

PRINCIPLES ON THE RELATIONSHIP BETWEEN STRUCTURE  
AND BEHAVIOR OF DYNAMIC SYSTEMS

by

Alan Karl Graham

B.S., B.S., Massachusetts Institute of Technology  
(1973)

S.M., Massachusetts Institute of Technology  
(1973)

SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE  
DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August, 1977

Signature redacted

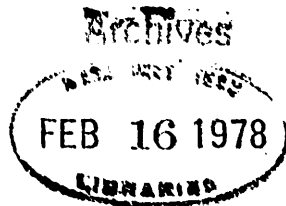
Signature of Author.....  
Department of Electrical Engineering and  
Computer Science, July, 1977

Signature redacted

Certified by.....  
Thesis Supervisor

Signature redacted

Accepted by.....  
Chairman, Department Committee





PRINCIPLES ON THE RELATIONSHIP BETWEEN STRUCTURES  
AND BEHAVIOR OF DYNAMIC SYSTEMS

by

ALAN KARL GRAHAM

Submitted to the Department of Electrical Engineering and Computer Science  
August, 1977, in partial fulfillment of the requirements  
for the Degree of Doctor of Philosophy

ABSTRACT

A large gap exists in the body of knowledge about the relationship of system structure and system behavior. The gap exists between formal mathematical methods and the conceptual or experiential tools that interact to give a "gut feel" for the behavior of a system. The traditional means of narrowing this gap is to work with feedback systems for several years until the analyst acquires a trained intuition. This thesis gives another means of narrowing the gap between formal mathematics and intuition: clearly enunciating explicit principles or rules of thumb about the relationship between a system's structure and its behavior. Conceptual knowledge of that relationship allows one to begin from a behavior mode and identify the kinds of structures that could have generated it. Thus, in formulating and restructuring a model of a social or economic system, such principles can provide a criterion for selecting the known cause-and-effect relationships which are to be explicitly included in the model. Also, the concepts with which such principles are expressed provides a means of explicating and communicating further experiences with dynamic systems. Finally, in testing a model of a social or economic system, principles on the relationship between structure and behavior can provide a basis for identifying structural changes that can render the model's behavior more realistic.

As an example of using a principle, suppose one is testing a model whose natural period is unrealistically long, and whose oscillations seem too heavily-damped. What sorts of structural additions or modifications might render the behavior (and structure) more realistic? The analyst here is searching for structural changes that shorten the period and decrease the damping. The principle in Section 3.3. states that "If a minor loop with a delay is added around a level already on an oscillatory loop, the added loop forms another pathway through which disturbances in the level can propagate back to the level. When the

additional disturbance returns to the level, it moves the level more rapidly to and past its steady-state value, which results in a shorter period and less damping." In this case, the principle in Section 3.3 directly identifies one type of structural change that can render the model behavior more realistic. One can then search within the cause-and-effect relationships of the real system to (possibly) find such a structure (in this case, a minor negative loop with a delay).

The principles in this thesis are empirical and conceptual rather than theoretical and mathematical: the principles explain what has usually happened in real applications, rather than describing what must happen of mathematical necessity. (Even so, Chapter 5 validates an empirical and conceptual principle with mathematically-derived results.)

The principles explicated in this thesis cover a diversity of structure-behavior relationships, although the focus is on oscillatory systems and the effect of structural changes on oscillations. The topics of the principles include: the origin of oscillations, damping, reduction to an effectively-first-order system, adding cross-links, adding a minor negative loop with a delay, adding a positive loop, adding a minor positive loop with a delay, and examining the entrainment of two similar but not identical systems subjected to a common exogenous random input.

DEDICATION

*This thesis is dedicated to my parents, who have supported my education from the very beginning.*

## ACKNOWLEDGEMENTS

This thesis marks the end of a period in my life, so in one sense, everyone involved in my life up until now has made some form of contribution to me and therefore to my thesis. Let me start with the most immediate people, and work my way back.

My work is my own, and at the same time, my colleagues create an environment that nurtures my work. Jay Forrester has created a discipline in which creation is the rule rather than the exception. Day-to-day contact with numerous people around M.I.T. has created the epistemological and factual body of knowledge from which my thesis is derived. These people include, in alphabetical order, Louis Alfeld, David Andersen, Professor Timothy Johnson, Professor Gilbert Low, Professor Nathaniel Mass, Barry Richmond, Professor Dale Runge, Professor Fred Schweppe, and Peter Senge.

Writing a thesis is a rite of passage, marking the transition from a sheltered learning environment to a profession. Of course I was in a learning environment before I came to the System Dynamics Group, and even before I came to M.I.T. My parents have provided me with a learning environment from birth onward, and for that I acknowledge them. Having to take a parent's role with a child of my own, I can state unequivocally that my own parents worked very hard to bring me up, and only now when it's all over can I realize that.

CONTENTS

CHAPTER ONE: INTRODUCTION	11
1.1. Purpose, Usefulness, and Audience of Thesis	13
1.2. Organization of Thesis	20
1.3. Needs for a Conceptual Understanding of Dynamic Models	22
The System Dynamics National Model	22
Formulating and Testing the National Model	24
Useful Answers to "What Causes the System Behavior?"	27
My Difficulties in Analyzing Entrainment	30
1.4. Comments on the Nature of Conceptual Understanding	36
Normal Intuition: Implicit Use of Experiences	37
Conceptual Understanding: A Descriptive Language for Less Implicit Use of Experiences	38
Insight	41
Explanation and Description	44
Communication of Concepts and Insights	46
CHAPTER TWO: OSCILLATIONS	49
2.1. Origin of Oscillations	51
An Example: A Spring-Mass System	51
Explanation in Terms of Equation Solutions	57
Explanation in Terms of Departures from Equilibrium Values	62
Explanation in State Space	67
Oscillation from a Purely Positive Loop	76
Damped Oscillation in a Managerial System	81
A Principle	94
Further Examples: Wire Memory and Glucose Regulation	100
2.2. Damping	106
An Example: A Damped Spring-Mass System	106
Explanation in Terms of Diminution of Disturbances	109
Explanation in State Space	110
A Principle	114
Another Example: Influence of Backlog on Ordering	117
A Further Example: Rabbit-Coyote System	124

CHAPTER THREE: FURTHER PRINCIPLES ON OSCILLATIONS	132
3.1. Reduction to a First-Order System	134
Characterizing the Effect of Parameter Changes:	
Time Constants	134
A Principle	147
Another Application: Orders in a Pipeline	151
Further Examples: The Spring-Mass System and Time to Correct Backlog for Orders TCBO	155
3.2. Cross-Links between Subsystems	161
A Descriptive Example: Vehicle Control	162
Another Descriptive Example:	
A Production-Distribution Chain	165
A Principle	167
An Optimal Control Viewpoint	174
3.3. Adding a Minor Negative Loop with a Delay	181
An Example: Influence of Availability on Ordering	181
Explanation in Terms of Phase Shift	189
A Principle	198
CHAPTER FOUR: POSITIVE LOOPS	205
4.1. Positive Loops Changing Response Time	207
An Example: Saving	207
Another Example: A Driven Spring-Mass System	213
A Further Example: Price Maintained by Tradition	216
A Principle	221
Further Examples: Cold Hands and The Flu	226
4.2. Adding a Minor Positive Loop with a Delay	229
An Example: The Influence of Backlog on Ordering	229
A Principle	238
Symmetrical Effect of Positive and Negative Loops	240
4.3. Use of Principles When They Don't Hold	248
An Example: Time to Average Output TAO	248
Another Example: Changing TCB and TAO	253



CHAPTER FIVE: SEPARATE SYSTEMS WITH A COMMON RANDOM INPUT	260
5.1. A Heuristic Development of the Principle	262
The Configuration and Its Significance	262
An Example in the Employment-Backlog System	265
A Heuristic Analysis	270
A Principle	272
Other Examples. Menstruation and Squirrels	274
5.2. A Mathematical Development of the Principle	276
Mathematical Approach	276
General Linear Case	277
Equations for a Simple System	282
Numerical Results for the Simple System	288
A Restatement of the Principle	307
Observations on Mathematical Techniques	309
CHAPTER SIX: CONCLUSION	314
6.1. Summary of Concepts Introduced	316
6.2. Summary of Principles Developed	320
6.3. Consequences of Thesis	323
Fulfilling the Purpose: Value to the Reader	323
Avenues for Further Development of Principles	324
Role in Curriculum Development	332

APPENDIX A: DOCUMENTED EQUATIONS FOR THE SPRING-MASS SYSTEM	336
APPENDIX B: ANALYZER LISTING FOR THE SPRING-MASS SYSTEM	337
APPENDIX C: EQUATION LISTING FOR THE SPRING-MASS SYSTEM	338
APPENDIX D: DOCUMENTED EQUATIONS FOR THE EMPLOYMENT-BACKLOG SYSTEM	339
APPENDIX E: ANALYZER LISTING FOR THE EMPLOYMENT-BACKLOG SYSTEM	344
APPENDIX F: EQUATION LISTING FOR THE EMPLOYMENT-BACKLOG SYSTEM	347
APPENDIX G: EQUATION LISTING FOR NOFPER	351
APPENDIX H: EQUATION LISTING FOR NOFTIME	354
APPENDIX I: DEFINITIONS OF VARIABLES IN CALCULATIONS OF N	357
BIOGRAPHICAL NOTE	360

CHAPTER ONE:

INTRODUCTION

Chapter 1 provides the context and motivation for the remaining chapters. It reviews the process of formulating and testing system dynamics models, and shows how and why experience is needed to perform these tasks. Chapter 1 examines experience, conceptualization, insight, and communication of experiences in some detail. This discussion motivates Chapters 2 through 6, which attempt to effectively communicate a specific set of useful experiences of how system structure determines system behavior.

CONTENTS

1.1.	Purpose, Usefulness, and Audience of Thesis	13
1.2.	Organization of Thesis	20
1.3.	Needs for a Conceptual Understanding of Dynamic Models	22
	The System Dynamics National Model	22
	Formulating and Testing the National Model	24
	Useful Answers to "What Causes the System Behavior?"	27
	My Difficulties in Analyzing Entrainment	30
1.4.	Comments on the Nature of Conceptual Understanding	36
	Normal Intuition: Implicit Use of Experiences	37
	Conceptual Understanding: A Descriptive Language for Less Implicit Use of Experiences	38
	Insight	41
	Explanation and Description	44
	Communication of Concepts and Insights	46

1.1. PURPOSE, USEFULNESS, AND AUDIENCE OF THESIS

The purpose of this thesis is to effectively communicate to the audience my present experiences with oscillation and entrainment, which are embodied in a series of principles.

