THE UNAVOIDABLE A PRIORI*

Donella H. Meadows
System Dynamics Group
Dartmouth College
Hanover, N.H. 03755

ABSTRACT

This paper is a summary of the major assumptions underlying the field of computer modeling and the specific assumptions that differentiate four modeling methods used to represent social systems: system dynamics, econometrics, input-output analysis, and optimization.

The primary conclusions are:
1. Each modeling method is based on a set of techniques and priors that suit it well to some sorts of policy problems and poorly to others.
2. Misunderstandings between different kinds of modelers and between modelers and clients often arise from failures to recognize these implicit priors and the various strengths and weaknesses of the various modeling schools.
3. Some modeling schools, especially system dynamics and econometrics, are based on such different basic world views and assumptions about the nature of human knowledge that communication from one school to another is almost impossible.

---

*Title suggested by a section heading in Gunnar Myrdal's Asian Drama.

TABLE OF CONTENTS

> Questions are necessarily prior to answers, and no answers are conceivable that are not answers to questions. A "purely factual" study---observation of a segment of social reality with no preconceptions---is not possible; it could only lead to a chaotic accumulation of meaningless impressions. Even the savage has his selective preconceptions by which he can organize, interpret, and give meaning to his experiences.
> (Myrdal, 1968, p.24)

Although the field of computer modeling has existed for only a few decades, a number of different methodological schools based on distinct technqiues have already appeared. They include linear programming, input-output analysis, econometrics, stochastic simulation, and system dynamics. All these modeling schools share a number of common concepts about the properties of systems, the process of modeling, the use of the computer, and the role of models in decision making.

In addition to the shared concepts general to all mathematical modeling, each methodological school also employs its own set of theories, mathematical techniques, languages, and accepted procedures for constructing and testing models. Each modeling discipline depends on unique underlying and often unstated assumptions; that is, each modeling method is itself based on a model of how modeling should be done.

These deep, implicit, operating assumptions at the foundation of each modeling method are sufficently important that they should be re-examined more often than they actually are. Practitioners of each method learn its operating assumptions once

and thereafter may reflect on them only occasionally. Typically, the assumptions of each modeling school are part of the subconscious rather than conscious reasoning that goes into the making of models. Physicists rarely rethink the laws of algebra or the second law of thermodynamics as they work, practicing econometricians seldom stop to question the use of statistics to measure model validity, and system dynamicists regularly use the principle of feedback control without redefining it each time.

Such time-tested and rarely-examined preconceptions seem to fit the concept of a paradigm as defined by Thomas S. Kuhn:

> Scientists work from models acquired through education and through subsequent exposure to the literature often without quite knowing or needing to know what characteristics have given these models the status of community paradigms....Paradigms may be prior to, more binding, and more complete than any set of rules for research that could be unequivocally abstracted from them....(Paradigms) are the source of the methods, problem-field, and standards of solution accepted by any mature scientific community at any time....In learning a paradigm the scientist acquired theory, methods, and standards together, usually in an inextricable mixture. Therefore, when paradigms change, there are usually significant shifts in the criteria determining the legitimacy both of problems and of proposed solutions....Paradigm changes...cause scientists to see the world of their research engagement differently. In so far as their only recourse to that world is through what they see and do, we may want to say that after a (paradigm) revolution scientists are responding to a different world. (Kuhn 1970, pp.46-111)

Different modeling paradigms cause their practitioners to define different problems, follow different procedures, and use different criteria to evaluate the result. In a very real sense the paradigm biases the way the modeler sees the world and thus influences the content and shape of his models. As Abraham

Maslow says, "If the only tool you have is a hammer, you tend to treat everything as if it were a nail." (Maslow 1966) The selective blindness induced by any operational paradigm has both unfortunate and fortunate results. Unfortunately, it often leads to sterile arguments across paradigms, each school criticizing the problems, assumptions, and standards of the other, from the biased perspective of its own problems, assumptions, and standards. On the other hand, paradigm-directed research seems to be not only psychologically necessary but exceptionally fruitful. "Within those areas to which the paradigm directs the attention of the group, normal science leads to a detail of information and to a precision of the observation-theory match that could be achieved in no other way." (Kuhn 1970, pp.64-65)

Because of the inescapable effect of methodological paradigms on modelers' thoughts and perceptions, any comparison or evaluation of models must begin with an understanding of the underlying paradigms within which the models were made. Furthermore, in order for social-system modeling to produce a cumulative understanding of social processes and to contribute significantly to social policy making, the problems of selective blindness and of cross-paradigm communication must be dealt with. Computer modeling would be more effective, both as a science and as a useful art, if each modeler could recognize the assumptions behind his own modeling school and try continuously and respectfully to understand other schools.

To become actively aware of the deep and implicit operating assumptions that guide one's daily professional activities is a surprisingly difficult task. It is even more difficult to discover someone else's operating assumptions, since they are usually also implicit and not directly observable, and since they are often antithetical to one's own habitual way of viewing the world. The rarity of this kind of paradigmatic overview is evident from the misunderstandings and even occasional hostility that various kinds of modelers exhibit toward each other, and the distrust that other professionals sometimes exhibit toward modelers in general.

This paper is my attempt to expose the bedrock assumptions that underlie the entire field of social-system modeling and the more specific assumptions that define four modeling schools--system dynamics, econometrics, input-output analysis, and optimization.

My viewpoint here must necessarily be that of a one-time physical scientist turned system dynamicist, one who is relatively new to that field and who has theoretical knowledge of, but little practical experience with, the other modeling schools described. The biases associated with this viewpoint will undoubtedly be readily apparent to everyone but me. Anyone who undertakes this task will bring some set of biases to it. Mine will serve as well as anyone's to begin the discussion and to bring forth the clarifications and rebuttals necessary to produce a balanced view.

The Preconceptions of Modeling

Although modelers may disagree vehemently about their specific methods or models, they are unified by some very basic assumptions that define the whole modeling approach to problem-solving. First of all, social system modelers generally come from or were educated in a Western culture, one where attempts at rational, logical, scientific mode of thought predominate. Whatever happens is not believed to be random; it is assumed to have a cause that can be understood and probably altered. Careful measurement, clever experimentation, and logical deduction should reveal that cause.

Furthermore, modelers share a basically managerial world view. Problems should be actively solved, not passively endured, and problems can be solved if their causes can be understood. One does not ride along with the process of social evolution, one strives to direct that evolution. This managerial world view is generally acceptable to engineers, businessmen, some scientists, and some politicians, but not to most artists, theologians, or other humanists, or to those educated in traditional Eastern cultures:

> By one Chinese view of time, the future is behind you, where you cannot see it. The past is before you, below you, where you can examine it. Man's position in time is that of a person sitting beside a river, facing always downstream as he watches the water flow past....
>
> In America and other Western countries, the commonest view of abstract time seems to be the opposite of the old Chinese one. In this, man faces in the other direction, with his back to the past, which is sinking behind him, and his face is turned upward to the future, which is floating down upon him. Nor can this man be static: by our ambitious Western convention, he is supposed to be rising into the future under his own power, perhaps by his own direction. He is more like a man in a plane than a sitter by a river.
>
> (Peck, 1967, pp.7-8)

Although computer modelers use historic observations to form their hypotheses, their faces are primarily turned upstream, toward the future, and with a belief that the future can and should be shaped by decisions and actions based on scientific understanding.

The assumptions that distinguish all computer modelers from other managerial types center around the tools modelers choose to help them analyze problems; mathematics and the computer. A computer modeler assumes that the computer augments the human brain as a steam engine augments human muscle, and thus it is the obvious tool for dealing with matters that are too complex for the unaided single mind. Furthermore, he assumes that human actions and purposes can be categorized, quantified, and represented by mathematical equations. This postulate does not necessarily imply, as many non-modelers believe it does, a belief that human beings or the systems they create are totally predictable. It does require a belief that they are predictable in the aggregate and on the average, however. As E. F. Schumacher says:

> In principle, everything which is immune to the intrusion of human freedom like the movements of the stars, is predictable, and everything subject to this intrusion is unpredictable. Does that mean that all human actions are unpredictable? No, because most people, most of the time, make no use of their freedom and act purely mechanically. Experience shows that when we are dealing with large numbers of people, many aspects of their behavior are indeed predictable; for out of a large number, at any one time, only a tiny minority are using their power of freedom, and they often do not significantly affect the total outcome.
>
> (Schumacher, 1973, p.217)

Modelers believe not only that aggregate human actions can be quantified into computer equations and that computer equations can be grouped into representations or models of social systems, but also that these models are at least potentially better representations than any others that might be used as a basis for social decisions. Most computer modelers seem to agree with J.W. Forrester's postulate that individual and social decisions must be made on the basis of some model, most usually the "mental model", which is the set of unexpressed assumptions and generalizations about the world that exists in each person's mind. There is no Absolute Truth upon which to base one's actions, there are only more or less simplified models, derived from education, culture, and personal experience. Given that decisions must be based on some sort of uncertain model, a computer model may be preferable to a mental model because:

1. It is precise and rigorous instead of ambiguous and unquantified.
2. It is explicit and can be examined by critics for inconsistency or error.
3. It can contain much more information than any single mental model.
4. It can proceed from assumptions to conclusions in a logical, error-free manner.
5. It can easily be altered to represent different assumptions or alternate policies.

Very few computer modelers can claim that their models actually do exhibit all these advantages. Models can easily become so complex that they are impenetrable, unexaminable, and virtually unalterable. They can also be less complete than mental models, if requirements of mathematics or data inputs

prohibit the inclusion of certain kinds of relationships. However, the five advantages listed above are considered at least potential characteristics of computer models, and to the extent that they are realized, computer modelers believe they can provide unique and superior information for social decision making.

Characteristics of Different Modeling Methods

Upon the rock of these basic assumptions about rationality, the scientific method, the computer, and the advantages of mathematical models, a number of different modeling schools have been erected. Each was originally developed in response to a specific social need, each has developed its own methods and languages, and each shapes the procedures and perceptions of its adherents in a distinct way.

The different kinds of modeling are usually classified along a number of dimensions, some of which are partially overlapping and some of which are totally incommensurate. For example, models may be distinguished by their information bases:

social statistics
laboratory experiments
economic theory
ecological observations
etc.

by the mathematical procedures they employ:

random number generation
differential equations (analytical models)
difference equations (simulation models)
simultaneous equations
optimization procedures
statistical estimation procedures

or by the nature of the model relationships:

stochastic or deterministic
continuous or discrete
linear or nonlinear
lagged or simultaneous

These properties of models are combined in practice into a limited number of fairly consistent sets. For example, engineering models of physical systems tend to be based on information from controlled laboratory experiments, to be solved analytically, and to consist of linear, deterministic, continuous relationships. Econometric models are based on statistical data, typically contain both simultaneous and lagged relationships, are highly linear, and are solved by iterative simultaneous-equation techniques. Any complex dynamic model with nonlinear and lagged equations necessarily must be solved by simulation techniques (difference equations). Information base, disciplinary preconceptions, and mathematical necessity interact to form the philosophical view and the procedural rules that characterize each modeling school.

I could organize the following discussion around any one of the properties of models listed above, but instead I shall choose as the primary point of distinction another property that is not very often mentioned in model classifications. That is the use to which the model is to be put. I am concerned here only with models that contribute to the understanding and management of social systems. Therefore, I shall classify models according to the stage of social decision-making at which they are most applicable.

At the very first identification of a problem there may be a need for general understanding. The system producing the problem may never have been critically studied, or past studies may have been incomplete or faulty. Important data may be missing, interconnections that had been considered absent or unimportant may suddenly appear significant, or old theories may be called into question by new and unexpected behavior. The problem must be understood in its long-term historical perspective and in a wide enough boundary to include all of its causes and consequences. This is typically the point where learned study commissions are established or basic research projects are funded. Current models, mental or otherwise, may need revision, updating, or complete overhaul before the problem can be tackled.

