, 1969.
- Chorafas, Dimitris, Systems and Simulation, Academic Press, New York, 1965.
- Churchman, C. West, "An Analysis of the Concept of Simulation," in Hoggart, Austin, and Balterston, Fred, Symposium on Simulation Models: Methodoloty and Applications to Behavioral Sciences, South Western Publishing Company, Cincinnati, Ohio, 1965.

- Cope, Ross, Resume on Work on Migration, May 10, 1971, Urban Dynamics Memo D-1568, Sloan School, MIT.
- Cornell Land Use Game, by Feldt, Allen, Center for Environmental Studies, 1966.
- Crecine, John, Governmental Problem-Solving: A Computer Simulation of Municipal Budgeting, Rand McNally, Chicago, 1969.
- de Neufuille, Richard, and Stafford, Joseph, Engineering Systems Analysis, School of Engineering, Cambridge, Massachusetts, 1970.
- Dean, Stuart, "Simulation Model for a Complex Production/Distribution Network," 1971 Winter Simulation Conference, ACM, New York, 1971.
- Duncan, Otis, in Hatt and Reis, Cities and Society, Free Press, Glencoe, Illinois, 1957.
- Emshoff, James and Sisson, R., The Design and Use of Computer Simulation Models, MacMillan, New York, 1970.
- Emshoff, James, "Roles for Simulation in Behavioral Research," 1971 Winter Simulation Conference, ACM, New York, 1971.
- Evans, Wallace, and Sutherland, Simulation Using Digital Computers, Prentice Hall, 1971.
- Fishman, and Kiviat, Digital Computer Simulation: Statistical Considerations, Rand Corporation, RM-5287-PR, 1967.
- Flaugherty, Joe, Managing Mailer, Dell, New York, 1971.
- Forrester, Jay, Industrial Dynamics, MIT Press, Cambridge, Mass. 1963.
- "Industrial Dynamics - A Response to Ansoff and Slevin," Management Science, Vol. 14, No. 9, May, 1968.
- Principles of Systems, Wright-Allen Press, Cambridge, 1971.
- Urban Dynamics, MIT Press, Cambridge, 1969.
- Gordon, Geoffrey, Systems Simulation, Prentice Hall, 1969.

- Green, H. A. John, Aggregation in Economic Analysis, Princeton Press, Princeton, 1969.
- Greenberger, Martin, Computers and the World of the Future, MIT Press, Cambridge, Massachusetts, 1962.
- Feigenbaum, Edward (ed.), Computers and Thought, McGraw Hill, New York, 1963.
- Guetzkow, Harold, Simulation in Social Science, Prentice Hall, 1962.
- Halbegewachs, "Optimal Simulation of Nose Cone Billets of Material," 1971 Summer Computer Simulation Conference, New York, 1971.
- Hamilton, H.R., et al., Systems Simulation for Regional Analysis; An Application to River Basin Planning, MIT Press, Cambridge, 1969.
- Harris, Briton, "The Uses of Theory in the Simulation of Urban Phenomena," Journal of the American Institute of Planning, September 1966.
- Hester, James, Systems Models of Urban Growth and Development, Urban Systems Lab, Cambridge, Massachusetts 1969.
- Hoggart, Austin, and Balterston, Fred, (eds.) Symposium on Simulation Models, South-Western Publishing Company, Cincinnati, Ohio 1963.
- Kleindorfer, Paul: Lectures in 15.702.
- Krasnow, Howard, "Simulation Languages," 1971 Winter Simulation Conference, ACM, 1971, New York.
- Larsen, Uno and Lundin, Rolf, "Some Methodological Views on Digital Simulation," in Stockhaus, Haken, Models and Simulation.
- Lindblom, The Policy Making Process, Prentice-Hall, Englewood, New Jersey, 1968.
- Ling, Timothy, "Statistics and Dynamics of Simulation," in Naylor, Design of Computer Experiments, Duke, Durham, 1971.
- Long, Norton, The Policy, Rand McNally, Chicago, 1962.

- Meier, Robert, et al., Simulation in Business and Economics, Prentice-Hall, Englewood, New Jersey, 1969.
- Middleton, "Simulation of the Basic Oxygen Steelmaking Process," 1971 Summer Computer Simulation Conference, Boston, 1971.
- Milsum, J., "Vancouver Regional Health Planning Model," 1971 Winter Conference on Simulation, ACM, New York, 1971.
- Naylor, Thomas, The Design of Computer Experiments, Duke University Press, Durham, 1969.
- Computer Simulation Techniques, Wiley, New York, 1966.
- Naylor, T., and Hunter, "Experimental Designs for Computer Experiments," Management Science, Vol. 16, No. 7, March 1970.
- Naylor, T., and Finger, "Verification of Computer Simulation Models," Vol. 14, No. 2, October 1967.
- Orcutt, Guy, Microanalysis of Socioeconomic Systems, Harper and Row, 1961.
- Poole, Ithiel de Sola, et al., Candidates, Issues, and Strategies: A Computer Simulation of the 1960, 1964 Presidential Elections, MIT Press, Cambridge, Massachusetts, 1965.
- Raser, Simulation and Society, Allyn and Bacon, Inc. 1969.
- Rosenthal, Robert, Pygmalion in the Classroom, Holt, Reinhart, and Winston, New York, 1968.
- Royko, Mike, Boss: Richard and Daley, New American Library, New York, 1971.
- Russell, Bertrand, History of Western Philosophy.
- Schechter, Mordachasi, Simulation and Games, Vol. I, No. 1.
- Schon, Don, Beyond the Stable State, Random House, New York, 1971.
- Technology and Change, Dell, New York, 1967.
- Schlager, Ken, "How Managers Use Industrial Dynamics," Industrial Management Review, Vol. 6, No. 1, 1964.

Simon, Herb, Science of the Artificial, MIT Press, Cambridge, 1969.

Simon, Herb; Fisher, Franklin; Ando, Albert, Essays on the Structure of Social Science Models, MIT Press, Cambridge, 1963.

Skinner, B. F., Walden II, MacMillan, New York, 1966.

Smith, John, Computer Simulation Models, Hafner Press, London, 1968.

Stockhaus, Hakan (eds.), Models and Simulation, Scandinavian University Books, Gotesburg, 1970.

Taylor, John, Instructional Planning Systems, - A Gaming Approach to Urban Problems,

Toucher, K.D., The Art of Simulation, D Van Nostrand Co., Princeton, New Jersey, 1963.

Van Horn, Richard, "Validation," in Naylor, T., Design of Computer Experiments, Duke Press, Durham, 1969.

Wright, Richard, "On Validating Dynamic Models," Sloan School Paper.

"An Industrial Dynamics Implementation: Growth Strategies for a Trucking Firm," Sloan Management Review, Vol. 13, No. 1, p. 71, 1971.

1971 Winter Simulation Conference, ACM, New York, 1971.

1971 Summer Computer Simulation Conference, ACM, Boston, 1971.