The principles in this thesis provide conceptual summaries of empirical relationships between the structure and the behavior of a number of dynamic systems. These principles grow out of the author's experiences with social and economic systems, whose oscillations tend to be moderately damped.\*1\* For the most part, nonlinearities in the structures of these systems have not played a major role in determining their behavior, so the rules of thumb given in this thesis do not focus on oscillatory phenomena uniquely generated by nonlinearities. The behavior characterized in this thesis is usually a step, impulse, or initial-condition response of a model, as these responses have been the forms of analysis during the formulation and testing of the models from which the principles are drawn.

---

\*1\* One can distinguish two types of principles that one can extract from one's experiences. The type of principle not dealt with in this thesis is principles pertaining to a particular field, discipline, or area of study. For example, one could develop principles of inventory management in the field of business, or principles of land-use policy in urban studies. These are principles about systems with shared, and rather specialized, structure. In contrast, the principles developed in this thesis apply to systems in a diversity of fields or disciplines. The object of study in this thesis is oscillations in general, regardless of whether they occur in a management, biological, social, or even an astronomical system.

In brief, the explication of principles in this thesis is intended to increase the ability of the reader to characterize a change in a system's behavior given a change in its structure, or vice-versa. The need for such characterizations can arise either during the formulation of a model, or during policy testing. The principles are concepts that embody both the experiences described in succeeding chapters, and the reader's own experiences. Thus, this thesis should have the effect of making a modeler seem more experienced, at least in situations where the principles apply. I expect this to have four desirable results:

- (1) In some situations, the reader of this thesis should be able to use a principle to arrive quickly at hypotheses about the structural features of a system that do or could cause a given behavior.

As an example of using a principle relating structure and behavior, I was once discussing with a friend the behavior of a model of corporate activity.\*2\* The model was a production sector in the System Dynamics National Model, which is described in the following section of this thesis. He pointed out a seemingly strange phenomenon concerning the price charged by the corporation for its output. If some disturbance raised this price above its long-term equilibrium value, the price returned smoothly to that value, but with a time constant of decline very much longer than any of the time constants of the negative loops that regulated price in that model! To make a long story short, I had

---

\*2\*The purpose and structure of the System Dynamics National Model are summarized in Forrester, J.W., N.J. Mass, and C.J. Ryan, "The System Dynamics National Model: Understanding Socio-Economic Behavior and Policy Alternatives," System Dynamics Group Working Paper D-2243-3, Alfred P. Sloan School of Management (Cambridge, Mass.: MIT, 1976).

already outlined Section 4.1 of this thesis, so I was able to immediately hypothesize the cause of this behavior, whereas my friend had taken a few days to arrive at the same conclusion. (As a result of that conversation, Section 4.1 now contains as an example the analysis of how price can decline much more slowly than any of the time-constant parameters of the system might seem to indicate.)

- (2) In other situations, the reader of this thesis will use the principles to form hypotheses and expectations about the relation between a system's structure and behavior. When those expectations are not met (and the principle breaks down), the modeler is alerted that there is something unusual in the model's structure and behavior.

The principles in this thesis have exceptions. They are presented in this thesis by virtue of being useful and often true. Even when the principles are violated, the principles provide tangibility and direction to the analysis: the analyst can begin by asking "what caused this principle to break down?" rather than "is there anything here I should know about?" Section 4.3 provides two specific examples of the utility of violated principles. In this respect, the analysis of model structure and behavior seems to follow a classic dictum for writing, which states that if the rules are going to be broken, one had better know the rules and why they are being broken.

- (3) The concepts within the principles add to the vocabulary with which systems can be described. If people share the same descriptive vocabulary (and the similar experiences that underlie the vocabulary), they can communicate with one another more effectively than without the descriptive vocabulary.

An incident occurred during the time this thesis was being written that clearly illustrates the usefulness of properly-choosen

concepts and principles in communicating about systems. I told a student of the curious result of Section 3.3, namely that when two loops, each independently oscillatory with periods of 4 and 12 years respectively, are coupled through a shared level, the composite system oscillates with a period around the period of the faster loop. The student asked if there were circumstances under which the slower loop could dominate the system behavior, and cause oscillations at closer to the longer period. I answered immediately with close to these words: "Yes. If you consider the dominant loop as the loop around which a disturbance can propagate most easily (and between you and me, we know that "most easily" has something to do with phase and gain), then it is natural that the faster loop, which oscillates with less damping than the slower loop, should propagate whatever disturbances were present in the system around it, and dominate the oscillatory behavior. If the slower-oscillating loop were markedly less damped than the faster loop, the disturbances would propagate the longest around it, and it should dominate the oscillatory behavior." At least three things about this incident are remarkable to me. First, because I considered the question in terms of properly-chosen concepts, it was easy for me to arrive at the spontaneous insight just described. Second, the compactness of the answer is remarkable, considering the complexity of the issue. Third, and most remarkable, the short explanation reproduced above communicated to the student a clear sense of how that third-order system behaved.



- (4) By utilizing principles, the reader of this thesis should become more aware of his or her own half-conscious rules of thumb that relate structure to behavior, to the extent that they can be explicated as principles. This both adds to the body of communicable experiences with systems (embodied in the principle), and creates the opportunity for still less-conscious conceptualizations of experiences to emerge.

My experience is that, although the present principles provide satisfactory answers to a number of questions, the present principles point the way toward the development of many, many more principles. Section 6.3 describes the main avenues of future investigation I have uncovered in the process of developing the present principle.

The operational test of whether or not the thesis achieves its purpose is to see whether or not a reader dealing with some model has the appropriate principle come to mind at the appropriate time. If a reader is able to call to mind the experiences represented by the principle at the appropriate moment, the experiences have been "effectively communicated." Note that this degree of internalization on the part of the reader sets a standard for quality of presentation much higher than the usual technical presentations. Most technical presentations can be called successful if the readers understand it well enough to judge the material correct. This thesis will be adequately presented only if readers can spontaneously bring the material to mind and apply it to their own work. Obviously, it is impossible for me, the author, to guarantee that my experiences will be effectively communicated to you, the reader. If it happens, it happens. I can say that at least the prerequisites for communication are present: numerous examples drawn from a variety of fields and disciplines, concepts and principles gradually and explicitly evolved from the examples, and continual relation of the principles to normal, everyday situations as

well as well-known System Dynamics models, all should serve to recreate my original experience of the principles within the reader -- that is, to communicate. (Section 1.4 discusses communication of experiences, concepts, and principles in detail.)

By now, the vocabulary used in the examples has probably begun to convey a sense of the background and motivations required to read this thesis profitably.

This thesis addresses an audience of people who want or need to increase their qualitative understanding of the oscillatory behavior of a model of a dynamic system, and who are also familiar with the basic tools, concepts, and examples of System Dynamics and differential equations.

More specifically, I assume that the reader is familiar with DYNAMO equations and flow diagrams, and concepts of modeling and system structure. Such material can be found in Industrial Dynamics, or in one or two introductory courses. I assume that the reader has seen (but not necessarily mastered) models of physical systems, managerial systems, predator-prey systems, and endocrine systems, so that the examples used in this thesis will not seem foreign or unusual. I also assume that the reader is familiar with elementary differential equations. (Section 5.2 requires more sophisticated mathematical preparation, but is not essential to understand any of the principles.)

Another, equivalent, way of characterizing the intended audience is anyone who is in the same frame of mind as the student who has taken a few System Dynamics courses and a smattering of control theory, sees that they are related, wants to relate them, and doesn't know how. That student will have perceived that there is an intuitive skill or art

in tying together system structure and behavior, and that concepts from control theory are helpful but not sufficient. Where does the student proceed from there? The conceptual principles in this thesis are aimed precisely at unifying and expanding the body of knowledge desired by that person.

Note that this thesis is not intended as a piece of curriculum material per se, but, if anything, as a basis for future curriculum materials. The presentation of the principles assumes that the reader is familiar with state space, diverse applications of System Dynamics, characteristic equations, phase and gain, and undoubtedly several other pieces of knowledge. In curriculum materials, each of those pieces of knowledge would need to be introduced with its own text, examples, and exercises for a student to gain familiarity with it. One can consider this thesis as a guideline for selecting and organizing the pieces of knowledge within a curriculum aimed at teaching relationships between structure and behavior of dynamic systems. In addition to lacking introductory and preparatory material, this document also lacks exercises and tests to ensure that the reader understands the material. So, although reading this thesis may be educational, the thesis itself does not constitute curriculum materials, but a possible basis for future curriculum materials.

## 1.2. ORGANIZATION OF THESIS

The preceding commentary specifies in a very brief and abstract way what conceptual principles on the relationship between structure and behavior are, and why they are useful. The remaining two sections of this introductory chapter provide more detailed description of the nature and need for such principles. The remaining chapters explicate the principles themselves.

The principles each deal with some aspect of oscillation, but each principle does so in a different way. Chapter 2 develops two principles that establish a conceptual basis for explaining oscillations. Chapters 3 and 4 develop corollary principles based on the concepts of Chapter 2, that describe the effects of various types of structural and parametric changes on oscillatory behavior. Chapter 5 develops only one principle for a methodologically significant configuration: similar oscillatory systems with similar random inputs. Chapter 6 concludes the thesis with a number of suggestions for further principles linking system structure to system behavior.

The true building block of this thesis is the presentation of each principle, one in each succeeding section. As Section 1.4 will discuss, the section presenting each principle contains considerably more material than a statement of the principle and one example. Each section begins with a simple example of the behavior to be explained. The section takes several different viewpoints on the behavior--several different ways of explaining the behavior. For example, the section on the origin of oscillations explains oscillations in terms of differential

equations, conserved flows, and state-space trajectories. For each explanation, each section examines the shortcomings of the explanation (the questions upon which it sheds no light), and identifies the concepts from it that do seem valuable. The principle synthesizes these concepts, after which the section concludes by giving diverse examples of applications of the principle, and discussing implications of the principle for other situations.

Having been forewarned of the volume of material for each principle, the reader should expect to be covering more material than may at first seem necessary. The reader's approach to reading the remaining chapters should probably not be the same as he or she would use to read a technical report. Instead, think perhaps of an autobiography, where the reader stores away interesting and possibly relevant experiences for later use. These numerous experiences are only summarized and made accessible by the principles themselves.

### 1.3. NEEDS FOR A CONCEPTUAL UNDERSTANDING OF DYNAMIC MODELS

This thesis presents principles (conceptual rules of thumb) that are intended to be used in the process of modeling social and economic systems. Since such modeling has proceeded for years virtually without such principles (or at least such principles explicitly recognized as such), one might well ask, "Are they necessary?" or "What can principles do that ordinary mathematics or intuition can't do just as well?" In order to provide tangible answers to these questions, this section discusses an example in which conceptual explanations of the relationship between structure and behavior assumes paramount importance, and to which the principles in this thesis can make a significant contribution.