Models that can contribute to an improved general understanding obviously must be easily understood. They should make clear exactly how their assumptions lead to their conclusions, and they should provide new insights about the working of some real-world system. Quantitative precision is unnecessary and probably unattainable at this point; it is difficult enough to decide what system elements are even qualitatively important and how they are related. Because the problem being addressed is new and may have sprung from an unsuspected source, the model must have very broad boundaries, usually crossing many disciplines. General-understanding modeling projects tend to be more process oriented than product oriented;

that is, the very process of making the model, asking questions systematically, and defining new concepts may itself improve understanding so much that by the time the computer model is finished, it is no longer needed. Its concepts and conclusions have been integrated into the mental models of both modelers and clients.

If general understanding allows some agreement about the cause of the problem or the nature of the system generating it, then the second phase, which I will call <u>policy formulation</u>, begins. Theories about the cause of the problem will lead directly to suggestions about the general directions in which a cure might be found. Several broad policy choices must be evaluated and integrated to identify possible tradeoffs or synergies. The policy questions to be answered by a model are still imprecise and generic at this stage, but the examination can be limited to those points in the system that have been identified as potential policy foci. Should family planning or health care be given more emphasis? What would happen if government control of domestic grain prices were released? A model that can help in policy formulation should be able to reproduce the real system's behavior under a variety of conditions, it should be easily altered to test a wide variety of possible policies, and it should clarify why different policies lead to different results. Quantitative precision is more important here than it was at the level of general understanding, but the emphasis is still primarily qualitative and process-oriented.

When a basic policy direction has been formulated, a whole host of new questions arise concerning the <u>detailed implementation</u> required to carry out that policy. A policy to promote family planning engenders numerous further decisions about budgets, personnel training, geographic distribution, and educational techniques. A policy to stabilize grain prices by creating a buffer stock will require the creation of new organizations to establish and maintain the stock, a precise set of rules for buying and selling, and a plan linking markets, warehouses, transportation systems, and final consumers so that the greatest stabilization can be realized at the least cost.

As these detailed implementation decisions become complex and require the organization and processing of many pieces of information, mathematical models may be very useful. Such implementation-stage models typically must be detailed and highly accurate, but each one need represent only one basic policy direction, so its boundary can be narrow. Detailed-implementation modeling schools are usually <u>product oriented</u>; they aim to produce a model that can be used again and again to transform new input data into specific predictions or operating instructions. Product-oriented modelers rarely need to involve the client in the modeling process or try to make clear all the model's assumptions. Probably most computer models now being made are directed to this stage of detailed decision making.

Different people sit at the various stages of the policy process, asking different sorts of questions requiring different kinds of models. Each of the methodological paradigms described below can be regarded as one useful tool in a tool box. Knowledge of all their properties is essential in deciding which is the best tool for a given specific purpose. It is possible to use each of these methods for several different purposes, and evey by stretching things a bit, for all purposes. A saw could be used to pound in a nail, if necessary. But a hammer would do the job better and faster, and so an essential aspect of wisdom in making, sponsoring, promoting, or criticizing models is knowing when each kind of model is most useful and when it is being pushed beyond its range of applicability.

Four modeling schools will be discussed below, beginning with those best suited to a general-understanding phase of decision-making and moving toward those used for detailed implementation. After a brief summary of the historical development of each field, the most important characteristics and assumptions of the method will be discussed. Examples of actual policy applications will be given. Finally, the most common problems and limitations of the method will be described. These problems are not necessarily present in all models; in fact, the best models are often recognized as good because they have managed to avoid them. Nevertheless, every method has its most common pitfalls, into which students often fall and against which advanced modelers must continually guard. Understanding these potential limitations as well as the strengths of the various modeling schools may be one of the most effective steps toward better modeling and better use of models.

System Dynamics: General Characteristics

System dynamics was developed at MIT during the 1950's, primarily by Jay W. Forrester. He brought together ideas from three fields that were then relatively new -- control engineering (the concepts of feedback and system self-regulation), cybernetics (the nature of information and its role in control systems), and organizational theory (the structure of human organizations and the forms of human decision-making). From these basic ideas, Forrester developed a guiding philosophy and a set of representational techniques for simulating complex, nonlinear, multiloop feedback systems. He originally applied these techniques to problems of industrial firms, and the first system dynamics models addressed such general management problems as inventory fluctuations, instability of labor force, and falling market share (see Forrester, 1961).

The methods worked out by Forrester and his group have since been applied to a wide variety of social systems (see, for example, Forrester, 1968; Hamilton et al., 1968; and Forrester, 1971). The field is still dominated by engineers, industrial managers, and physical scientists, with a world view that is basically problem-oriented. The literature of system dynamics contains many more descriptions of models addressed to policy questions than theoretical discussions about modeling techniques.

As its name implies, system dynamics is a method of dealing with questions about the dynamic tendencies of complex systems--what kinds of behavioral patterns they generate over time. System dynamicists are generally unconcerned with specific values of system variables in specific years. They are much more interested in general dynamic tendencies; whether the system as a whole is stable or unstable, oscillating, growing, declining, or in equilibrium. To explore the dynamic tendencies of systems, they will consult and include in their models concepts from any discipline or field of thought, with special emphasis, however, on the physical and biological sciences and some tendency to discount (or rediscover and rename) theories from the social sciences.

The primary assumption of the system dynamics paradigm is that the persistent dynamic tendencies of any complex system arise from its causal structure--from the pattern of physical constraints and social goals, rewards, and pressures that cause people to behave the way they do and to generate cumulatively the dominant dynamic tendencies of the total system. A system dynamicist is likely to look for explanations of recurring long-term social problems within this internal structure rather than in external disturbances, small maladjustments, or random events. For example, a system dynamicist is led by his paradigm to explain the energy problem in terms of reserve depletion, systematic underpricing, and rising material aspirations, rather than Arab oil embargoes or bad weather. He is likely to look for a solution to the problem through changing the goals and the information that influence people's decisions, not through one-time adjustments in taxes, research expenditures, environmental standards, or foreign policy. This basic assumption does not necessarily imply that all problems originate from faulty system

structure; just that system dynamicists are more likely to see and become interested in the ones that do.

The central concept that system dynamicists use to understand system structure is the idea of two-way causation or feedback. It is assumed that social or individual decisions are derived from information about the state of the system or environment surrounding the decision-maker. The decisions lead to actions that are intended to change the state of the system. New information about the changed state (or unchanged, if the action has been ineffective) then produces further decisions and changes. (See Figure 1.) Each such closed chain of causal relationships forms a feedback loop. System dynamics models are made up of many such loops linked together. They are basically closed-system representations; most of the variables occur in feedback relationships and are thus endogenous. Relatively few variables are determined exogenously (influence the system but are not influenced by it).

The element in each feedback loop that represents the environment surrounding the decision-maker is referred to as a state variable or level. Each level is an accumulation or stock of material or information. Typical levels are population, capital stock, inventories, and perceptions. The element representing the decision, action, or change (often, but not always, induced by human decision-makers) is called a rate. A rate is a flow of material or information to or from a level. Examples are birth rate, death rate, investment rate,

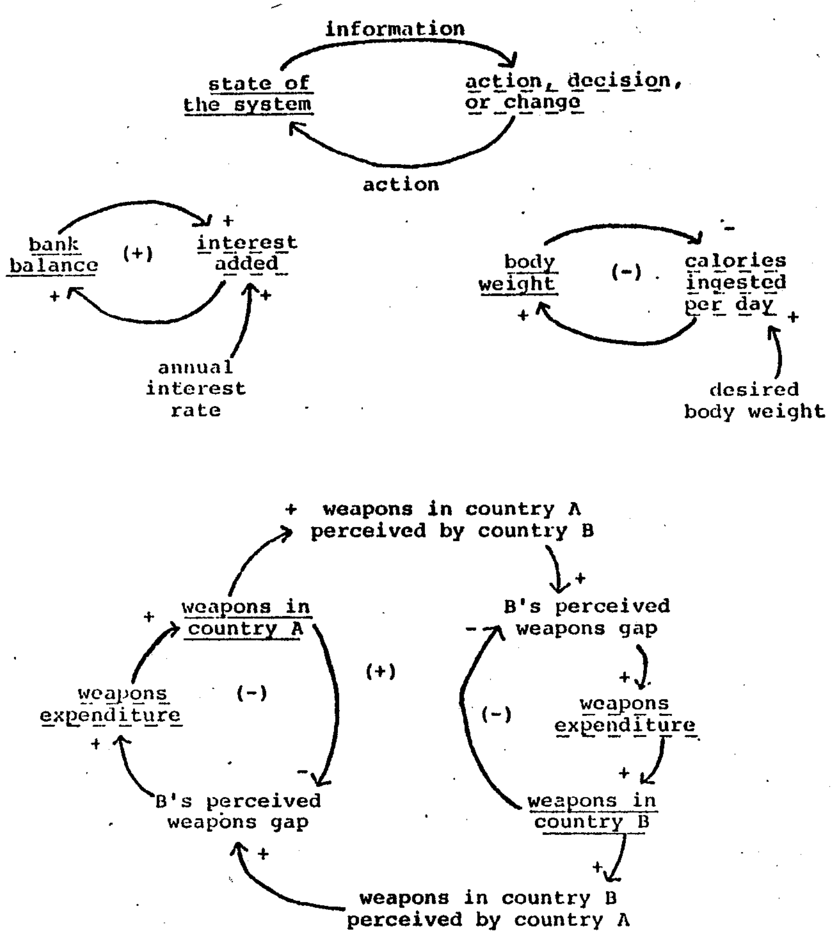
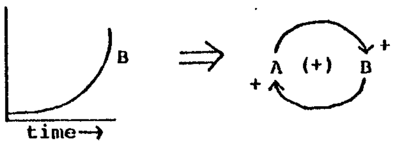


Figure I: Examples of feedback loops. Arrows indicate causal influence. Positive loops are designated by (+) and negative loops by (-). Levels are underlined with a solid line, rates with a dashed line. Elements not underlined are goals, perceptions, or other information affecting rates.

or rate of sales from inventory. Figure I illustrates several levels and rates and shows how they are causally linked into feedback loops.

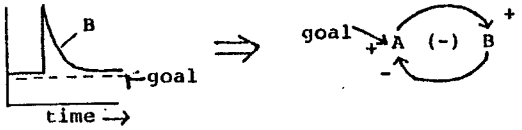
The concepts of feedback, levels, and rates require a careful distinction between stocks and flows of real physical quantities and of information. In the system dynamics paradigm physical flows are constrained to obey physical laws such as conservation of mass and energy. Information flows obey their own particular laws: information need not be conserved, it may be at more than one place at the same time, it cannot be acted upon at the same moment it is being generated,* it may be systematically biased, delayed, amplified, or attenuated.

Two kinds of feeback loops are distinguished. A positive loop tends to amplify any disturbance and to produce exponential growth. A negative loop tends to counteract any disturbance and to move the system toward an equilibrium point or goal. Certain combinations of these two kinds of loops recur frequently and allow system dynamicists to formulate a number of useful generalizations or theorems relating the structure of a system (the pattern of interlocking feedback loops) to the system's dynamic behavioral tendencies. For example, exponential growth indicates the presence of a dominant positive feedback loop.

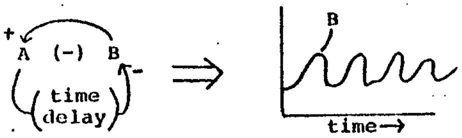


*From this assumption about the nature of information is derived the system dynamics use of difference equations and time delays, rather than differential equations or simultaneous equations.

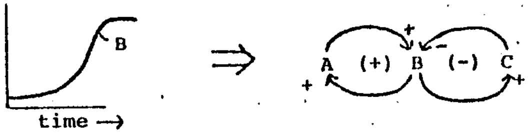
A tendency for a system to return to its original state after a disturbance indicates the presence of at least one effective negative feedback loop.



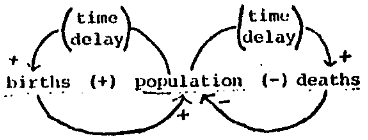
A single negative feedback loop with a time delay in it can produce oscillatory behavior.

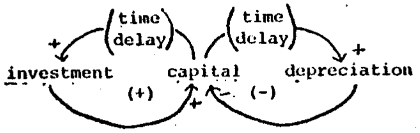


Sigmoid or S-shaped growth can only result from linked positive and negative loops that respond to each other nonlinearly and with no significant time delays.



These and other structure-behavior theorems are the main intuitive guides that help a system dynamicist interpret the observed dynamic behavior of a real-world system and detect structural insufficiencies in a model. They permit identification of isomorphisms in very different systems that can be expected to have similar behavioral patterns. For example, to a system dynamicist, a population with birth and death rates is structurally and behaviorally the same as an industrial capital system with investment and depreciation rates. They look like this:





and from their structure can be expected to grow exponentially, decline exponentially, or oscillate, but not to exhibit sigmoid growth (because of the time delays).

As these examples illustrate, time delays can be crucial determinants of the dynamic behavior of a system. System dynamics theory emphasizes the characteristics and consequences of different types of delays, both in information and in physical flaws. System dynamicists expect and look for lagged relationships in real systems.

Nonlinearities are also believed to be important in explaining system behavior. A nonlinear relationship causes the feedback loop of which it is a part to vary in strength, depending on the state of the system. Linked nonlinear feedback loops thus form patterns of shifting loop dominance--under some conditions one part of the system is very active, and under other conditions another set of relationships takes control and shifts the entire system behavior. A model composed of several feedback loops linked nonlinearly can produce a wide variety of complex behavior patterns.

Nonlinear, lagged feedback relationships are notoriously difficult to handle mathematically. Forrester and his associates developed a computer simulation language called DYNAMO that allows nonlinearities and time delays to be represented with great ease, even by persons with limited mathematical training. DYNAMO is a very specialized language developed to express the basic postulates of the system dynamics paradigm and to be easily understandable to laymen. It is widely used by system dynamicists because of its convenience, and, therefore, it is often thought to be an identifying characteristic of a system dynamics model. But any system dynamics model can be written in a general purpose language, such as FORTRAN, and, conversely, DYNAMO can be used to program linear open-system models that are not philosophically system dynamics models at all. In other words, DYNAMO is tool used by many system dynamicists but it is not exclusively a system dynamics tool.

System dynamics models are usually intended for use at the general-understanding stage of decision-making.* Therefore, they tend to be fairly small, aggregated, and simple. Most fall within the range of 20-200 endogenous variables. The individual model relationships are usually derived directly from mental models and thus are intuitive and easily understandable. The paradigm requires that every element and relationship in a model have a readily identifiable real-world counterpart; nothing should be added for mathematical convenience or historical fit. Great emphasis is placed on careful model documentation and on involving the client as much as possible in the modeling process.

Some questions that have been addressed with system dynamics models include:

> What has caused American cities to experience a 100-200 year life cycle of growth, followed by stagnation and decay? (Forrester 1969)
>
> How do primitive slash-and-burn agricultural societies control their populations and their land use practices to ensure a stable pattern of life in an ecologically fragile environment? (Schantzis and Behrens 1973)
>
> Why do agricultural and mineral commodities exhibit oscillating price and production trends, and why does each commodity exhibit a characteristic period of oscillation? (Meadows 1970)
>
> What has caused the decrease in the number of economically viable dairy farms in Vermont, and what policies might halt that decrease? (Budzik 1975)
>
> What policies will help the U.S. energy system make a smooth transition from a petroleum base to other energy sources? (Naill,et.al. 1975)

The first three of these studies fall in the category of general understanding; the fourth and fifth include both general understanding and policy formation. All of the studies have a time horizon of 30 years or more.

System Dynamics: Problems and Limitations

System dynamics modelers, particularly when using the DYNAMO compiler, must supply knowledge and judgment about interconnections in the real-world system, but not extraordinary mathematical or programming skill. The well-developed DYNAMO software package has many obvious advantages, but it also has

---

*However, system dynamics models have been made to address problems of detailed implementation, and several successful consulting groups regularly work with decision-makers at the implementation stage of problem-solving.

several disadvantages. First, it makes modeling look so easy that beginners who know the language but not the underlying philosophy of the method, are likely to become overconfident and to oversell their skills and their models. Second, because additions, alterations, and policies are readily added and analyzed within minutes, beginners and advanced modelers alike are tempted to play with endless model variations, rather than analyze carefully the experiments they have tried and the lessons they have learned. Finally, the mechanical simplicity of adding new elements and relationships to a model enhances the natural tendency of all modelers to create an overcomplex, opaque, uncontrollable structure.

The ease with which models can be overelaborated is common to many modeling schools, but is a special problem in system dynamics. Both the philosophy and the general-understanding purpose of the system dynamics method require simplicity and transparency. System dynamicists recognize the problem of overcomplex models and greatly emphasize, both in training and in publication, the necessity and difficulty of creating simple models. System dynamicists tend instinctively to criticize complex models and to admire simple ones. In fact, the pains that are taken to instill and reiterate the goal of model simplicity may reflect the very real difficulties in achieving it.

The emphasis on simplicity in system dynamics is consistent with the purposes for which this technique is usually intended, but it has also limited its range of application

primarily to questions that involve aggregate quantities. System dynamicists tend to avoid questions of distribution. Distribution of income, resources, opportunity, pollution or any other quantity is represented in almost any modeling method by the "brute force" method of disaggregation. Each class, person, or geographical area concerned is represented explicitly, and the flows of goods or bads among them is accounted for. Disaggregation into even a few classes or levels can complicate a model tremendously. A modeler striving for clarity and simplicity will try to avoid disaggregation as much as possible, and thus may be likely to discount or simply not perceive questions of distribution. This does not mean that system dynamicists are unable to deal with distribution questions, just that their paradigm gives them a certain reluctance to disaggregate.

Three problems that recur in all modeling techniques but that are relatively less bothersome in system dynamics than in other modeling schools are estimation of parameters, sensitivity testing, and assessment of model validity.

Parameter estimation is less important in system dynamics, and statistical estimation procedures are used less, for three reasons. First, most system dynamics models are not directed to problems of detailed implementation or precise prediction, but to problems of general understanding that do not require highly accurate numbers. Second, because of the long-term nature of most system dynamics problem statements, parameters are likely to exceed historic ranges, so estimation based on historic data alone would be insufficient. Third, the nonlinear feedback structure of system dynamics models renders them less sensitive to precise refinements of parameter values.

The general insensitivity of system dynamics models is partly a result of their feedback structure, but it is also partly due to the way sensitivity is defined in the system dynamics paradigm. Model output is read not for quantitative predictions of particular variables in particular years, but for qualitative behavioral characteristics. A model is said to be sensitive to a given parameter only if a change in the numerical value of the parameter changes the entire behavior of the model (from growth to decline, for example, or from damped oscillation to exploding oscillation). Sensitivity of this kind is extremely rare, both in system dynamics models and in social systems, but it does occasionally occur. In fact, detection of a particularly sensitive parameter is an important result of the modeling process, because it earmarks that parameter as one that must be estimated carefully or one that might be an effective site for policy input.

No rigorous theory or procedure exists in system dynamics for performing sensitivity analysis, and this is a weakness of the field. On the other hand, the informal structure-behavior theorems that characterize the paradigm sometimes permit an experienced dynamicist to locate possibly sensitive parameters by inspection of the model structure and thus to eliminate the necessity of testing every possible parameter in the system.

This intuitive approach to sensitivity testing is very effective in small models but almost unusable in large ones.

The system dynamics paradigm also handles the problem of model validity qualitatively and informally. There is no precise, quantitative index to summarize the validity of a system dynamics model. In fact, system dynamicists do not usually use the term validity. Reference is made to model utility; is the model sufficiently representative of the real system to answer the question it was designed to answer? A system dynamicist begins to have confidence in his model when it meets these conditions.

1. Every element and relationship in the model has identifiable real-world meaning and is consistent with whatever measurements or observations are available.
2. When the model is used to simulate historical periods, every variable exhibits the qualitative, and roughly quantitative, behavior that was observed in the real system.
3. When the model is simulated under extreme conditions, the model system's operation is reasonable (physical quantities do not become negative or exceed feasible bounds, impossible behavior modes do not appear).

These standards are imprecise and do not lend themselves to quick evaluation. They are also quite difficult to achieve in practice. The issue of model validity is an unresolved one in every modeling field. System dynamics approaches it by admitting the indeterminancy of the very concept of validity and by establishing performance standards that are qualitative but demanding.

The most difficult problems in system dynamics appear in the process of modeler-client interaction. The system dynamics paradigm leads the analyst naturally to a long-time horizon and wide-boundary approach to any problem. This viewpoint often overlooks the very real short-term pressures and constraints felt by most decision-makers. The result may be an impasse; the client cannot take the broad perspective of the modeler, and the modeler is convinced that no other perspective will lead to a problem solution.

Since system dynamicists assume that most problems, like most model elements, are endogenous to the system, they will look for and often find internal decisions to be a major cause of problems. Thus a system dynamics study will often (not always) lead to the conclusion that the problem is caused by the internal structure and the decisions being made in current systems. The recommended solution often requires structural change. This change may be as simple as bringing new information to bear on a decision, but it may also involve revision of goals, reward structures, or areas of authority. These recommendations are often politically unacceptable. This problem is intrinsic to the basic paradigm of system dynamics and the nature of public decision-making and will probably always be a factor hindering the practical use of system dynamics in the policy world.

Econometrics: General Characteristics*

Econometrics is defined as the use of statistical methods to verify and quantify economic theory. A set of theoretical

---

*Econometrics is a more widely-practiced and more varied field than systems dynamics, and no general description can cover the diversity of individual practitioners. The following description tends to capture the common characteristics of the majority of econometric models.

relationships that has been verified and quantified for a particular economic system constitutes an econometric model of that system. The model can be used for structural analysis, for forecasting, or for testing the effects of policy alternatives.

The field of econometrics combines tools and concepts from the two older fields of statistics and economics. Therefore it shares aspects of both those paradigms, as well as adding its own special perspectives to the world-view of its practitioners. Statistical economics developed in the 1930's as a result of a rising interest in the quantitative behavior of national economic variables, especially aggregate consumption, which was postulated to be a major cause of the problems of the great depression. The journal Econometrica was begun in 1933. By the late 1930's Jan Tinbergen had constructed the first dynamic models of the Dutch, United States, and British economies (Tinbergen 1937). Much theoretical and practical work had already been done by the early 1950's, when the development of the computer permitted a great expansion in the scope and complexity of econometric models.

The dominating characteristic of the econometric paradigm is its reliance on statistical verification of model structure and model parameters. Econometricians are forced by their paradigm to tie their models firmly to statistical observations of real-world systems. The formulation of an econometric model may be divided theoretically into two sequential phases, specification of structure from economic theory, and estimation of parameters by statistical analysis. The second phase is the center of concern, however, occupying most of the modeler's time and attention and most of the pages in econometric textbooks and journals. To some extent the mathematical and data requirements of the estimation phase enter into the specification phase as well.

The information base from which an econometrician can draw his model structure is the same one underlying system dynamics or any other modeling technique--abstractions, intuitions, personal experiences, statistical data, established wisdom, experimentation, and guesswork. In practice, most econometricians are attracted to questions about the precise, short-term values of economic variables. They find most of the concepts they need in traditional economic theory. They tend to make only limited use of theories from other disciplines, and when they do, their bias tends to be as much toward the social sciences as the system dynamicists' bias is toward the physical sciences. No special distinction is made between the properties of physical and information flows in econometric models. For example, many of the common variables in econometric models are expressed in units of unconserved monetary stocks and flows, even when they stand for conserved physical stocks and flows (examples are production, consumption, capital, investment, depreciation, imports, and exports).

The underlying economic theory from which econometrics is drawn is much richer in static concepts than dynamic ones, perhaps

because much of the theory was developed before computer simulation allowed dynamic analysis of complex, nonlinear systems. Much attention is paid in economics to the optimum or equilibrium points in a system, comparitively little to the path of approach to equilibrium or the time required to attain it. Although many econometric models are dynamic, they maintain their parent field's emphasis on optima and equilibria rather than on dynamic characteristics.