The System Dynamics National Model. The System Dynamics Group at MIT is constructing a simulation model of the U.S. economy. The System Dynamics National Model addresses a broad range of economic and social issues: inflation, economic fluctuations (at periods ranging from four to fifty years), the nature of economic growth, tax policy, energy-resource management, and the long-term prospects for the agriculture and education sectors of the economy. The National Model is divided into seven sectors, which are being developed separately. Broadly, these sectors represent production, labor, consumption, finance, government, demographic changes, and international trade. The production sector is a generic model of corporate activities, including production, shipping, hiring, inventory management, capital investment, accounting, payments, price-setting, and financing. It will be replicated once for each producing segment being represented

in the economy. Parameter values will be set to reflect the different characteristics of each individual production sector. (For example, capital goods, durable and non-durable manufacturing, or service sectors are characterized by different pricing policy characteristics, different requirements for factors of production, and different probabilities of individual business failure. Each of these characteristics can be represented by specifying the value of one or more parameters within the production sector.) Similarly, the labor and household sectors will be replicated once for each socio-economic group explicitly represented in the Model.

The National Model is being constructed by first constructing, testing, and revising individual sectors, and then testing and revising larger and larger agglomerations of sectors.\*3\* Testing and revising the structure of the National Model requires considerable time and thought, because the formulations must give realistic responses both for short-term transient conditions, and for the long sweep of a two-hundred-year depiction of the development from an agricultural to an industrial economy. Each formulation must contribute to a number of distinct behavior modes: hyperinflation, "stagflation," depression, short-term labor-inventory cycles, long-term capital-adjustment cycles, and so on. Each formulation must respond correctly to a variety of

-----

\*3\*The current status of the National Model is described in "The System Dynamics National Project Annual Report 1976," System Dynamics Group Working Paper D-2453-4, Alfred P. Sloan School of Management (Cambridge, Mass.: MIT, 1976).

extreme conditions: depression, war, resource shortage, hyperinflation, and the other vicissitudes that plague real economies.

Formulating and Testing the National Model. One possesses two kinds of information at the beginning of model formulation: the behavior one wishes to explain or control, and a vast number of real cause-and-effect relations. For example, behavior addressed by the National Model was described briefly above. With the National Model, as in many other models, one also begins with much more knowledge about real cause-and-effect relationships than can possibly be put into a model. All of one's knowledge about work, how to do things, and all of one's observations about how things influence other things, fall into the category of known real cause-and-effect relationships.\*4\* Therefore, the problem of formulating a model is really the process of selecting the real cause-and-effect relationships associated with the behavior mode being explained and controlled, and then embodying those cause-and-effect relationships in realistic equations.

-----

\*4\*On an epistemological level, it is quite difficult to distinguish between "real" cause-and-effect relationships and correlations. Suffice it here to say that there seems to be an axis along which we can characterize relationships. At one extreme along that axis are the highly regular, multisensory correlations (which I am here calling "real cause-and-effect relationships") that we all experience as a part of our physical existence. At an intermediate point on that axis are models of those real cause-and-effect relationships, which abstract, quantify, and correlate the individual cause-and-effect events into mathematical relationships between two or more variables. At the other extreme along that axis are purely correlative relationships, where, if there exists any interpretation of cause and effect for the relationships, the interpretation is much less significant than the statistical fact of correlation.



What basis does one have for selecting the real cause-and-effect relationships that are to go into a model, prior to the formulation of any equations? All that the modeler has is intuition (nonconscious concepts relating structure and behavior) and principles (conscious concepts relating structure and behavior). Even if the modeler proposes to use numerical data, there must be some prior conceptual basis for determining which numerical data are appropriate.

Principles--explicit patterns relating concepts of structure to concepts of behavior--can streamline the process of selecting the real cause-and-effect relationships to be represented explicitly in a model. For example, suppose one is seeking to analyze the three- to seven-year business cycle. The principle in Section 2.1 of this thesis suggests that one should look for loop structures to explain cyclic behavior. Principle 10.3-1 in Principles of Systems suggests that the coupling time constants appropriate for generating a three- to seven-year fluctuation are  $1/2\pi$  of those periods, or on the order of two to five months.\*5\* These principles provide considerable focus to the search for cause-and-effect relationships that can produce the observed behavior. Even when these principles fail to identify structures that fluctuate with the observed period, other principles can suggest classes of structural changes that could move the period into a realistic range. The principles in Sections 3.3, 4.1, and 4.2 all identify such classes of structural changes that alter the period of a system's oscillation. Such conceptual principles, then, provide considerable guidance during model formulation.

---

\*5\*Forrester, J.W., Principles of Systems (Cambridge, Mass.: Wright-Allen Press, 1968, pg. 10-20).

Testing a model consists of carefully scrutinizing the behavior of each model variable, identifying the structural features that cause the behavior, and evaluating whether or not the behavior and the structure that causes it seem plausible. For example, suppose one is testing the labor sector of the National Model by subjecting it to a step increase in the demand for labor, and the step response of unemployment overshoots its equilibrium value before equilibrating. One might attribute the overshoot to a delay in the perception of wage conditions by the unemployed, using whatever intuition, principles, or algorithms one possesses to make that attribution. Only when one has attributed the behavioral features to structural causes can one begin to evaluate the model's realism and how the realism might be improved: For example, one can start by asking whether it seems possible for unemployment to remain as high as it did, given other variables in the labor market. One can ask whether the overshoot points toward missing cause-and-effect relationships, or an unrealistically long perception delay.\*6\* Thus, in both the processes of model formulation and testing, the modeler needs to know what structural features are causing the system behavior, and principles can aid in knowing that.

---

\*6\*There are numerous criteria for judging the realism or validity of model behavior too numerous to discuss individually here. For comments on criteria for non-statistical model testing, see Forrester, J.W., Industrial Dynamics (Cambridge, Mass.: MIT Press, 1961), pp. 122-129.

Useful Answers to "What Causes the System Behavior?" One legitimate but trivial answer is always possible: "The whole system structure causes the system to behave as it does." This answer, however, is not very useful. It is model-bound, in the sense that it is tied to exactly one particular way of representing the system. That answer cannot be transferred to similar models or similar systems. In particular, this answer does not give even loose indications of the parameters to which the dynamics are and are not sensitive (by whatever definition of sensitivity is appropriate to the task at hand). This answer provides no guidance during model development as to which relationships in the model can be altered to give more realistic behavior. This answer provides no beginning point for policy analysis other than trial and error.

Clearly, more useful answers are possible to "what causes the system behavior?" Consider what form the answer should take: The answer should be a general concept or concepts, transferrable to a family of systems, rather than just one system. The answer should provide a conceptual basis for policy design--a way of selecting the policies that are likely to improve the behavior of the system. The answer should help the modeler to understand the system by indicating why some elements of the system are important in producing the system behavior mode, and why other elements are relatively unimportant.

Urban Dynamics provides examples of general, conceptual principles linking system structure behavior.\*7\* An explanation of virtually all of the behavior modes and policy results can be based upon only three principles, which can be briefly characterized as follows:

First, if an urban area is more attractive for inward migration than its surrounding environment, people will move into the area, expand its population, and eventually cause a shortage of some resource of the area (jobs, housing, open space, transportation, for instance) and deter further population growth.

Second, the balance of land use between residential and commercial uses determines whether a shortage of housing or a shortage of jobs will be the ultimate deterrent to population growth.

Third, new construction within a growing urban area keeps the average age of houses and commercial and industrial buildings relatively low, so that a growing area has a relatively new housing stock, suitable for occupancy by upper- and middle-income groups, and high employment densities in commercial and industrial structures to provide adequate jobs. When the land area begins to be fully-occupied by structures, aging and obsolescence reduce the employment densities of the commercial and industrial structures, while at the same time aging and obsolescence increase the proportion of relatively aged housing, attractive only for lower-income groups. When limited land ends urban growth, aging and

---

\*7\*For those readers unfamiliar with Urban Dynamics, these concepts are explicated in Forrester, J.W., Urban Dynamics (Cambridge, Mass.: MIT Press, 1969), Chapter 7, and in Alfeld, L.E., and A.K. Graham, Introduction to Urban Dynamics (Cambridge, Mass.: Wright-Allen Press, 1976), Chapter 10.

obsolescence thus attract lower-income groups with the availability of lower-income housing, while aging and obsolescence at the same time reduce the number of jobs available for these groups. The three principles--land-use balance, attractiveness for migration, and aging and obsolescence--here linking urban cause-and-effect structure to urban behavior are conceptual devices (mental tools for considering urban problems), which do not appear as a part of the formal model structure. They abstract the myriad details of the system (such as a city) into the essential structural features that give rise to the behavior mode (which is itself an abstraction of many specific instances of behavior). Perhaps a principle of system behavior might best be thought of as similar to the final summary paragraph of a book: while it in isolation appears plausible, it in concert with all of the other materials allows the reader/modeler an effective mastery of the entire body of knowledge about structure and behavior.

Conceptual principles linking system behavior to system structure permeate the entire process of modeling. In model formulation, the modeler uses a conceptual understanding of the dynamic hypothesis to select the cause-and-effect relationships that are relevant to the problem being addressed. In model testing, the modeler uses a conceptual understanding of dynamics to scrutinize the model structure and the model behavior to

perceive the less-than-realistic formulations.\*8\* Similarly in policy design, the modeler uses a conceptual understanding of the system dynamics to select effective policy leverage points out of the plethora of possible policy changes. The final and perhaps most critical role for concepts and principles of system behavior occurs in the area of implementation. In most cases, a faith in the competence of the analyst/modeler is not sufficient grounds for implementing the recommended policy changes. The decision-maker must have a clear understanding which, independently of the particular mathematical structure used in the model, explains why and how the recommended policy changes will improve the behavior of the system. That understanding is communicated by the modeler in the form of a concept or principle of system behavior. Such principles, then, seem essential in effectively solving real problems. For a more specific set of examples, let us return to one aspect of the System Dynamics National Model.

My Difficulties in Analyzing Entrainment. At one point in the development of the System Dynamics National Model I undertook to create conceptual answers to a relatively simple question: What causes the multitudinous sectors of the economy to rise and fall so closely with one another over the three- to seven-year business cycle?

---

\*8\*A conceptual understanding is not the only tool available for analyzing model behavior; Section 5.2 discusses mathematical methods as an alternative to the qualitative, conceptual approach described in this section. In particular, Section 5.2 discusses the limitations which sometimes render mathematical approaches inefficient or inappropriate in treating the dynamics of socio-economic systems.