Economic theory also leads econometric modelers to create structures that are partially open--driven by many exogenous variables that must be forecast independently from the model--rather than entirely closed into feedback loops that drive the system through time. Economics evolved as an open-system body of theory for several reasons. Economic systems are strongly driven by forces outside the disciplinary boundary; resources come from the domain of geology, weather fluctuations from meteorology, consumer motivations from psychology and sociology, labor availability from demography. Furthermore, the relatively short-term focus of many economic problem statements means that analysts often need not take into account feedback processes with long time delays.

When two-way causation does appear in econometric models, it is typically represented by means of simultaneous equations. The simultaneous-equation formulation is equivalent to assuming that system equilibrium will occur within one calculation interval, Although most econometric models contain simultaneous-equation formulations and are driven dynamically by exogenously forecasted variables, many models also contain some feedback through lagged endogenous variables. These formulations are not essentially different from those in system dynamics models. The distinction between the two approaches is one of relative emphasis, not absolute contrast. Econometric models contain some feedback relationships, some of which are lagged; system dynamics models are composed almost entirely of feedback relationships, all of which are lagged.

Even within the disciplinary boundary of economics, the variables that can be included in econometric models are restricted to a subset of all conceivable elements, because of the necessity for statistical validation. Each element in an econometric model must be observable, and sufficient historic observations of it must exist to permit precise estimation of its quantitative relationship to other variables. That requirement tends to eliminate the inclusion of most of what system dynamicists call the information components of any system, especially the motivations behind human decisions. These motivational components are not absent from economic theory, which contains many inherently unobservable concepts such as marginal utility, indifference curves, and the profit motive. But none of these ideas are easily measured or contained as explanatory variables in econometric models.

The requirement of observability is not as confining to econometricians as a system dynamicist might think. In the long

run it creates the pressures that are already improving and expanding data-collection efforts around the world. Furthermore, useful but unmeasured concepts eventually can become sufficiently well defined to be measured and included in data bases. No GNP statistics were available until economists devised the concept, found it useful, and figured out how to measure it.

Econometricians can often represent an unobservable concept by means of a closely-correlated tangible substitute or proxy. Literacy may suffice as a stand-in for degree of modernization, rainfall may be a proxy for all the effects of weather on crop production, or advertising expenditures may be used to represent some of the perceived-utility assumptions underlying a consumer demand equation. In other words, an econometrician can transform a direct causal hypothesis, such as "in early stages of modernization people's material aspirations rise and so they consume less and save more for investment to increase their future consumption," into an indirect hypothesis about correlation of observables, such as "at low income levels literacy is inversely correlated with consumption." The use of correlated rather than direct causally-related variables allows econometricians to proceed in spite of the requirement of empirical validation, but it also reinforces that requirement because a double set of assumptions has been made. A relationship has been hypothesized not only between modernization and consumption, but also between modernization and literacy. Both relationships are tenuous and always subject to change, and therefore they must be rechecked continuously against real-world data.

The principal technique used to obtain parameters for econometric models is least squares estimation, a method that generates the set of numbers that best fits a postulated general relationship to historic observations and that also provides a quantitative measure of how good that fit is. The theoretical and mathematical requirements of this method impose several conditions that cause econometric models to depart from economic theory. For example, it requires that the equations be convertible to a form in which all parameters to be estimated enter linearly. As a consequence, most relationships in an econometric model are linear or log-linear. The assumed relationship between literacy and consumption is most likely to be expressed as:

$$\text{consumption} = \beta_o + \beta_1 \text{ (literacy)} + E$$

or perhaps as:

$$\log \text{ (consumption)} = \beta_o + \beta_1 \text{ (literacy)} + E$$

where β_o and β_1 are constants called <u>structural coefficients</u>, to be determined by fitting historical data for consumption and literacy. The "error term" E measures the observed variation in consumption that cannot be accounted for by variations in literacy.

Another requirement of least squares estimation is that the variation in each explanatory variable must not be linearly dependent on the variation in any other variable and must be strictly independent from the error term. Thus if consumption were postulated to be a function of both literacy and income, [consumption $= \beta_o + \beta_1$ (literacy) $+ \beta_2$ (income) $+$ E], the statistical procedures for estimating β_o, β_1, and β_2 will be accurate only if there is no high degree of correlation between income and literacy or between either of those and any of the omitted factors that might influence the error term. The effect of this requirement on model specification is a subtle psychological one; in order to avoid the numerical biases that result from co-variance of variables with each other or with the error term, econometricians tend to include relatively few explanatory variables in their equations.

> By the very existence of a large number of intercorrelations among all economic variables we can estimate but a few partial coefficients with tolerable precision. This accounts for the contrast between economic theory and empirical research. The theory is comprehensive: if we list the determinants of, say consumption or investment that have been discussed by economists, we may easily find some ten or twenty distinct effects. But in econometric research we rarely try to estimate more than four or five coefficients. (Cramer 1971)

The structural coefficients in an equation like the one relating literacy and consumption are estimated for the system of interest by finding observed values for all variables over some historical period or over some cross-section of subsystems (families, nations, firms, etc.). Ideally, the observed values are used to estimate the structural coefficients of the model, and then the model with its estimated parameter values is used to generate or simulate the values of system variables for another time period or over another cross-sectional sample. The entire procedure depends upon the assumption that the underlying causal mechanisms do not change in form, strength, or stochastic properties, from the estimation period to the forecasting period or from one cross-sectional sample to another. Various statistical indices, such as the square of the multiple correlation coefficient (R^2), are used to summarize the extent to which the model-generated values can be expected to duplicate the variance in the observed values.

Econometric models tend to deal with highly aggregated quantities, even more aggregated than those in system dynamics models. Typically, few variables are included and the models are small compared to other kinds of models.

Econometricians tend to represent systems as highly linear, partly open, at or near equilibrium, and centered around variables that fall within the disciplinary boundary of economics. The real-world systems that are most congruent with this image encompass flows of economic goods and services, money and prices, over a fairly short time horizon. In these systems many important influences are indeed exogenous, and many relationships are constrained within ranges that are very nearly linear. Also, over the short term the numerical coefficients derived from historical observations are still likely to be valid. If appropriate data are available, econometric methods can provide very precise information about such systems. Thus econometric models are mostly short-term prediction of aggregate economic variables. They are least applicable to policy questions that may range across disciplines, over long time horizons, or into circumstances that have not been historically observed.

Examples of questions addressed by recent policy-oriented econometric studies include:

> Will a change in the oil import quotas of the U.S. aggravate the shortage of domestic natural gas from now to 1985, and if so, what wellhead natural gas price would alleviate the shortage? (Spann and Erickson 1973)

> What will be the effects on U.S. economic growth from now to 2000 of more or less government spending, faster or slower population growth, sustained or decreased technical progress? (Hudson and Jorgenson 1974)
>
> What will be the quarterly consumer price index for food over the next four quarters? (Barr and Gale 1973)
>
> How many acres will be planted in wheat in the U.S. next year if government acreage restrictions and loan programs are altered in various ways? (Hoffman 1973)
>
> How will fiscal and monetary control decisions of the Federal Reserve Board affect the U.S. macro-economy over the next few years? (Modigliani, et.al. 1973)

Most but not all of these policy questions fall within the short-term, narrowly-bounded range of the implementation stage of decision making.

Econometrics: Problems and Limitations

The greatest strength of the econometric paradigm is its insistence on continuous, rigorous checking of theoretical hypotheses against real-world data. This strength leads, however, to two problems already noted: the statistical methods used for estimation impose artificial restrictions on the initial formulation of the model, and the data necessary for proper verification are seldom available.

The mathematical requirements of estimation cause econometricians to represent economic systems as linear, mostly simultaneous relationships connecting a few aggregate economic variables by means of historically-observed coefficients. A system dynamicist's bias causes me to suspect that real economic systems are nonlinear, multivariable, time-delayed, disaggregate, and ecological-socio-economic, and they may respond to policy decisions in ways that are not represented in historical data. However, there must certainly be parts of these systems that fit the narrow domain of econometrics quite well. Within these areas econometric techniques can produce accurate, informative, precise, and useful predictions. The major problem in econometric modeling is to recognize the limits of the congruent areas and resist the temptation to push outside them. Thoughtful econometricians know these limits well and seem to conclude that general macro-economic forecasting purposes are outside the limits:

> What then is econometrics....best suited for? I myself would place economic problems of the firm in the fore.. ..The problems confronting a firm are in general much less complex than those confronting the economy as a whole: they often are truly of a partial nature. Secondly, the number of observations of the same social system can here frequently be increased. We do not have to face the dilemma of the need for large samples in a world changing rapidly during sampling.
> If we wish to use econometrics for macro-economic purposes--to which it first turned, perhaps because disciplines as the outflow of human perversity always first turn to the field of application least suited for them--I would think that it can well be used to test which of a large number of economic hypotheses can best explain an economic reality precisely defined as to time and place. (Streissler 1970, pp.73-74)

Nearly every econometrician would list the lack of good empirical data as the most annoying and constricting problem in his field. Econometric researchers pay great attention to data problems and have developed names and categories for the most frequently occurring ones:

> Among the more important problems are that there is simply not enough data (the degrees of freedom problem); that the data tend to be bunched together (the multicollinearity problem); that because changes occur slowly over time, the data from time periods close together tend to be similar (the serial correlation problem); that there may be a discontinuous change in the real world so that the data refer to different populations (the structural change problem); and that there are many inaccuracies and biases in measuring economic variables (the errors of measurement problem). (Intriligator 1972, p.157)

Econometric techniques include a number of ingenious methods for recovering from data problems and for extracting maximum possible information from minimal real-world observations. Unfortunately, none of these methods can create more information than is already there, and a process that overcomes one data problem usually makes another one worse:

> For example, replacing annual data by quarterly data increases the number of data points but tends to aggravate both the multicollinearity and the serial correlation problems; eliminating data points referring to unusual periods such as during war years, overcomes the structural change problem but aggravates both the degrees of freedom and the multicollinearity problems; and replacing variables by their first differences overcomes the serial correlation problem but aggravates the errors of measurement problem. (Intriligator 1972, p.157)

Another criticism econometricians commonly voice about their own field is that econometric modeling is often done badly. In part this may be a byproduct of widespread use of econometrics and of very convenient computer software--the same mixed blessing we have encountered in system dynamics and will encounter again when we go on to discuss other modeling techniques. In the case of econometrics, the statistical packages that are now standard equipment at most computing centers can be used rather easily by skilled analysts, and also by those who have never understood or who have entirely forgotten the assumptions underlying the regression techniques. The result can be a blind manipulation of data and an overconfident belief in computed results. Mechanical application of statistical techniques may be substituted for experience with the real-world system, for knowledge of economic theory, and for thoughtful evaluation of conclusions. This is not an inevitable problem; it can be overcome by better training of modelers, better self-regulation of the econometric profession as a whole, and continuous questioning and review of econometric modeling efforts by modelers, clients, and sponsors.

Because econometric models are partially open systems, they tend to be more sensitive to parameter variation than are system dynamics models. The difference in sensitivity is magnified by the fact that econometricians are usually striving for much more precise statements about the future than are system dynamicists. One would expect, therefore, that sensitivity analysis would be a central concern of econometricians. However, the procedures for carrying out and reporting sensitivity tests, especially for alternate forecasts of the usually numerous exogenous variables, do not seem to be formalized or regularly reported in model documentation. Testing every believable combination of values for exogenous variables would be an impossible task, and the intuitive structure-based hunches that system dynamicists use to detect sensitive points are less applicable to econometrics.

Econometricians determine the validity of their models by the use of statistical tests of model-generated data against real-world data and by the informal comparison of model results with their mental models of "reasonable" values for economic variables. These two validity tests are probably as good as any other when the statistical tests are done honestly and skillfully, and when the modeler has a deep understanding of the workings of real economies. A less honest, skillful, or knowledgeable modeler, however, can produce with these tests evidence of validity for almost any model. In other words, although econometrics techniques include a number of sophisticated statistical validity tests, establishing confidence in a model's output is as difficult and uncertain in this modeling school as it is in the others.

A Note on Input-Output Analysis and Optimization

The next two modeling methods discussed here--input-output analysis and optimization--are not easily characterized as self-contained, world-shaping paradigms, as system dynamics and econometrics have been. Each of these methods originally dealt with only a specific part of social systems; input-output analysis with production functions, and optimization with the nature of certain decisions, especially investment and allocation decisions. Neither provides in itself a complete method of representing social systems, and therefore each is perhaps more correctly defined as a sub-paradigm that provides its own way of looking at systems but also takes on the characteristics of whatever paradigm it is combined with. As we shall see, both methods are regularly integrated with econometric models, and therefore they tend to take on some aspects of the econometric paradigm. In theory they could be combined with system dynamics models as well.

Input-Output Analysis: General Characteristics

The first semblance of an input-output analysis appeared in 1958 when Francois Quesnay constructed his Tableau Economique representing the interdependence of various wealth-producing activities on a single farm. The chain of development can be traced for nearly two hundred years, through such economists as Leon Walras and Vilfredo Pareto, until it reaches Wassily Leontif, who published his original paper laying the foundation of modern input-output analysis in 1936 (Leontief 1936). The first official input-output table for the United States economy was compiled by the Bureau of Labor Statistics for the year 1947. By 1963 at least 40 countries had completed their own national input-output tables, and now at least one collection of more than 300 such tables from 80 countries has been assembled.* Input-output tables are widely used for national economic planning in planned-economy countries and for forecasting and policy analysis in market-economy countries. Although input-output techniques arose from the economic paradigm and were originally intended for analyzing inter-industry flows of money and goods, more recently the field has expanded to include flows of other quantities, such as energy and pollution (see, for example, Leontief 1970).

An input-output analysis begins with a set of data measuring the internal flows of money or goods among various sectors of an economy over a given year. These flows are summarized by an input-output table, which is nothing more than an array of the purchases made by each sector (its inputs) and the sales of each sector (its outputs) from and to each other sector of the total economy. The inputs and outputs might be expressed in physical units but they are more often expressed in terms of monetary value. In that case the table would be a summary of dollar flows to and from each sector of the economy.

---

*By A. Bottomley of the University of Bradford, England.

An example of a hypothetical input-output table is shown in Figure II. In this economy six industries have been distinguished, labeled A-F. The flows of inputs and outputs among these six industries are shown in the upper left-hand corner in units of billions of dollars. Thus Industry A used 10 billion dollars worth of its product itself, and sold 15 billion dollars worth to Industry B, 1 billion to Industry C, etc. The final demand for the economy's products (sectors that purchased goods but did not transform and resell them) is shown in columns 7-11. Final demand consists of additions to inventory, exports, government purchases, additions to capital plant, and households, which

		Processing Sector						Final Demand					
	Outputs¹ → Inputs²	(1) A	(2) B	(3) C	(4) D	(5) E	(6) F	(7) Gross inventory accumulation (+)	(8) Exports to foreign countries	(9) Government purchases	(10) Gross private capital formation	(11) Households	(12) Total Gross Output
Industry Producing — Processing Sector	(1) Industry A	10	15	1	2	5	6	2	5	1	3	14	64
	(2) Industry B	5	4	7	1	3	8	1	6	3	4	17	59
	(3) Industry C	7	2	8	1	5	3	2	3	1	3	5	40
	(4) Industry D	11	1	2	8	6	4	0	0	1	2	4	39
	(5) Industry E	4	0	1	14	3	2	1	2	1	3	9	40
	(6) Industry F	2	6	7	6	2	6	2	4	2	1	8	46
Payments Sector	(7) Gross inventory depletion (−)	1	2	1	0	2	1	0	1	0	0	0	8
	(8) Imports	2	1	3	0	3	2	0	0	0	0	2	13
	(9) Payments to government	2	3	2	2	1	2	3	2	1	2	12	32
	(10) Depreciation allowances	1	2	1	0	1	0	0	0	0	0	0	5
	(11) Households	19	23	7	5	9	12	1	0	8	0	1	85
	(12) Total Gross Outlays	64	59	40	39	40	46	12	23	18	18	72	431

¹Sales to industries and sectors along the top of the table from the industry listed in each row at the left of the table
²Purchases from industries and sectors at the left of the table by the industry listed at the top of each column

Figure II

From William H. Miernyk, <u>The Elements of Input-Output Analysis</u>, New York: Random House, 1967, p.9.

means final purchases by private domestic consumers. Industry A added 2 billion dollars worth of production to its inventories, exported 5 billion dollars worth, sold 1 billion to the government, etc. Reading down a column gives a record of any industry's expenses for inputs. Thus Industry C bought 1 billion dollars worth of Industry A's output, 7 billion from Industry B, and used 8 billion dollars worth of its own product. Industry C also depleted its inventory by 1 billion dollars, imported 3 billion dollars worth of materials from abroad, paid 2 billion in taxes to the government, depreciated its capital by 1 billion, and paid 7 billion to households in the form of wages.

If input-output analysis ended here, it would just be a handy way of displaying and communicating historical information about the complex interdependence of many subsectors in an economic system. However, a table like this is the beginning of the analysis, not the end. The next step is to assume that the numbers in the table arise from continuous, linear relationships between the inputs and outputs of each sector. If that assumption can be made, the entire table can be rewritten in more general terms; for any quantity of production, how much input is required to produce one unit (or one dollar's worth, if the table is in monetary terms) of output. This rewritten table is called the structural matrix for the economy, and the numbers in it are referred to as the technical coefficients defining the linear relationships between inputs and outputs.

The structural matrix can be used to indicate for any hypothetical output of any sector what the direct inputs to that

sector must be. One more transformation of the matrix allows another sort of question to be answered--what will be the effect on the total economy of a change in the demand for one item? For example, if the final demand for Industry A's product goes up by 1000 units per year, the structural matrix indicates that Industry A must buy more inputs from the other industries. Each of those industries must then produce more and buy more inputs, including some inputs from Industry A. So 1000 units more final demand for A will actually require more than 1000 units of added output from A, and will also increase the production of all other industries.

In order to account for this interdependence of input factors, the structural matrix is inverted. The inverse matrix is called the table of direct and indirect requirements. Each entry in this matrix indicates the total (direct and indirect) output from the row industry that is required for one unit of production of the column industry. This table can be used to derive much useful information about the economy. It can indicate how much total production of all intermediate and final goods would be needed to satisfy any desired pattern of final demands. If final demand for some item suddenly shifts, the necessary changes in production of all supporting industries can be traced through the economy. The columns of the table can be used by individual firms for cost planning, and the rows for market analysis. Above all, since the necessary interlinking of industries is clearly represented, consistent planning and analysis on a fairly detailed scale becomes possible.

For example, expansion of automobile production can be discussed taking into account not only additional steel production as a direct input to the automobiles, but also additional steel production for more railroad cars to deliver the additional steel, and still more steel production to build oil refineries to provide more oil to run the additional railroad cars.

It is important to remember that the logical step from the input-output table, which is a summary of the operation of an actual economy in a particular year, to the structural matrix, which is a generalized model that allows planning and forecasting, depends on three assumptions:

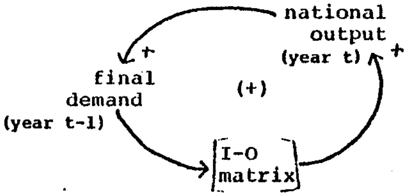
1. Linearity. The numerical relationships between inputs and outputs in each industry must remain constant over all ranges of inputs and outputs. This is equivalent to assuming constant returns to scale and no significant technological changes. Linear relationships are mathematically necessary in order to invert the structural matrix.

2. Continuity. Each industrial sector must be able to expand or contract output marginally while maintaining the same relationship between inputs and outputs. Thus no input or output must occur in the form of large indivisible lumps.

3. Instantaneous adjustment. Since there is no time dimension in an input-output table, there is no way of representing delays in the availability of inputs or in the production of outputs. Using such a table to investigate the effects of changes in final demand, technological conditions, or other

factors can give no information about the time necessary to achieve those changes. Thus use of an input-output analysis is essentially equivalent to assuming that shifts in inter-industry use of intermediate products, capital, and labor can occur quickly and with no bottlenecks.

The linearity and continuity assumptions are intrinsic to input-output analysis and can be weakened only at the cost of great mathematical complication. The instantaneous adjustment assumption holds for <u>static</u> input-output analysis of the sort we have discussed so far. Input-output analysis can be made <u>dynamic</u> by combining it with some other modeling method that provides a way of moving the matrix forward through time. For example, an input-output table for 1975 might be used to calculate total national output (GNP); an econometric analysis might relate national output to final demand (consumption) in various sectors in 1976. The final demand prediction can then be used in the table of direct and indirect requirements to calculate all intermediate production levels, which will add up to a prediction of 1976 national output. The process can be iterated to carry the forecast further into the future. The causal assumptions behind this dynamic analysis form a positive feedback loop that will generate exponential growth in national output.



The technological coefficients may also be assumed to change with time as a result of exogenous technological developments or relative price shifts causing substitution. New industries may also be added to the system, enlarging the structural matrix.

Any dynamic application of input-output modeling adds to the assumptions in the static analysis another set of assumptions that express the dynamic relationships of inputs, outputs, and technical coefficients. This second set of assumptions may be simple extrapolation, intuition, or guesswork; or it may be derived from one of the dynamic modeling paradigms such as econometrics or system dynamics.

Input-output analysis shares with econometrics the use of directly observable economic data rather than the attempt to represent underlying causal mechanisms. In fact, input-output analysis is even less concerned than econometrics with why things happen; it seeks only to represent what has happened. The decision rules that determine the interindustry flows remain implicit. The entire basis for the input-output table is the actual performance of an economic system in a given year. No information is available about whether that performance was typical, optimal, efficient, or desirable, nor whether the system was in equilibrium.

Assumptions of linearity and continuity are most applicable to systems that are not greatly different from the system that generated the initial data. Thus input-output analysis is

most useful for analyzing marginal changes in economic systems over the short term. Input-output models can add considerable detail to economic analysis and forecasts, and, as the following examples of recent studies indicate, they can represent complex flows of dollars, goods, and even energy and water through industrial production systems:

> A forecast for the U.S. economy, disaggregated into 90 industries, from 1965 to 1975. (Almon 1966)
>
> An analysis of water requirements by industry in the state of California (technical coefficients expressed in acre-feet of water per million dollars of output). (Lofting and McGauhey 1963)
>
> A description of the effect over 60 industries and 19 geographic regions of a proposed shift in U.S. government spending from military to non-military procurements. (Leontief, et.al., 1966)
>
> A record of the energy flows within the U.S. economy (357 sectors) in 1963 and 1967, used to analyze the inter-industry effects of various national energy policies. (Herendeen and Bullard 1974)

Input-Output Analysis: Problems and Limitations

Static input-output models are limited in scope and fixed in structure. Therefore the structural-conceptual problems of econometrics and system dynamics are absent from this method, at least in its static form. The analyst need spend little time wondering about the unseen mechanisms by which variables might be interrelated. This gain in conceptual simplicity is realized at a cost in range of applicability--many pressing policy questions cannot be addressed with an input-output model. For those questions that do fall within the range of applicability, however, structural ambiguity is not a problem.

The only conceptual problem an input-output analyst must face is the degree of aggregation of his model. The strength of input-output analysis and the reason for its existence is its ability to provide a disaggregated picture of a complex system. Practically, however, disaggregation has its limits. More sectors are included in an input-output table only at the cost of obtaining more data and using very much more computer time and space. Since a firm-by-firm disaggregation of a large national economy would probably strain any existing computer's capacity, some aggregation of productive activities is necessary.

Aggregation of entities that are actually unalike with respect to some characteristic of crucial importance is a danger in all schools of modeling. Ultimately, the decision about what quantities to aggregate can only be resolved by reference to the purpose of the model. Unfortunately, the construction of a major input-output model is so time-consuming that new models are not likely to be prepared for special purposes. Most modelers begin from some standard "general purpose" model, usually made for a national economy. Such models are difficult to construct and can be deceptive to use. Any amount of detail will be useful to someone, and any degree of aggregation may bury some important distinction. The degree of disaggregation in these models is decided not by purpose, but by data availability and computer capacity.