("Entrainment" is the general name given to such phenomena.) I proposed to answer the question by examining all of the channels through which two production sectors could entrain each other in a short-term business cycle.\*9\* I initiated a two-pronged attack on the problem. I began to analyze, in detail, the dynamics of a single production sector, and also immediately began simulating two sectors interacting with each other through one sector ordering and purchasing the other sector's output. Even after many months, however, neither activity yielded much insight into entrainment. The investigation of a single production sector expanded severalfold as the complexity of the structure and behavior of even one production sector manifested itself. At the same time, I was unable to formulate a satisfactory conceptual explanation for the behavior of two production sectors coupled in a chain of production (that is, one buying from the other). Although I had amassed a large number of observations of the behavior of the two sectors with various parameters and inputs, the data were not coalescing into any clear, easy-to-communicate description of entrainment behavior in a chain of production.

The research described above was not infertile, despite failing to shed light on the causes of entrainment in a chain of

---

\*9\*There seem to be only four channels: buying and selling to one another, a common source for a factor of production (such as labor, capital, or financing), a common source of demand for production (such as consumer demand), and common susceptibility to exogenous disturbances (weather, tax changes, wars, and so on). This research is described in more detail in Graham, Alan K., "Proposal for Thesis Research: Entrainment of Economic Sectors in Cyclic Behavior," System Dynamics Group Working Paper D-2256-3, Alfred P. Sloan School of Management (Cambridge, Mass.: MIT, 1976).

production. Some of the simulations had important implications for leading-indicator analysis.\*10\* More importantly for this thesis, I generalized what I had observed over the course of the research into a series of principles that related system structure to system behavior. For example, many of the principles predict changes in the fluctuations of an oscillatory system that result from activating positive or negative loops, with or without delays in them. I needed to develop the principles to consolidate what I had learned to date, and hopefully to allow me to analyze faster the behavior I did not as yet understand.

On examining the list of principles and the work remaining to be done on entrainment, it became obvious that I would need considerably more principles (embodying a great deal more experience into many more descriptive concepts) before I could offer an adequate explanation of entrainment dynamics for production sectors. In other words, I discovered that there are very few explicit rules of thumb, concepts, or images that describe the characteristics of oscillatory systems, in comparison to what I would need to describe entrainment. I recognize that the reader may not fully grasp what it means to not have enough concepts to fully describe something. Here is an example: If I had enough concepts about thinking and modeling, I would have been able to describe and communicate

-----

\*10\*In brief, I discovered another theoretical flaw in using leading indicators as predictors of system performance. A noise input in one sector in the chain of production causes it to lead the other sector. The apparent lead-lag relationship between the two sectors thus depends on if and when each sector happens to be disturbed by some random event.



my difficulties with entrainment concisely and clearly. But I don't, so I haven't. Perhaps the reader can gain some experience of what it is like to be both with, and without, adequate concepts by comparing the fragmented, incomplete portrayal of the urban problem given in the popular press to the compact, orderly description of the origin of urban problems given above.

There is a test for whether or not one has enough concepts to understand some system's behavior: Is one able to predict what the system will do in broad terms 50 percent of the time, without recourse to computation or mathematics? Note that this test requires one to attempt to understand the behavior before examining the behavior. Many people can claim to understand behavior after seeing it, but demonstrating an understanding without first knowing the behavior is much more difficult. The 50 percent mark is probably a very high standard. At any rate, I couldn't claim to understand the entrainment behavior anywhere close to that mark.\*11\*

Entrainment is not the only research area in the National Model Project in which adequate rules of thumb are missing. A capital production sector ordering capital from itself generates a fifty-year oscillation remarkably similar to the empirically-observed Kondratieff Cycle.\*12\*

---

\*11\*I have not, however, gone back to analyze entrainment behavior after having developed the principles reported in this thesis. My suspicion is that with the fully-developed concepts and principles in hand, entrainment behavior would be much more comprehensible.

\*12\*The structure and behavior of the fifty-year oscillation are described in more detail in both Forrester, J.W., "Business Structure, Economic Cycles, and National Policy," Futures, June 1976, pp. 195-214, and Mass, Nathaniel J., "Modeling Cycles in the National Economy," Technology Review, vol. 78, no. 5, 1976, pp. 1-12. For a discussion of the real implications of this Kondratieff wave, see "National Project Annual Report 1976," op. cit.

There is no well-developed body of knowledge that would allow one to accurately hypothesize which one of the hundreds of feedback loops in the production sectors are responsible for the fifty-year cycle, and why the other loops are not essential. How can one know how to alter the fifty-year cycle? This lack of a well-developed body of knowledge (that can render complex oscillatory systems reasonably intuitive) will impede the understanding of nearly every important behavior mode of the economy. Which structural features and policies produce inflation? What can be changed to retard or reverse inflation? What structural changes in the economy lead to energy independence? And even before those questions can be posed, how is one to find the causes of any unrealistic behavior that may arise during the formulation and testing of the System Dynamics National Model?

It seems natural that such a body of knowledge about complex oscillatory systems in general can exist; if not, it would imply that no oscillatory system has anything in common with any other oscillatory system. Even though very few generalizations about oscillatory systems are always true, many generalizations are true often enough to serve as useful guides. For example, there is an implicit principle among engineers that says increasing the gain of a feedback controller decreases the stability of the system. Even though this is not always true, the principle is quite useful to engineers (and public speakers using microphones and amplifiers, who reduce the gain to eliminate feedback "howl").

In brief, from my difficulties in analyzing entrainment of production sectors, I concluded that the body of qualitative knowledge describing oscillatory systems is far smaller than is necessary to give

any clear qualitative understanding or description of a complex socio-economic system such as the U.S. economy. This thesis attempts to expand that body of knowledge. Before beginning on that task, we must first examine the nature of such a body of knowledge in more detail.

#### 1.4. COMMENTS ON THE NATURE OF CONCEPTUAL UNDERSTANDING

The following material describes mental processes such as intuition, concepts, insight, explanation, and communication; although the material may seem excessively psychological, it seems appropriate to examine what is being conveyed (insights leading to a conceptual understanding) before conveying it in the succeeding chapters. Just as it would be foolhardy to design a tool without knowing the use to which the tool would be put, it would be less than wise to attempt to facilitate someone in generating insights without first examining the form and substance of insights.\*13\* The material is definitely introspective and "soft"; perhaps the best way to think of the following is not about whether or not the material is true in any scientific sense, but about whether or not the material provides a useful conceptual framework for considering the insightful use of principles linking system behavior to system structure.

-----

\*13\*The descriptions of psychological events concerning insights into the structure and behavior of feedback systems arise from my own experience with such systems, both in attempting to address the problem of entrainment and in working with previous systems. However, the way in which I describe my personal experience has been influenced by numerous persons and literature. Among the foremost contributors are personal contacts with Jay W. Forrester and Werner Erhard, and reading Koestler, A., The Act of Creation (New York: MacMillan, 1964) and Kuhn, T. W., The Structure of Scientific Revolutions (Chicago: University of Chicago Press, 1964).

Normal Intuition: Implicit Use of Experiences. "Normal intuition" in the present context denotes taking some action without a conscious reason for the action. For example, a control engineer may use "engineering judgement" to specify initial values for gains in a feedback controller. The initial guesses may in fact be quite close to the final values derived from repeated experimentation, even though the control engineer did not consciously utilize any algorithms or rules of thumb to obtain gains from the known structure of the plant being controlled. Clearly, there was some reasoning or some mental process which produced the initial guesses at gain values; mental processes were merely below the level of consciousness. This nonconscious reasoning is normally called "intuition." (I will use "nonconscious" rather than confusing anything with Freud's "unconscious.") The existence of intuition is widely acknowledged; educational institutions even seek to inculcate it by exposing the students again and again to engineering problems and solutions.\*14\* In this way, the student begins to amass a

-----

\*14\*This is not to imply that engineering is entirely a matter of intuition. The engineering literature consists principally of formal algorithms. However, even the selection of the appropriate formal algorithm is usually a process that is nonconscious, hence, intuitive. There is constant interplay between intuition and more explicit, formal methods. In this respect, engineering is little different from classical arts such as painting and sculpture. There are numerous explicit guidelines regarding form and design that an artist may consciously utilize, and yet much of the composition of a work of art is highly intuitive.

body of experiences that can be drawn on, consciously or not, to solve problems.\*15\*

Conceptual Understanding: A Descriptive Language for Less Implicit Use of Experiences. There is a spectrum of mental tools used in problem-solving. At one end is nonconscious intuition. At the other end of the spectrum are completely explicit formal algorithms that can be automated totally, such as finding the gain matrix for a linear quadratic Gaussian control problem. Between these two extremes, there are descriptive concepts. For example, control engineers have a concept that says, "Systems controlled by feedback can be made unstable if the feedback gain is too high, so if the system is unstable, try reducing the feedback gain." In general, the most obvious component of a conceptual understanding consists of conscious statements in a language that describes the specific situation in general terms. The general terms also describe similar past experiences, so that the conceptual understanding is an explicit way of connecting a current problem to past experiences with similar problems. ("I stabilized that other system by reducing the gain.") When a current problem is explicitly described in general conceptual terms, the mind automatically connects with other experiences that have been described with the same general terms. In

-----

\*15\*An "experience" is used in this thesis to denote a highly specific event: what happened at a particular place and time, and what one sees, hears, thinks, and feels at that particular time. Note that an experience can be a mental event such as a thought or insight, as well as a physical event.

contrast, intuitive understandings have no explicit and conscious device to connect the current problem with past experiences, so the connections may not be made as reliably.

If the relationship between a concept and the experiences that it describes were to be cast in geometrical terms, the concept would appear as the peak of a pyramid, supported by the far more numerous and specific experiences, images, and subsidiary concepts. Thus, when a new problem is described with a concept, the mind establishes a connection with part of the base of the pyramid and its peak (See Figure 1-1). Because the concept at the peak of the pyramid has been used already to describe other experiences, the peak of the pyramid is connected to other specific experiences relevant to the problem at hand. Thus, utilizing a conceptual description of a problem allows more efficient mobilization of the previous experiences necessary to solve the problem.\*16\* Concepts are therefore devices that allow one to access

-----

\*16\*The importance given here to mobilization of experiences as a means to arrive at an insight may strike some engineers as odd. For models of social or economic systems, the structure of the system (and even many of the parameter values) are well-known to the participants in the system. In contrast, in many engineering applications, the principal barrier to problem-solving seems to be acquiring sufficient data about the system structure and behavior to be able to reason out a solution. Even then, experiences must be mobilized and manipulated to give the engineer a concept of what needs to be known.

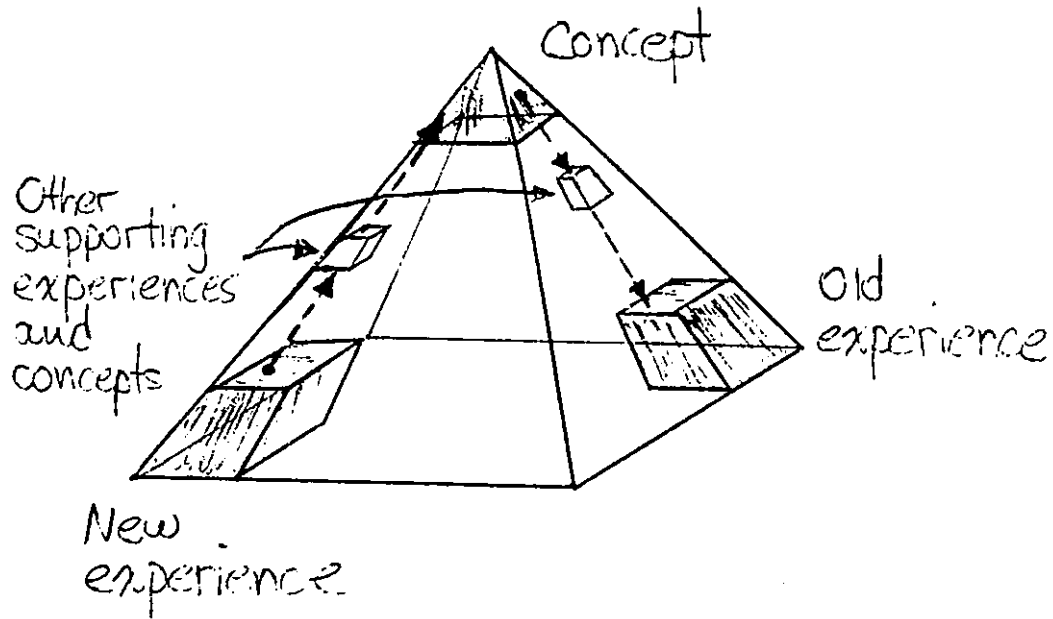


Figure 1-1.  
RELATIONSHIP AMONG A CONCEPT, AND OLD AND NEW EXPERIENCES

and manipulate detailed, tangible experiences in a way that produces new insights into the problem at hand.

Figure 1-1 shows a concept imposing an organization or pattern among experiences. Figure 1-1 also suggests that a concept can impose organization or pattern among other concepts. In fact, the principles developed in this thesis are such concepts, which show consistent relationships among concepts such as "oscillation," "loop," "state," "propagation," and so on. So principles are both a type of concept, and made up of concepts.



Insight. The aim of this thesis is to communicate principles (which are concepts) about relationships between system structure and system behavior. The principles are to be used to give a modeler insight into the structural causes of the behavior he or she must deal with. The introductory chapter has thus far examined the need for principles relating system structure to behavior in practical modeling situations, and the function of principles in such situations. With this background, it is now possible and appropriate to examine more closely the process of applying these particular kinds of concepts by generating an insight. "Insight" can be operationally defined as the reorganization and redescription of existing experiences (or concepts, for that matter) in a satisfactory manner. (An equivalent and more compact definition is changing concepts in a satisfactory manner. The next subsection characterizes the "satisfaction" involved in an insight.)

Experiences do not reside within the mind independent of one another. They are connected by our ideas of what is relevant to what--a conceptual organization. When an insight changes, this organization of experiences changes, so that things thought to be irrelevant turn out to be relevant, and things thought to be relevant turn out not to be. Figure 1-2 diagrams such a change in the relationships among experiences; the experiences are the same, but they assume a different pattern of importances. For example, when a child has the insight that its parents emphasize cleanliness to prevent the entry of germs into its body, many actions which the child had thought were unimportant, such as handling food with dirty hands, become important. In contrast, before the insight, a child is sometimes upset by fingerprints on the outside of glasses--the glass is "dirty." After the insight, such "dirt" is irrelevant to

cleanliness. As another example of how previously unrelated facts become related to the problem at hand, consider the principle in Section 2.1, which implies that in order to oscillate a system needs two subsystems with comparable time constants. This principle integrates several of my experiences with oscillations; for example, that increasing the time constant of an integral controller increases stability, or that either increasing or decreasing a parameter can stabilize a system.\*17\*

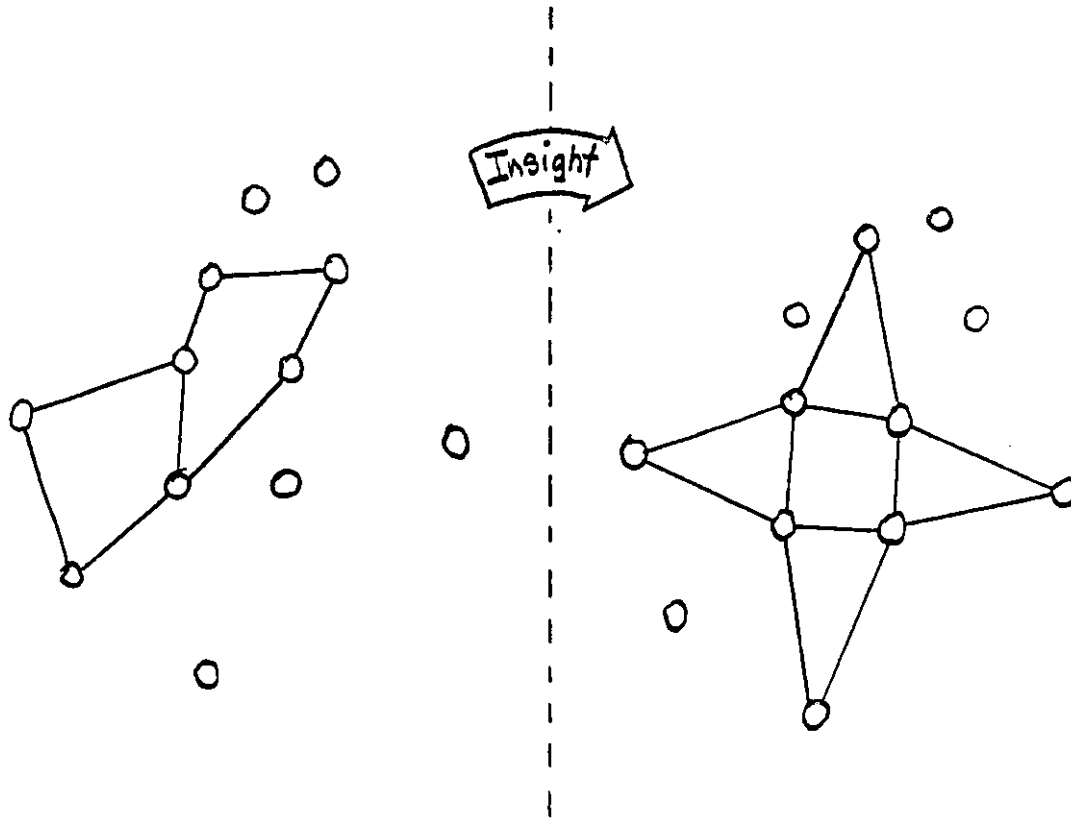


Figure 1-2.  
REORGANIZATION OF EXPERIENCES BY AN INSIGHT

---

\*17\*William Bode derived this result for feedback radio amplifiers in the 1930s, and Dale Runge discusses an economic example in "Understanding Simple System Behavior: An Examination of Job Vacancy - Employment Interactions," D-2244-1, Alfred P. Sloan School of Management, E40-253 (Cambridge, Mass.: MIT, 1976).

I have called the reorganization of experiences that occurs during an insight a redescription of experiences, to emphasize the relationship between a descriptive concept ("the system is unstable") and the specific experiences that the concept gathers together by implying some similarity between ("perhaps its feedback gain is too high, just like that other system"). Although an insight is just a change in concepts, calling it a redescription emphasizes the substance of an insight (experiences or concepts) in addition to its form (concepts). For example, the Phillips Curve is usually described as the curve that specifies the structural or causal relationship between unemployment and inflation, and which shifts due to unknown factors. Forrester redescribes the curve as describing only a correlative (behavioral) relationship between unemployment and inflation over a four-year business cycle, with shifts in average unemployment due to the long-term fifty-year Kondratieff Cycle, and shifts in average inflation due to monetary policy.\*18\* This redescription totally changes the range of policy issues that are relevant to unemployment. Long-term economic development planning becomes as germane to unemployment as next year's federal budget.

Another way of describing the reorganization that occurs during an insight is to claim that the insight creates a new context for considering the difficulties. Forrester's insight into the Phillips Curve clearly provides a substantially different context for considering difficulties with unemployment. In the case of many system dynamics models, the context is a clear and simple explanation of troublesome behavior. For example, I have given a reasonably simple explanation for

---

\*18\*Forrester, J. W., "Business Structure, Economic Cycles, and National Policy," Futures, June 1976, p. 209.

the onset of puberty in males.\*19\* That explanation provides a completely new context for designing hormone therapy for precocious or delayed puberty, by providing a description or image of how such therapeutic programs act on the internal hormone system. The same explanation of puberty also provides a simple framework that renders a substantial number of previously inexplicable phenomena logical and straightforward: why the age at which puberty occurs is correlated with body weight, or why precocious puberty sometimes occurs as a byproduct of other hormone therapy. The central insight creates a context for these previously obscure and irrelevant facts that renders them straightforward corollaries. Insights, then, can be thought of as creating a new context for already well-known content.

Explanation and Description. Most insights concern explaining something that previously was only described--knowing why something happens rather than the previous knowledge that something happens. An explanation characterizes a wider range of situations (including what happens when

---

\*19\*The theory is that simple bodily growth is the "biological clock" for puberty. Briefly, my theory suggests that sex hormone levels are usually in an equilibrium maintained by a nonlinear negative feedback system. Before puberty, sex hormones are maintained at low levels. Somatic growth causes blood volume to expand, so hormone secretions become progressively more diluted as the child grows. The dilution effect gradually shifts the nonlinear feedback system to a point where it is no longer able to stabilize hormones at the low prepuberal levels. Puberty eventually occurs when the feedback system, unable to maintain low hormone levels, allows the hormone levels to rise to normal adult equilibrium. (See Graham, A.K., "Feedback Processes Underlying the Onset of Puberty in Males," unpublished M.S. Thesis (Cambridge, Mass., MIT, 1973).

we change things), has fewer apparent exceptions or contraindications, and usually uses a smaller number of basic concepts than a description i.e. Newton's law of universal gravitation characterizes both the elliptical trajectories of planets and the parabolic trajectories of earth-bound projectiles. That law has very few observable exceptions, and uses a small number of basic concepts, such as mass, acceleration, momentum, and gravitational force. But what of later theories, such as Einstein's relativistic mechanics? Compared to it, Newtonian mechanics are merely a description, and sometimes an awkward one at that because, for example, Newtonian mechanics treats mass and energy as two separate entities.