If an input-output analyst saves time in the conceptualization stage of model-making, he spends it many times over in assembling the data to fill all the entries in the input-output table. The kinds of data required are relatively straightforward--

they are measurable physical or monetary flows rather than the unmeasured attitudinal variables the system dynamicist must often deal with. But extracting these numbers from actual accounting records and making them consistent is far from easy.

Finding appropriate and complete industry records at the proper degree of aggregation is the first problem. For some countries or regions they simply do not exist. Even in statistically advanced countries such as the U.S., it takes five to ten years to assemble a national table, a great handicap for a short-term linear method whose assumptions become less acceptable as the time between data-base year and forecasting period increases.

Input-output tables require internal consistency, since total inputs must equal total outputs within each industry and for the system as a whole. Assembling a table is an excellent way to check on the consistency of national accounts. Unfortunately, the data are almost never actually consistent; all inputs and outputs are not accurately recorded and they are unlikely to match each other as they should. While this is a useful lesson to learn about economic data, it threatens to stop the input-output analysis in its tracks unless the table can be "reconciled". Reconciliation of an input-output table relies on the modeler's judgement, intuition, and knowledge of the real system. It is a "fudging" step that is rarely documented, and it introduces assumptions that are rarely examinable. Fudging of some sort occurs in all types of modeling, and it may well be that the rigorous structure of the input-output table restricts the degree of fudging freedom to the point where the effect of reconciliation is negligible. That is a hard conclusion to prove or disprove, however, since the reconciliation process is rarely discussed, and since methods of sensitivity analysis are as primitive in input-output analysis as they are in most other modeling schools.

Like econometric models, input-output models are strongly affected by mathematical requirements, especially in the central assumption of linearity. This assumption may be more unrealistic for input-output analysis than for econometrics because of the greater degree of disaggregation--individual firms or industrial sectors may run into diminishing or increasing returns to scale, supply bottlenecks, or discontinuous, lumpy inputs before aggregated economies do. As we have already indicated, linear assumptions may be entirely acceptable in the short term; the problem is to refrain from pushing the technique beyond its range of applicability.

Optimization: General Characteristics

During World War II the planning and coordination of U.S. military operations became so complex that several experiments were begun to compute mathematically the deployment of personnel, supplies, and maintenance activities that would best achieve wartime objectives. After the war the Air Force set up a research group for the Scientific Computation of Optimum Programs to continue working out methods for calculating optimal allocation of resources. In 1947 this group, led by G.B. Dantzig, developed the first linear programming model and the Simplex

method for finding optimal solutions. Linear programming spread rapidly to many fields of application, particularly to engineering, management, and economic analysis. As computers and mathematical understanding improved, extensions to nonlinear optimization methods appeared.

Optimization models select from a large number of possible choices the single one that allows maximum achievement of some objective. For example, extremely complex optimization programs are used regularly in oil refineries to choose that combination of feedstocks, operating sequences and conditions, blending methods, storage locations, and shipping routes that will supply a large variety of products to a large number of widely-dispersed markets at minimal cost. Optimization may be the computer modeling method most often used as an input to actual decision-making, especially in industry.

The optimization method requires that problems be stated in a simple and unvarying format:

Maximize or minimize: objective function
By manipulation of: control variables
Subject to: constraints

The objective function is an expression either of the welfare of the system (such as profit, output, or per capita income), which is to be maximized, or of the cost to the system, to be minimized. The control variables are all the generic policy choices available to the decision-maker. For example, in an agricultural planning problem the control variables might be land area planted to each kind of crop and fertilizer and irrigation water applied to each kind of crop. The constraints express the desired or necessary relationships among the control variables. In the agricultural example, the areas allocated to each crop must be equal to or less than the total cultivable land available, or the total fertilizer budeget cannot exceed a certain amount.

The constraints define a complex, multidimensional surface upon which the desired minimum or maximum must lie. A major part of optimization theory is devoted to finding efficient techniques for searching out maxima and minima and for ensuring that a discovered maximum is an absolute extreme point, not just a local one. As the dimensions and complexities of the possible surfaces increase, the mathematics can become very complicated and the search processes so tedious that they can only be done by a computer.

The mathematical difficulties of optimization are simplified if the objective function and constraints are expressed as linear equations and the variables are continuous. If these conditions can be met, the problem is one of <u>linear programming</u>. The multidimensional surface defined by the constraints is reduced to a faceted surface--in three dimensions it can be imagined as a polyhedron. Any maximum or minimum on such a surface must be at a corner; that is, at an intersection of two or more constraints. Thus the search procedure can be confined to a few points on the perimeter of the problem surface, and the location of an absolute maximum or minimum becomes much more tractable. Standard search techniques exist for linear programming problems, and these have been incorporated into computer software packages that permit almost effortless solution to any problem, once it is stated in the linear programming format.

The conceptual paradigm of optimization, like that of input-output analysis, is rigid and limited in applicability, yet powerful and widely useful because the limited circumstances within which it is applicable recur frequently in the decision-making process. Optimization techniques can only be used when a clear objective function can be stated, when all the control variables available to the decision-maker can be specified, and when the constraints in the system can be defined precisely. These conditions are rarely met at the general understanding or policy formulation stages of decision-making. On the other hand, within the final stage of detailed decision-making, when a problem has been narrowed down to a choice among well-defined options to achieve a clearly-stated goal, optimization is uniquely useful. Furthermore, at earlier stages of problem definition, the identification of objectives, policy variables, and constraints provides a powerful set of organizing concepts that may be helpful in sorting out the complexities of problems even if they are not yet well-structured enough to be thoroughly analyzed by optimization techniques. In particular, the normative view of the world imposed by the optimization paradigm encourages discussion of concrete goals, which may in itself be a worthwhile exercise.

Within the relatively strict format of the optimization paradigm, an imaginative modeler has in fact a wide range of freedom. The objective function can be expressed, for example, as a minimization of cost, labor, or use of a scarce resource, or as a maximization of profit, output, productivity, or some measure of social welfare (such as life expectancy). Two objectives cannot be optimized at once, but secondary objectives can be expressed as constraints; in fact, constraints and objective functions are essentially interchangeable. For example, one may seek to maximize industrial output while insuring that energy use does not exceed a certain limit, or minimize energy use under the constraint that industrial output does not fall below a given target.

In addition to generating the best decision for achieving a given objective function, an optimization program can provide a clear picture of the trade-offs that are implied by that decision. By strengthening or weakening various constraints slightly or by changing parameters in the objective function, the analyst can investigate how different social priorities might shift the optimal decision point. The model can also indicate sets of objective functions and constraints that have no mathematical solution, and rule these out as inconsistent or unrealistic sets of goals.

Specification of an objective function is an obvious value statement; it is often dictated by the model's client.

Specification of constraints is a more disguised value statement as well as a representation of the environment within which the optimization decision is made. Here the judgement and knowledge of the modeler are particularly important. In engineering, transportation, and other physical optimization models, the constraints on the system may be numerous and complex, but they are usually conceptually straightforward expressions of physical laws, material properties, or the actual spatial arrangement of the system. In social optimization models, the constraints must be expressed by a static or dynamic model of the important interrelationships of the social system. Linear programming and other optimization methods do not provide the basic concepts for constructing this model. In practice, therefore, the most important assumptions in social optimization models are derived from some other paradigm, often from economics. For example, common constraints in social-system optimization models are requirements that supply must equal demand or that output must be a Cobb-Douglas function of capital and labor. Optimization procedures have been combined with all three of the modeling techniques we have already described--system dynamics, econometrics, and input-output analysis--and in each case the assumptions, strengths, and weaknesses of the other paradigm were dominant influences on the representation of constraints in the optimization program (see, for example, Oerlemanns, et.al. 1972; Bruno 1966; Chenery and MacEwan 1966).

Optimization models tend to be highly disaggregated (especially when combined with input-output models or when written for detailed industrial production decisions). They may contain hundreds or even thousands of equations and take significant computer time to run (several minutes to a few hours of central-processing time, depending upon search procedures and starting positions). Although the basic organizing scheme of an optimization program (objective function, activities, constraints) is an intuitive and helpful way of expressing a policy problem, the actual representation of these quantities is often highly abstract, complex, and opaque. A large linear program is difficult to construct, debug, adjust, and run, and it is usually designed to be used and re-used for ongoing decisions--for annual investment allocations, for example, or for continuous adjustment of inventories or production processes. In other words, the optimization modeling process is product- rather than process-oriented, and the models are usually used as black-box inputs to recurrent, detailed decision-making. As an outsider reading the optimization literature, I rarely find clear explanations of model assumptions in nontechnical language as a part of model documentation.

Optimization models are most frequently used in engineering and industrial management. Examples include models to plan lowest cost transportation routes, to specify the most effective sites for sewage treatment plants on major rivers, to allocate electricity demand among various generating units, and to establish inventory ordering policies. These applications fit the optimization format very well; the objective function is clear and the constraints and activities are precisely known.

Virtually all these applications occur at the detailed implementation stage of the decision-making process.

Less typical are optimization models representing more general and long-term social decision problems. Some examples within the population/resources/development subject area are:

> A determination of the allocation mechanism for all types of energy resources in the Untied States that would minimize the total discounted costs of meeting a projected set of final energy demands from 1970 to 2170. (Nordhaus 1973)
>
> A model to plan educational system development by selecting the optimal number of students to enter the educational system at various times in order to meet forecasted manpower needs while minimizing costs. (Balensky 1976)
>
> A simulation model to test short-term government economic policy in Mexico, assuming that the market mechanism operates to maximize the sum of producers' and consumers' surplus. (Goreaux and Manne 1973)

Optimization: Problems and Limitations

Optimization models suffer from many of the same problems we have already encountered in other modeling techniques. Most of them are linear and static. When they are dynamic they take on the limitations of whatever dynamic paradigm they adopt. Linear programming search routines have been packaged into widely-available, easy-to-use software that can be misused by unskilled modelers. Data sources for optimization models are the same as those for other models, with the same problems. Validation of optimization models, like all other kinds of models, is a vague and uncertain process.

Two problems especially serious for optimization modeling are computer-time limitations and sensitivity. Because of the tedious process of searching for the optimum, computer costs for optimization models are generally high, often so high that cost is a limiting factor in testing the model. Disaggregation into regions or subsets adds to cost, as does stepping through time, since finding an optimum decision at each new time period or for each region requires an additional search process. The tradeoff between disaggregation and dynamic solution is usually severe; highly disaggregated models are typically solved for only one time period, and long-term dynamic models tend to be quite aggregated.

Optimization models, and especially linear programming models, can be extremely sensitive to small parameter changes. The output of an optimization model is a single precise point (or series of points over time)--the minimum or maximum point of intersection of an objective function with a complex, multidimensional constraint surface. Small changes in the slope of the objective function or shifts in the constraint surface may move the optimum point long distances, to completely different policy choices. For example, in dynamic linear programming models of national investment policy, the "bang-bang" problem appears; small parameter changes will shift the optimal investment pattern either all to the early years or all to the late years of the projection (Kendrick 1972, pp.204-205).

Of course it is essential for the modeler to be aware of such sensitivities in his model, and optimization modelers are more likely than other kinds of modelers to worry about sensitivity analysis. They have developed a number of sophisticated techniques for determining model sensitivity without requiring total model reruns. However, a complete sensitivity analysis is as rare in optimization models as it is in other kinds of models. The problems of sensitivity analysis is most serious for large social models, where there may be hundreds of sensitive parameters, where few of them can be precisely known, and where the cost of testing each one may be prohibitive.

As we have mentioned in discussing other modeling tools, a major problem in using any of them is recognizing where they are and are not applicable. The narrower the range of applicability, the more difficult the match between modeling technique and policy problem, and the greater the temptation to extend the technique beyond its appropriate areas of usefulness. Optimization is the most specialized and precise modeling tool we have discussed. It is most effectively used for the last accurate refinements in decision-making, when general understanding of the circumstances surrounding a decision is good, when broad policy options have been defined and assessed, and when objectives and constraints have been stated clearly.