Is there any way, then, of distinguishing an explanation from a description? They certainly have the same form: collections of concepts that tie together specific experiences and other concepts. The only distinction that can be made is relative: description A is more of an explanation than description B. Description A is more general (it explains many more things), has fewer apparent exceptions, and uses a smaller number of concepts (that is, it's simpler).\*20\* It is these characteristics that in some sense make the insight of a new explanation more satisfying. For instance, Chapter 2 begins by describing the behavior of an oscillatory system, then explaining the behavior through solution of differential equations. But that "explanation" is a description in comparison to later "explanations." Chapter 2 arrives at better and better

---

\*20\*There will be cases where one explanation is not obviously superior to another; two examples are Freudian psychology (versus behaviorist psychology) and monetary economics (versus what might be called Keynesian or fiscalist economics).

explanations, but of course does not arrive at an ultimate and unimprovable explanation.\*21\*

Communication of Concepts and Insights. Figure 1-1 shows how a concept (a description of a specific situation in general terms) can facilitate the mental connection between a new situation and the appropriate previous experiences. The concept serves as a handle by which previous experiences can be manipulated and aligned with new situations so as to resolve the difficulty under consideration. The figure implies that a concept is useful only insofar as it is mentally connected to previous experiences. Therefore, one cannot communicate a conceptual understanding or an insight (a new concept) by merely repeating the words that describe the concept or insight. For example, a child can be told hundreds of times not to handle food with dirty hands, but until the child connects these words with some concept or experience of germs and disease, the words have no operational meaning and no motivating power. Similarly, people can peruse page 110 of Urban Dynamics, and

-----

\*21\*There can be no ultimate explanation of something, because the substance of explanations is concepts, which always differ from the experiences or events being conceptualized. For example, consider explaining why an upward-thrown ball returns to earth. Is gravity why the ball returns? Does gravity cause an upward-thrown ball to return to earth? Yes and no. Yes, because we possess a thoroughly-tested set of concepts (physical laws) that describe cause-and-effect relationships in kinetics. No, because gravity is a concept: an abstraction, a mental event. It seems clear that a thrown ball will return to earth regardless of whether or not I have the concept of gravity in my head. So gravity, strictly speaking, cannot be said to be the ultimate cause of the ball's returning to earth, and thus, not an ultimate explanation.

read that "Complex systems are remarkably insensitive to changes in many of the system parameters (constants in the equation)." But again and again, people cannot accept that statement as a fact until they have experienced it for themselves.

The only way to communicate a concept or insight is to recreate in the listener one or more of the experiences that the concept embodies, either by creating a new experience for the listener, or utilizing the listener's previous experiences.\*22\* Sometimes, the experiences of the speaker and the experiences of the listener have so much in common that relatively few words are needed to communicate. "I am sad" can be understood and acted upon by nearly everyone. But we do not look upon our everyday experience as experience with complex feedback systems, so communicating concepts or insights about such systems requires considerably more explicit communication of the numerous experiences that underlie the simple verbal expression of a concept or insight.

This thesis aims at allowing the reader to apply principles (which is the insightful application of a concept) to new situations. One method of accomplishing this might be to simply describe the

-----

\*22\* Sometimes, a listener can hear a concept and imagine (create) an experience of it mentally. However, that experience is not vivid and real enough to use the concept to understand further concepts. So, for example, someone is able to accept the idea that systems are often insensitive to parameter variations, but is unable to accept a consequence of that idea: that one should not attempt to statistically estimate all of a model's parameters (especially not the insensitive ones).

principle and one application, and hope that the reader can recreate the experience embodied in the principle in the context of a new problem. This method seems to place an undue burden on the insight of the reader. Another method is to use several examples and allow the reader to actually experience the transfer of the principle from one situation to another. Thus the principle does not become so tightly attached to a specific example. This thesis must of needs utilize a variety of examples to communicate the principles.



## CHAPTER TWO:

### OSCILLATIONS

Oscillations permeate the behavior of virtually any system one examines. In the sphere of the natural sciences, elementary particles, molecules, endocrine glands, planets (and perhaps even the universe itself) fluctuate. In the sphere of the social sciences, psychological well-being, organizational effectiveness, corporate sales, social attitudes, and long-term economic development all show cyclic behavior. This chapter concerns itself with understanding the structural causes of such oscillations. The chapter gives principles that give a general explanation of the origin of oscillations and some factors that affect the damping. The next chapter then gives more principles on the effect of various structural changes on oscillatory behavior.

CONTENTS

2.1. Origin of Oscillations	51
An Example: A Spring-Mass System	51
Explanation in Terms of Equation Solutions	57
Explanation in Terms of Departures from Equilibrium Values	62
Explanation in State Space	67
Oscillation from a Purely Positive Loop	76
Damped Oscillation in a Managerial System	81
A Principle	94
Further Examples: Wire Memory and Glucose Regulation	100
2.2. Damping	106
An Example: A Damped Spring-Mass System	106
Explanation in Terms of Diminution of Disturbances	109
Explanation in State-Space	110
A Principle	114
Another Example: Influence of Backlog on Ordering	117
A Further Example: Rabbit-Coyote System	124

## 2.1. ORIGIN OF OSCILLATIONS

An Example: A Spring-Mass System. To begin a description of oscillatory behavior, I will start by describing a very simple oscillatory system. Then, instead of taking advantage of its simplicity, the subsequent text will attempt to describe the system structure and its oscillatory behavior in general terms. In that way, the description should apply to very complex oscillatory systems, as well as the simple two-level system to be described here. Figure 2-1 gives a diagram of the system under consideration: a mass sliding on a surface, attached to one end of a spring. The other end of the spring is fixed to some stationary object. Assume for the moment that friction is negligible. This system is representative of a large class of pendulums, whose other members include the gravity pendulums in clocks, swinging doors, and suspension systems for cars (which bounce up and down). Although the structure of these systems can be obtained directly from well-known physical laws, the explanation of their oscillations does not depend on where the modeler obtains the system structure. Later examples in this section describe biological and corporate systems whose structures are obviously not obtained from physical laws, but whose behavior is oscillatory, and arises from the same structural features that characterize the physical system in Figure 2-1.

Figure 2-2 gives a preview of the structure of the system, by showing a DYNAMO flow diagram. The description of the equations of motion for the system could begin anywhere on the loop shown in

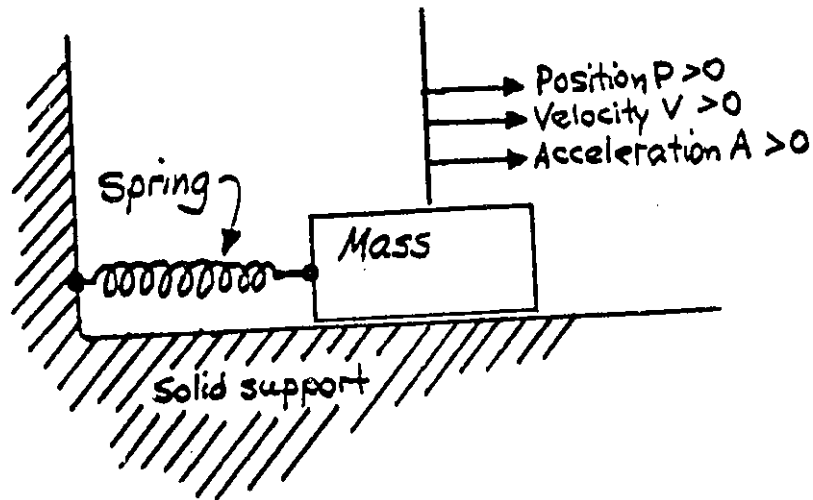


Figure 2-1.  
A SPRING-MASS SYSTEM

Figure 2-2; the position of the mass is a suitable beginning point. The position P of the mass is the accumulated result of the mass' movements, which is quantified as the velocity V. Thus, the position P is just the integral of velocity V. (This reverses the standard physics definition of velocity as the time-derivative of position.) The initial position has been set at half a foot to provide some disturbance to the system, so that it will not simply sit at rest.

$$P.K=P.J+(DT)(V.J)$$

$$P=IP$$

$$IP=.5$$

- P - POSITION (FEET)
- V - VELOCITY (FEET/SECOND)
- IP - INITIAL POSITION (FEET)

1, L

1.1, N

1.2, C

Newton's laws of motion say that an object in motion tends to remain in motion. Thus, the velocity V of the mass is a

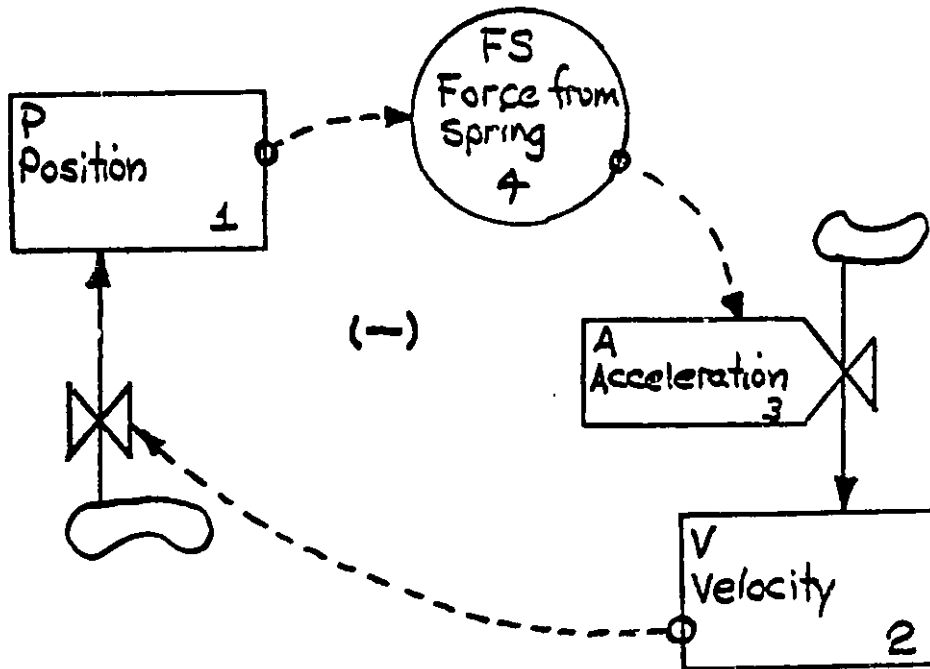


Figure 2-2.  
DYNAMO FLOW DIAGRAM OF SPRING-MASS SYSTEM

level: The velocity V accumulates acceleration A.

$$V.K = V.J + (DT)(A.JK)$$

$$V = IV$$

$$IV = 0$$

- V - VELOCITY (FEET/SECOND)
- A - ACCELERATION (FEET/SECOND/SECOND)
- IV - INITIAL VELOCITY (FEET/SECOND)

2, L  
2.1, N  
2.2, C

The acceleration A results from forces acting on the mass.

The modern formulation of Newton's laws gives the relationship as:

$$F = m*a$$

$$a = F/m$$

where a is the acceleration, m is the mass being accelerated, and F is the sum of the forces acting on the mass. But what is the mass? When we weigh something, that weight W is only the force exerted on the mass by gravitational acceleration G. Using Newton's law again,  $F = ma$  for gravitational force becomes

$$W = mG$$

or

$$m = W/G$$

The last formula allows us to write an equation for acceleration A in more familiar terms:

$$a = F/(W/G)$$

Force F is the sum of forces due to the spring, friction, and any exogenous forces exerted on the mass (the force from spring FS, the force from friction FF, and the force from driving FD). Of these, only the force from spring FS will be considered in this section. The other forces are set to zero in this section, and will be activated (i.e. made active or varying over time) and discussed in later sections. In DYNAMO, the equation for acceleration becomes:

$A.KL = (FS.K + FF.K + FD.K) / (W/G)$  3, R  
W=160 3.1, C  
G=32 3.2, C

A - ACCELERATION (FEET/SECOND/SECOND)  
FS - FORCE FROM SPRING (POUNDS)  
FF - FORCE FROM FRICTION (POUNDS)  
FD - FORCE FROM DRIVING (POUNDS)  
W - WEIGHT (POUNDS)  
G - GRAVITATIONAL ACCELERATION (FEET/SECOND/  
SECOND)

The force from the spring FS is simply proportional to the amount that the spring is lengthened or compressed (as measured by the position P); the further the spring is stretched, the more force it exerts to return to its original length. Because the force is exerted in the opposite direction from that in which the spring is moved, the proportionality is negative. (Historically, this formula is known as Hooke's Law.)

$FS.K = -SC * P.K$  4, A  
SC=5 4.1, C

FS - FORCE FROM SPRING (POUNDS)  
SC - SPRING CONSTANT (POUNDS/METER)  
P - POSITION (FEET)

The spring constant SC is a simple empirical parameter, characterizing the force exerted by the spring per foot of expansion or contraction of the spring. The value chosen for SC represents a spring about as powerful as a screen-door spring: if the spring is extended one foot, it exerts a force of 5 pounds in the opposite direction.

Figure 2-3 graphs a 10-second simulation of the spring-mass system described by the equations above. The behavior is clearly oscillatory. For the moment, let us concentrate on the most basic feature of this behavior: the same pattern of behavior of the variables

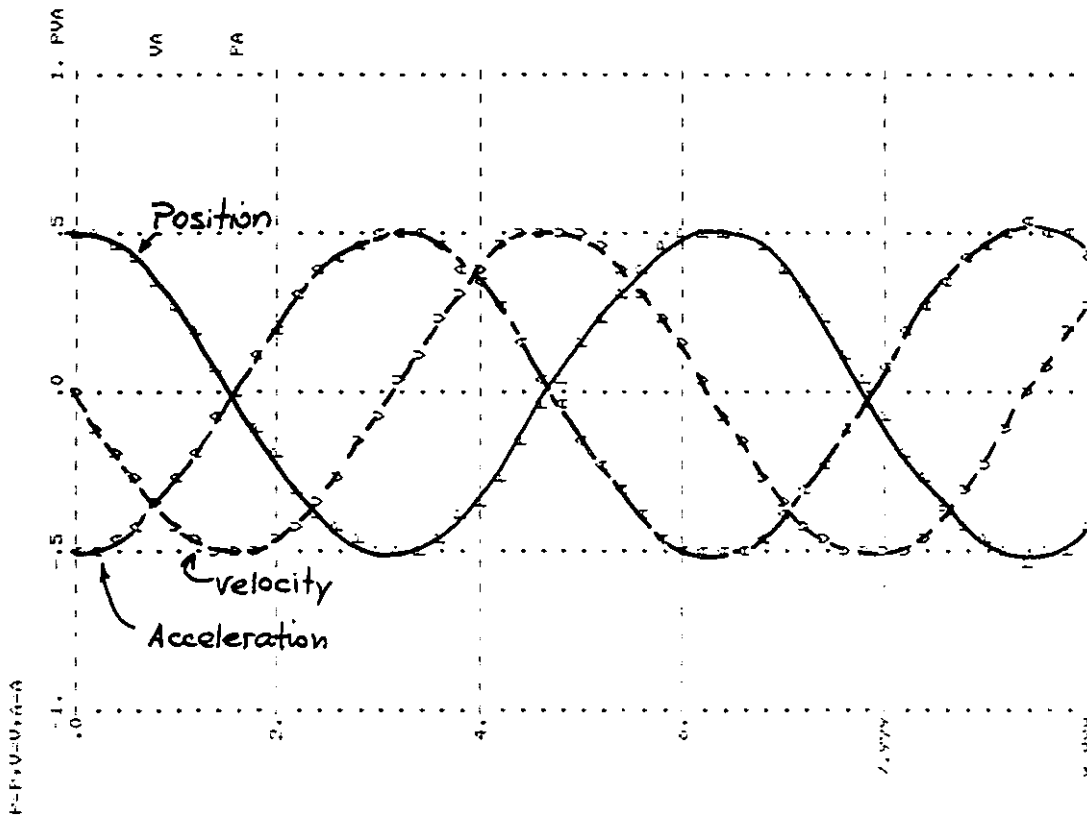


Figure 2-3.  
BEHAVIOR OF SPRING-MASS SYSTEM

being repeated at regular intervals. What is there about the system structure that causes this behavior? How can one characterize the fundamental features of the system which give rise to oscillatory behavior? The following subsections begin to work toward an answer to these questions.



Explanation in Terms of Equation Solutions. The classical answer to "Why does the system oscillate?" is to solve the differential equations that characterize the system, and then claim that the system oscillates "because" the eigenvalues of the system are imaginary. "Because" in this context is a very slippery concept; the eigenvalues are mathematical fabrications, and in fact they do not physically cause the oscillations. The cause of the system's oscillation is nothing more and nothing less than the system's structure. ("Structure" in the context of this thesis can denote all of the information necessary to generate the time-behavior being studied: equation form, parameters, initial conditions, and driving inputs.) The eigenvalues are only a way of describing the system's oscillations. A more pragmatic alternative to asking "Why does the system oscillate?" is to ask "How can we explain the system structure in a way that emphasizes the structural features essential to oscillatory behavior?"

An equation solution slightly more intuitive than eigenvalues is given in Forrester's Principles of Systems.<sup>\*1\*</sup> Translating that explanation of oscillation into terms of the spring-mass system, assume that the position P shows steady-state oscillations. That oscillation can be described by a cosine:

$$P(t) = .5*\cos((2*PI/PER)*t)$$

where t is time, PI is the constant 3.1415..., and PER is some period of oscillation. (Its value will be specified later.) Given the assumed behavior of P, we can trace the consequences of fluctuations around the

---

<sup>\*1\*</sup>Forrester, Jay W., Principles of Systems (Cambridge, Mass.: Wright-Allen Press, 1969), Section 10.3.

loop to the velocity V, which is the only rate affecting P. In this way, we can show that oscillations in the position P can be self-sustaining.

The fluctuations in position P create an opposing force exerted by the spring, whose magnitude can be computed from the model equations as follows:

$$\begin{aligned} A(t) &= FS/(W/G) \\ &= -SC*P/(W/G) \\ &= -5*P/5 \\ &= -.5*\cos((2*PI/PER)*t) \end{aligned}$$

Because of the constants chosen for the system, the acceleration A exerted by the spring has a magnitude precisely the negative of that of the position P. Figure 2-3 corroborates this calculation by showing position P and acceleration A at all times equal in magnitude but opposite in sign.

The velocity V is the integral of the acceleration A. The integration causes velocity V to lag 90 degrees behind acceleration A. Thus in Figure 2-3, when acceleration A is at its minimum (trough) value, the velocity V is declining through zero at its maximum rate of change, and when acceleration A is zero a quarter cycle later, the velocity V is unchanging at its minimum (trough) value. If the acceleration A is described by a cosine, the velocity V is described by a sine:

$$\begin{aligned} V(t) &= \int A(t) dt \\ &= -(.5*PER/(2*PI))\sin((2*PI/PER)*t) \end{aligned}$$

Integration also changes the magnitude of the fluctuation by a factor of  $PER/(2*PI)$ . Just as velocity  $V$  is the integral of acceleration  $A$ , the position  $P$  is similarly the integral of the velocity  $V$ . The second integration again produces a phase shift of 90 degrees, and changes the magnitude of the fluctuation by another factor of  $(PER/(2*PI))$ :

$$\begin{aligned} P(t) &= \int V(t) dt \\ &= .5*(PER/(2*PI))**2*\cos((2*PI/PER)*t) \end{aligned}$$

(The "\*\*\*" in the equation above denotes exponentiation, so the term preceding "\*\*\*2" is squared.) Thus, the fluctuation that propagates around the loop in Figure 2-2, beginning with the position  $P$ , and returning to the position  $P$ , has exactly the same phase it was assumed to have. Figure 2-4 summarizes: the position  $P$  caused an acceleration  $A$  phase-shifted by 180 degrees (i.e., multiplied by -1.0); the two integrations caused a phase-shift of 90 degrees each, summing to a total phase shift of 360 degrees, which is equivalent to no phase shift. If the period of oscillation  $PER$  is such that  $(PER/(2*PI))**2$  equals one, then the oscillation which was assumed to exist in the position  $P$  is completely self-sustaining. Indeed, Figure 2-3 indicates that the system has chosen a period apparently equal to that needed to sustain oscillation ( $2*PI$ , or about 6.2830).