An Example of Paradigm Conflict: Econometrics and System Dynamics

The four modeling techniques discussed here are complementary in several ways. For example, system dynamics provides a theory of causal structure and its relation to dynamic behavior that is a powerful guide to model specification. Econometrics offers numerous techniques for finding empirical parameters and for formal comparison of model results with real-world observations. One technique is particularly applicable to long-term analysis of possible changes in historic trends. The other is best suited to short-term precise prediction in situations that do not differ from the historic. It would seem that use of the two methods together might produce models that combine realistic structure with accurate parameters, models that are useful at every stage of the decision-making process, particularly for middle-term problems that are not easily analyzed by either method alone.

Unfortunately, this logical combination of two complementary modeling tools has not often been used. On the contrary, very few econometricians have bothered to learn system dynamics techniques, and those system dynamicists who have been schooled in econometrics do not regularly use its tools or concepts. Members of the two schools seem to regard each other as competitors rather than as potential collaborators, and find little to praise in each others' work.

In part this hostility may be due to the personalities of the methodological founders, the natural parochialism of academics, and inevitable jockeying for scarce funding resources. However, a closer examination of the two modeling paradigms reveals a deeper division, one that is not easily bridged. The basic world views upon which the two paradigms are built are quite different, as if they cut through reality with two perpendicular planes that only meet along one narrow line. Either paradigm, seen from the perspective of the other, looks

unrealistic and misleading. Methodological conversations between econometricians and system dynamicists tend to degenerate into classic cross-paradigm confusions. Key words such as "validation", sensitivity", and "prediction" are used in different ways based on different implicit assumptions.

Thomas Kuhn is not optimistic about building bridges across paradigm gaps:

> The proponents of competing paradigms are always at least slightly at cross-purposes. Neither side will grant all the non-empirical assumptions that the other needs in order to make its case....Though each may hope to convert the other to his way of seeing his science and its problems, neither may hope to prove his case. The competition between paradigms is not the sort of battle that can be resolved by proofs...
> The proponents of competing paradigms will often disagree about the list of problems that any candidate for paradigm must resolve. Their standards or their definitions of science are not the same....Communication across the revolutionary divide is inevitably partial....In a sense that I am unable to explicate further, the proponents of competing paradigms practice their trades in different worlds....The two groups of scientists see different things when they look from the same point in the same direction....That is why a law that cannot even be demonstrated to one group of scientists may occasionally seem intuitively obvious to another. Equally, it is why, before they can hope to communicate fully, one group or the other must experience the conversion that we have been calling a paradigm shift. Just because it is a transition between incommensurables, the transition between competing paradigms cannot be made a step at a time, forced by logic and neutral experience. Like the Gestalt switch, it must occur all at once...or not at all. (Kuhn 1970, pp.148-151)

I believe and hope that the paradigms of system dynamics and econometrics are not as totally incommensurable as Kuhn implies But there are certainly serious cross-paradigm translation problems that interfere greatly with attempts at synthesis. In this section I will look at both paradigms simultaneously, switching back and forth to see each from the point of view of the other. The resulting image will necessarily be a bit disjointed, since it will not have a constant reference point. It will also magnify the methodological division somewhat, because this description itself is an over-simplified model of reality. And, needless to say, it will not be a totally unbiased description, despite my efforts to make it so. If the following discussion does not induce mutual understanding in the "proponents of the competing paradigms", perhaps it will at least give uninvolved observers of the competition some idea of what each side is assuming as well as what it is saying.

As Kuhn says, the problem begins with the choice of a solvable problem. System dynamicists and econometricians are led by their paradigms to notice different problems and to strive for different kinds of insights into socioeconomic systems. Econometricians seem to feel that useful information must be detailed and precise--a picture that is not entirely in focus is not worth looking at. They see little substance in the ambiguous, qualitative, long-term output of system dynamics models. To achieve as much precision as possible, econometricians work with statistical methods, which require historic data bases, linear equations, and open structures. They develop little structure-behavior intuition, and not surprisingly, they feel that the long term is simply inaccessible to modelers. The lenses they use to look at the world are microscopic, not telescopic, and therefore they conclude that attempts to form a clear image of a happening far away can only be a waste of time.

System dynamicists regard any effort to gain precise predictions of social system as hopelessly naive. They regard human unpredictability as too dominant a factor in social systems to allow anything more than qualitative behavioral forecasts, even for aggregate systems where much unpredictability can be averaged out. Therefore they find it hard to understand the great effort econometricians go through to obtain better and better estimations or to quote their findings to six or seven significant digits. Especially when many exogenous variables must be predetermined, the whole econometric exercise looks to a system dynamicist like a transformation of one set of uncertain and unscientific guesses into a second set of equally uncertain guesses presented with deceptive, scientific-looking precision. System dynamicists should know from their own theorems of system behavior that most aggregate systems possess significant momentum, and that within a short time horizon the relatively simple structural hypotheses of econometrics are usually quite appropriate. But the system dynamics paradigm tends to reject not only the possibility but the utility of working within short time horizons. In the system dynamics world view the short term is already determined and thus unchangeable by policy. Furthermore, in system dynamics models policies designed only for short-term gain often lead to long-term loss.

These different ideas about what kinds of knowledge about the future are useful arise from basically different asusmptions about the nature of social systems. The econometric assumption reflects the common view of the policy-making world that the world is essentially dualistic and open. There is a sharp distinction between the economy and the environment (government, weather, Arab nations, consumers, investors, or whatever). The environment delivers specific inputs to which the system gives specific responses. Each system, input, and response may be unique, and thus particulars of different situations are more to be studied than similarities. The best strategy for policy is to foresee the next set of specific inputs and be prepared to give optimal responses to them. This view leads to policy questions about end states, rather than paths to those states, and about particular characteristics of the system under particular conditions:

> If the price of natural gas is deregulated this year, what will its equilibrium market price be? How much windfall profit would accrue to the gas companies?
>
> How much increase in income taxes would be required to reduce the current rate of inflation by 2%? What would that tax do to the unemployment rate?
>
> Given normal weather conditions, current fertilizer prices, and a subsidy of 5¢ per bushel, how much wheat will be produced in the U.S. next year? If no export embargoes are imposed, what will domestic wheat price be?

System dynamics, on the other hand, assumes that systems are primarily closed; not only does the environment influence them, but they influence the environment. In fact, the distinction between the system and its environment is rarely clear (except for obvious exogenous factors like incoming solar energy). Attention is focused on the general system reaction to general disturbances and on the dynamic path of a response rather than its end state.

> System dynamics...regards external forces as there, but beyond control and hence not worthy of primary attention...Instead, the focus is upon examining the organization's internal structure; the intent being to arrive at an understanding of how this structure...can be made more resilient to environmental perturbation. In adopting this approach, system dynamics is embracing the wisdom of the human body. The body, rather than forecasting--and then marshalling its forces in anticipation of--the arrival of each kind of solid and liquid input, remains continually poised in a state of general readiness for whatever may befall it. (Richmond 1976)

This assumption about the nature of systems would lead to a very different set of policy questions:

> How would deregulation of natural gas price affect the general depletion life-cycle pattern of U.S. natural gas reserves?
>
> What are the dominant positive feedback loops causing inflation? How could equally effective negative loops to counterbalance them be built into the economic system without causing unemployment?
>
> Why has wheat production fluctuated more in the past five years than in the preceding 15 years? Which sort of policy, direct price supports, increased buffer stocks, or increased exports, could induce stabilization of production while not increasing consumer prices?

After choosing different problems and dismissing the legitimacy and feasibility of each others problem areas, econometricians and system dynamicists go on to solve their problems with totally different procedures. The differences here have deep roots in conflicting theories of knowledge. Perhaps both sides would agree that the nature of the world and our perceptions of it produce a number of observable happenings that result from an underlayer of unseen causal motivations, events, and connections. The disagreement begins in deciding which part of that double-layered world to represent in a model.

Econometrics is firmly grounded in observable reality. Econometricians may speculate freely about unseen psychological and physical driving machinery, drawing on substantial causal theory from their parent paradigm of economics. But their models must contain explicitly not what they guess, but only what they know, and in their paradigm one can know only what one can measure. Therefore their models tend to represent surface phenomena only, with much causal structure implicit. There are no strong preconceived notions about the nature of that structure. It may be an interconnected web of feedback loops; it may be a series of unrelated stochastic forces; or it may be some combination of these. Whatever the underlying structure is, its nature and its relationship to the surface phenomena may change, therefore stochastic error terms must be added to equations, and a continuous stream of new observations must be obtained to verify that the system continues to run as it has in the past. Econometricians therefore feel a pressing need for more data, better measurements, more recent updating, better access to data bases.

System dynamicists, on the other hand, feel that statistical data represent only a small fraction of what one can know. They plunge enthusiastically into the lower layer of unseen causal relationships, armed with theories that help them relate visible dynamic variations in systems to invisible feedback-loop structure. They attempt to guess that structure, and to include it explicitly in their models. They are

searching for timeless general relationships; therefore they use data from any period and any subsystem, including, among many other sources, the same statistical data from which econometric models are derived. However, they generally prefer direct, qualitative observations of the physical processes and human actors in the real system to quantitative aggregate social indices. System dynamicists visualize a spectrum of increasingly precise information, ranging from intuitions, hunches, and anecdotal observations at one end to controlled physical measurement at the other, with social statistics somewhere in the middle. They declare that this spectrum offers far more information than is currently used, and that the real need is not for more data but for better use of the data already available. They point out that econometricians, by confining themselves to the narrow part of the spectrum consisting of social statistics, which contain no information about the operating policies, goals, fears, or expectations in the system, are hopelessly restricted in learning about how social systems work.

These two basic approaches to the interpretation and use of various kinds of knowledge result in continuous, fruitless cross-paradigm discussions about the relative importance of structure versus parameters. Econometricians probably spend 5% of their time specifying model structure and 95% estimating parameters. System dynamicists reverse that emphasis. Their long-term feedback models are prone to wild excursions if even one small information link is left unclosed but are often maddeningly unresponsive to parameter changes. Having worked with such models, system dynamicists find it difficult to imagine why anyone would bother to estimate most coefficients very accurately, especially when the coefficients are part of a model with an obviously defective open and linear structure. The econometrician, on the other hand, may find that in his models a 6.4% growth rate produces a very different result from a 7.0% growth rate, and his client may care a great deal about that difference. To him the system dynamicists' cavalier attitude about precise data seems both irresponsible and unsettling. Furthermore, since his paradigm provides no acceptable way of finding model parameters without statistical data, he cannot imagine how a system dynamics model becomes quantified. Since the numbers are not obtained by legitimate statistical methods, they must be illegitimate, made up, suspect.

The structure-parameter split is also revealed in the complaint often voiced by econometricians that "system dynamicists deliberately design their models to generate the results they want". System dynamicists do habitually specify in advance the dynamic behavior they will regard as a first test of confidence in the model, and do operate with some knowledge about what kinds of structure will produce what kinds of behavior. However, the task of making a complex dynamic simulation model behave in any reasonable way is surprisingly difficult, especially with a closed structure, with a paradigm requiring every constant and variable to have a recognizable real-world meaning, and with a bias against including time-dependent driving functions. When one has worked with models like that, one begins to regard the

relatively sensitive econometric model as much easier to manipulate. A system dynamicist would answer the econometrician's complaint this way. "Give me an open system with five dummy variables and 40 exogenous driving functions and I could design my model to generate the results I want." Both complaint and countercomplaint miss the essential point--the two kinds of models are each subject to rigid constraints of different sorts and are sensitive in different ways. A scrupuous modeler in either field will feel too bound by the characteristics of the real system to engage in conscious manipulation of results, and an unscrupulous modeler in either field can get away with outrageous fiddling. Unfortunately, neither field is sufficently self-monitored or self-critical to reward honesty or eliminate fiddling.

After each type of modeler has worked on an inherently unsolvable problem in the other's view, and has gone about it with entirely the wrong emphasis, the misunderstanding becomes complete when the finished models are examined for validity. Each kind of model fails to meet the other's criteria of validity or utility. The econometrician had a hard enough time understanding where the system dynamicist's numbers came from. Now he must evaluate the result without a single R^2 or t-test or Durbin-Watson statistic to help him along. He will find it impossible to calculate any statistical summary indices, because there will be multiple covariances and co-linearities and no data for many of the model's variables. The system dynamicist, who considers summary statistics either deceptive or meaningless, looks for the intuitive reality of the individual causal relations and the total dynamic behavior of the econometric model. He finds linearities, driving exogenous variables, and worst of all, dummy variables which correspond to highest-order cheating in his paradigm. The few instances of feedback he finds will be predominately positive feedback, which he knows will carry the entire model to ridiculous extremes if forecasts are generated for more than a few years into the future.

Even sincere efforts to understand each other's evaluation techniques tend to produce classic cross-paradigm conversations such as the following one between a system dynamicist and two mathematical economists, all of whom seem to be trying very hard to communicate.

HOWARD: We are used to seeing in the sciences one curve labeled "predicted" and another labeled "observed". These curves allow us to make evaluations such as "This is good" or "This is not so good". Is there any reason in principle....why you cannot take actual sales, production, and inventory data, use your model to obtain "predicted sales, production, and inventory figures for the corresponding period, and make a comparison?

FORRESTER: Yes, there is a reason why you cannot....Suppose you take two models, absolutely identical in structure and parameters, but both having different noise components in their decision mechanisms. If you start these models from identical initial conditions and let them run, their behaviors will diverge so quickly that there is no way of predicting what will happen on a specific day. Yet, the two models will exhibit similar qualitative performance characteristics. They will both be stable or unstable, for example....Thus one must predict, not the particular event, not the shape of the particular time history, but one must predict the change in the performance characteristics: profitability, employment stability, and characteristics such as these. The test you suggest of comparing a particular time history with the output of a model is not a test that you can expect to use, although it is a test that many people have been attempting for many economic models.

HOWARD: But I think that you have to have some quantitative measure of how good your model is....How can we possibly criticize you when you say, "It has the same qualitative behavior"? We both look at the same simulated history, and I say it does not look at all like the real thing, and you say it does. You say that you cannot with your model duplicate the actual sales data because of the noise in the system. All you can do is get a signal that has the same characteristics as the actual data. I say that this statement has no content....How can we get a quantitative agreement on what constitutes the same characteristics?

FORRESTER: This is a very troublesome question in the abstract, and yet in the actual specific case it is not answered in the rigorous objective sense that you speak of; neither is it in any of our real-life activities. I think you are trying for something here that we do not have in other areas of human endeavor. We do not have it in medicine or law or engineering. You are trying for something here that is more nearly perfect, more objective than in fact we know how to do anywhere else. I do not disagree with the desirability of it. I say we do not have it and we are not ready for it. Where we seem to have it in certain of the statistical model tests, I believe it is misleading and on an essentially unsound foundation.

HOLT: It is interesting to contrast Professor Forrester's willingness in model formulation to quantify such unstructured concepts as "integrity" with his unwillingness in model testing to accept quantitative tests of the models. Even where quantitative data are available for such variables as employment fluctuations both from the company and from the model, he accepts qualitative judgements on similarity as perfectly adequate. (from Greenberger 1962)

Can these two apparently antithetical ways of looking at and modeling social systems coexist within the mind of a single person? Can they coexist within the modeling profession? Or is it necessary, as Kuhn implies, that one paradigm must come to dominate the other totally?

Some people maintain that system dynamics and econometrics can indeed be merged within one person's mind and that in practice such mergers are appearing. System dynamicists can certainly be found who use statistical techniques to determine model parameters, and econometric models increasingly seem to contain state variables, distributed lags, and feedback. But these examples are just borrowings from each other's techniques, not shifts in world view. If the problem addressed by a computer model reflects an open system, basically static concern with particular responses to particular events, if the model variables are observables, if the validation procedure involves detailed matching with historic data, then I would say the model is in the econometric paradigm, no matter what mathematical techniques are used. If the problem is centered on generic dynamic behavior of a mostly-closed system, if the variables include motivations and goals, if the validation includes assessment of the realism of the model structure, then it is a system dynamics model. I cannot imagine how the two basic <u>philosophies</u> can be mixed or merged in one model, although the tools that have shaped and been shaped by those philosophies might be exchanged. Perhaps, however, using the tools of a paradigm can lead to a gradual, subconscious absorption of the paradigm itself.

On the level of the modeling profession as a whole, the outcome of the econometrics-system dynamics competition may be similar to the pattern of competition between species in an ecosystem. According to the competitive exclusion principle, two species struggling for the same ecological niche cannot coexist for long. One must eventually eliminate the other completely, as Kuhn says one competing scientific paradigm eventually eliminates the other from legitimate professional practice. However, when there are <u>diverse niches</u> available, it is entirely possible for one species to lose to the other in one kind of niche and dominate the other in a different kind. Econometrics and system dynamics clearly fit different niches in the modeling policy-making environment. As long as both short-term predictions and long-term perspectives are needed, these two techniques can both be actively pursued, probably with continued mutual hostility, at least until a better competitor comes along.

<u>Conclusion</u>

A comparative summary of some basic characteristics of the four schools of modeling, as I understand them, is shown in Figure III. Each school defines a particular way of looking at the world and provides a set of tools for working on particular kinds of problems. None is comprehensive enough to encompass all that might be observed about the world or to solve all problems. And of course very many observations and problems fall far outside the range of any of these modeling methods and outside the entire field of computer modeling. Modeling can certainly contribute greatly to human comprehension and control of complex systems. But like any other tool, it must be used with wisdom and skill, and that means with understanding of its appropriate uses, of its limitations, and of the way it influences its users' perceptions of the world.

FIGURE III: COMPARATIVE SUMMARY OF MODELING PARADIGMS

	System Dynamics	Econometrics	Input-Output	Optimization
usual purpose	to improve intuitive understanding of dynamic system properties; to make qualitative long-term forecasts	to predict precise future values of economic variables; to test economic theory	to trace effects of policy or prospective changes through a network of interlocking flows	to choose the optimal policy from a set of alternatives
typical time horizon	long	short	short	short
conceptual format or theoretical base	feedback loops, control theory	economic theory	interindustry accounting	objective function, activities, constraints
information base	broad and interdisciplinary	economic and social statistics	economic statistics, financial accounts	engineering, economics, or managerial data
degree of detail	aggregate	aggregate	disaggregate	usually disaggregate
mathematical restrictions	usually not precise	linear or loglinear no covariance of variables or error terms	linear, continuous	usually linear
other typical characteristics	dynamic descriptive largely closed	dynamic or static descriptive largely open	static descriptive open	dynamic or static normative open
sensitivity to errors or parameter changes	low	high	moderate	very high
typical number of equations	10-400	1-150	200-20,000	50-100,000
typical cost per computer run	50¢	estimation $10 simulation 50¢	$20-$200	$50-$15,000

BIBLIOGRAPHY

Almon, Clopper 1966, The American Economy to 1975, Harper & Row, New York.

Balinsky, Warren L. 1976, "Educational Models for Manpower Development", Technological Forecasting and Social Change, 8.

Barr. T.N. and Hazen F. Gale 1973, "A Quarterly Forecasting Model for the Consumer Price Index for Food", Agricultural Economics Research 25.

Bruno, Michall 1966, "A Programming Model for Israel" in Irene Adelman & Erik Thorbecke (eds.), The Theory and Design of Economic Development, Johns Hopkins Press, Baltimore.

Budzik, Philip 1975, "The Future of Dairy Farming in Vermont", Master's Thesis, Thayer School, Dartmouth College.

Chenery, H.B. and Arthur MacEwan 1966, "Optimal Patterns of Growth and Aid, The Case of Pakistan", Pakistan Development Review, Vol. VI #2.

Cramer, J.S. 1971, Empirical Econometrics, North Holland Publishing Company, Amsterdam.

Forrester, Jay W. 1971, World Dynamics, Wright-Allen Press, Cambridge, Mass.

Forrester, Jay W. 1969, Urban Dynamics, MIT Press, Cambridge.

Forrester, Jay W. 1961, Industrial Dynamics, MIT Press.

Goreaux, L.M. and L.S. Manne 1973, Multi-Level Planning: Case Studies in Mexico, North Holland Publishing Co., Amsterdam.

Greenberger, Martin, ed. 1962, Management and the Computer of the Future, MIT Press, Cambridge, Mass.

Hamilton, H.R., et.al. 1969, Systems Simulation for Regional Analysis, MIT Press, Cambridge, Mass.

Herendeen, R.A. & Clark Bullard, III 1974, "Energy Cost of Goods and Services, 1962 and 1967", Center for Advanced Computation, University of Illinois, Urbana 61801, Document #140.

Hoffman, Robert G. 1973, "Wheat-Regional Supply Analysis", The Wheat Situation Economic Research Service, U.S. Department of Agriculture, WS-225.

Hudson, Edward A. and Dale W. Jorgenson 1974, "U.S. Economic Growth: 1973-2000" in Long-Term Projections of the U.S. Economy, Data Resources, Inc., Lexington, Mass.

Intriligator, Michael D. 1972, "Econometrics and Economic Forecasting" in J. Morley English, ed., Economics of Engineering and Social Systems Wiley-Interscience, New York.

Kendrick, David A. 1972, "Systems Problems in Economic Development" in J. Morley English, Economics of Engineering and Social Systems, Wiley-Interscience, New York.

Kuhn, Thomas S. 1970, The Structure of Scientific Revolutions, University of Chicago Press, Second Edition.

Leontief, Wassily 1970, "Environmental Repercussions and the Economic Structure: An Input-Output Approach", Review of Economic Statistics, 52.

Leontief, Wassily, et.al. 1966, "The Economic Impact--Industrial and Regional--of an Arms Cut" in Input-Output Economics, Oxford University Press, New York.

Leontief, Wassily, 1936, "Quantitative Input-Output Relations in The Economic System of the United States", The Review of Economics and Statistics, 18.

Lofting, E.M. and P.H. McGauhey 1963, "Economic Evaluation of Water, Part III, An Interindustry Analysis of the California Water Economy", Contribution #67, Water Resources Center, University of California, Berkeley.

Maslow, Abraham 1966, The Psychology of Science: A Reconnaissance, Chicago: Henry Regnery.

Meadows, D.L. 1970, The Dynamics of Commodity Production Cycles, Wright-Allen Press, Cambridge, Mass.

Modigliani, Franco, Robert Rasche, and J. Phillip Cooper 1970, "Central Bank Policy, The Money Supply, and the Short-Term Rate of Interest", Journal of Money, Credit, and BAnking, Vol. 2; 1973, Board of Governors of the Federal Reserve System, Equations in the MIT-PENN-SSRC Econometric MOdel of the United States, Washington, D.C.

Myrdal, Gunnar 1968, Asian Drama, Pantheon, New York.

Naill, R.F., et.al. 1975, "The Transition to Coal", Technology Review, 78.

Nordhaus, William D. 1973, "The Allocation of Energy Resources" in Arthur Brooking's Paper on Economic Activity, M. Okum and George L. Perry (eds.), Vol. 3, Washington, D.C.: The Brookings Institution.

Oerlemans, T.W., et.al. 1972, "Dynamic Optimization of World 2", Project Globale Dynamics, Technische Hogeschool Eindhoven, Eindhoven, Netherlands.

Peck, Graham 1967, Two Kinds of Time, Houghton Mifflin Company, Boston, Second Edition.

Richmond, Barry 1976, "Conceptual Monograph No. 2", M.I.T. System Dynamics Working Paper.

Schantzis, S.G. and W.W. Behrens 1973, "Population Control Mechanisms in a Primitive Agricultural Society", in Meadows and Meadows (eds.) Toward Global Equilibrium, Wright-Allen Press, Cambridge, Mass.

Schumacher, E.F. 1973, Small is Beautiful, Harper & Row, New York.

Spann, Robert M. and Edward W. Erickson 1973, "Joint Costs and Separability in Oil and Gas Exploration" in Milton F. Searl, ed., Energy Modeling, Washington, D.C., Resources for the Future.

Tinbergen, Jan 1937, An Econometric Approach to Business Cycle Problems, Herman et Cie, Paris.