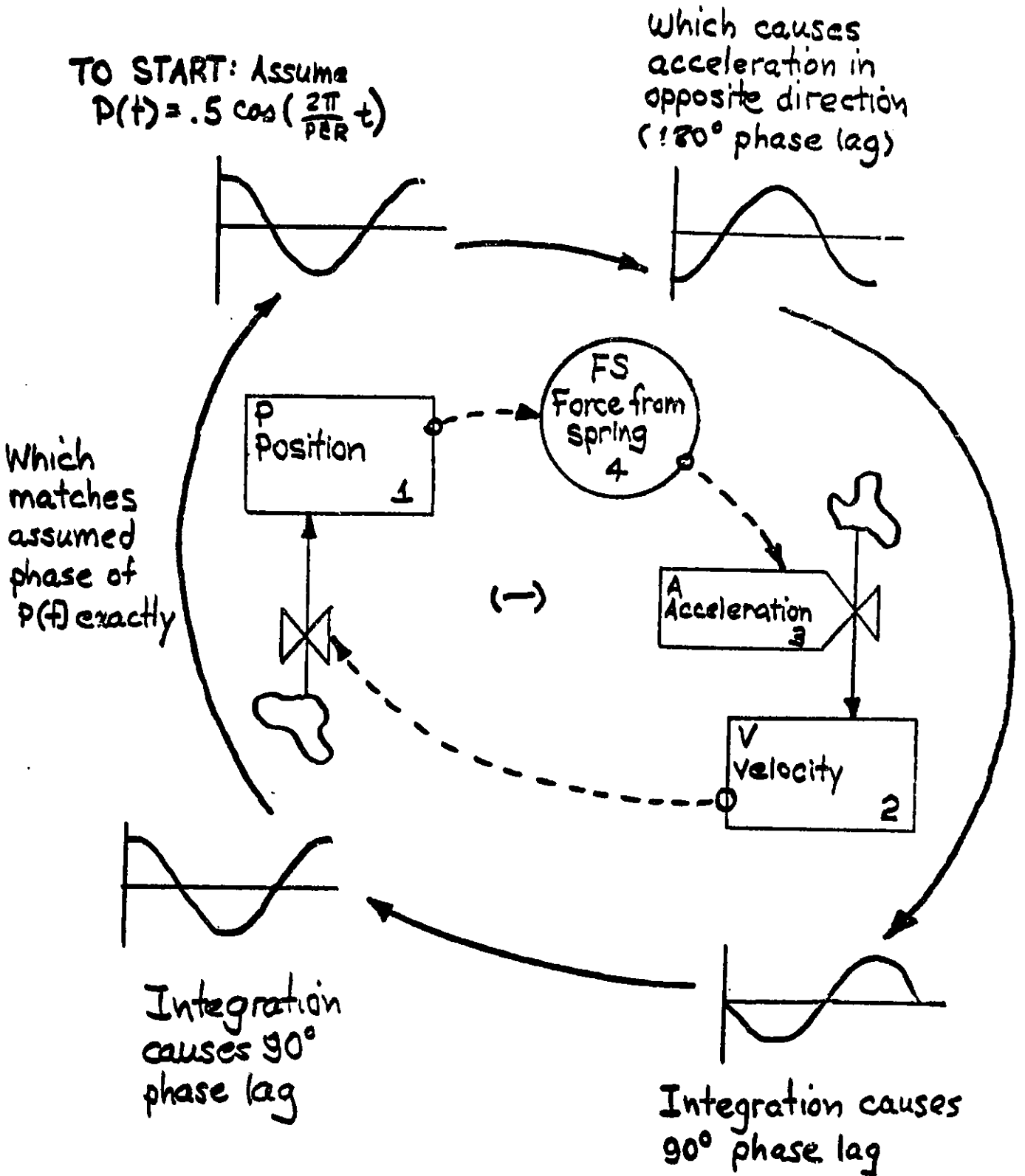


Figure 2-4.  
GRAPHICAL SUMMARY OF EXPLANATION  
IN TERMS OF EQUATION SOLUTION

Which features of the preceding explanation of oscillation are useful, and which are not useful? The most obvious shortcoming of this explanation is a lack of generality, for it capitalizes on a structural feature usually found only in very simple models: all of the rate-level interactions are pure integrations. No minor loops or nonlinearities modify the propagation of the sinusoids (and spoil the mathematics).

The explanation in terms of equation solution is clearly more of a description than an explanation. It relies on an intuition about mathematics, not an intuition about the system structure. The organization of the explanation is devious, for it shows that oscillations in a particular variable, if they exist, are self-sustaining. The explanation does not make it clear that the system will oscillate, merely that it can.<sup>\*2\*</sup> Also, the argument relies on examining open-loop behavior, then closing the loop by strictly mathematical arguments. This strategy of considering an open-loop system is certainly not very appealing conceptually.

One useful concept emerges from Figure 2-4. That figure segments the elements on the loop (levels, rates, and auxiliaries) into subsystems that produce phase lag between their input and their output. It is such phase-lag subsystems that transform the sinusoidal signal and allow it to continue propagating. We can observe that higher-order

-----  
<sup>\*2\*</sup>If one proves enough theorems about existence and uniqueness of solutions, one can indeed prove that the system will oscillate. Still, those theorems do not go very far toward explaining the oscillation.

structures such as a third-order delay can also constitute a phase-lag subsystem, so a subsystem can be a group of levels as well as one level.\*3\*

Explanation in Terms of Departures from Equilibrium Values. In 1975, Mass and Senge evolved an explanation of oscillation in terms of departure from equilibrium values, in addition to the previous concept of phase-lag subsystems.\*4\* Mass and Senge gave their explanation in terms of an industrial model whose two levels were inventory and workforce. That model is structurally analogous to the pendulum model discussed here. In terms of the pendulum model, the explanation runs as follows.

Figure 2-5 shows the first quarter-cycle of the oscillation. The position  $P$  greater than zero causes a negative acceleration  $A$ , causing velocity  $V$  to decline. As the position  $P$  approaches zero, the acceleration  $A$  goes to zero, and the velocity  $V$  temporarily ceases

-----

\*3\*Of course, one could contend that a third-order delay is just three first-order phase-lag subsystems in a row, so that subsystems can be defined as composed of just one level. But the whole purpose of defining concepts such as the phase-lag subsystem is to simplify descriptions of system structure. Therefore, let us define a phase-lag subsystem as a level or group of levels and the associated rates and auxiliaries that produce an output that is phase-lagged in relation to the input.

\*4\*Mass, Nathaniel J. and Peter M. Senge, "Understanding Oscillations in Simple Systems," System Dynamics Group Working Paper D-2045-2 (Cambridge, Mass.: MIT, 1975). Mass and Senge do not use the terms "phase-lag subsystem" or "departures from equilibrium values." These concepts do fit the explanation Mass and Senge give, however.

changing. At time  $t_2$  when the position  $P$  equals zero (its equilibrium value), the system has transformed the initial disturbance in the position  $P$  into a departure of the velocity  $V$  from its equilibrium value.

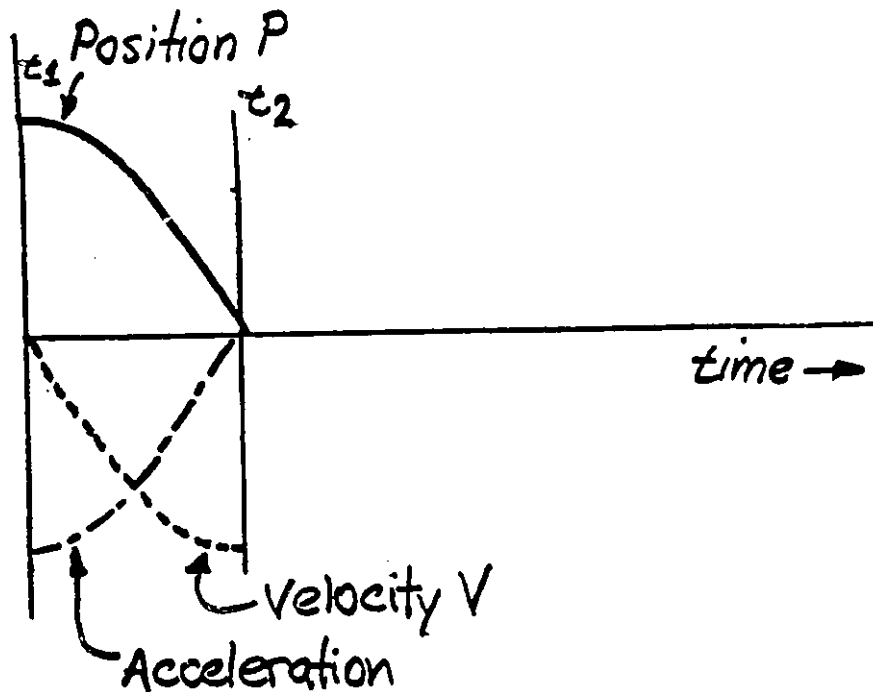


Figure 2-5.  
FIRST QUARTER-CYCLE OF SPRING-MASS OSCILLATION

Figure 2-6 shows how the velocity  $V$  of the mass continues to move the mass past the zero position, so that by the end of two quarter-cycles at time  $t_3$ , the velocity  $V$  is again zero, and the position  $P$  of the mass equals the negative of the initial position  $P$ . The disturbance in  $V$  has propagated around the loop to disturb position  $P$  away from its equilibrium value, this time in the opposite direction from that of the initial disturbance. After the first

half-cycle, the whole process reverses itself, as the disturbance of position  $P$  is transferred to the velocity  $V$  (and position  $P$  is zero) at time  $t_4$ . Finally, Figure 2-7 shows how the mass goes through the zero position and returns once more to its initial position as the disturbance returns the position  $P$  to its initial disequilibrium state at time  $t_5$ .



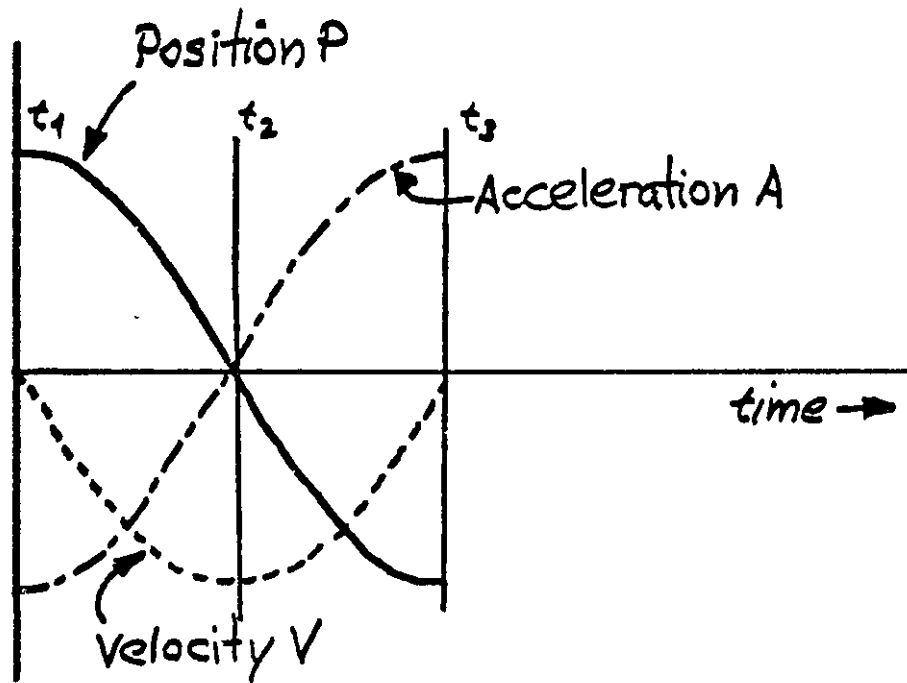


Figure 2-6.  
FIRST HALF-CYCLE OF SPRING-MASS OSCILLATION

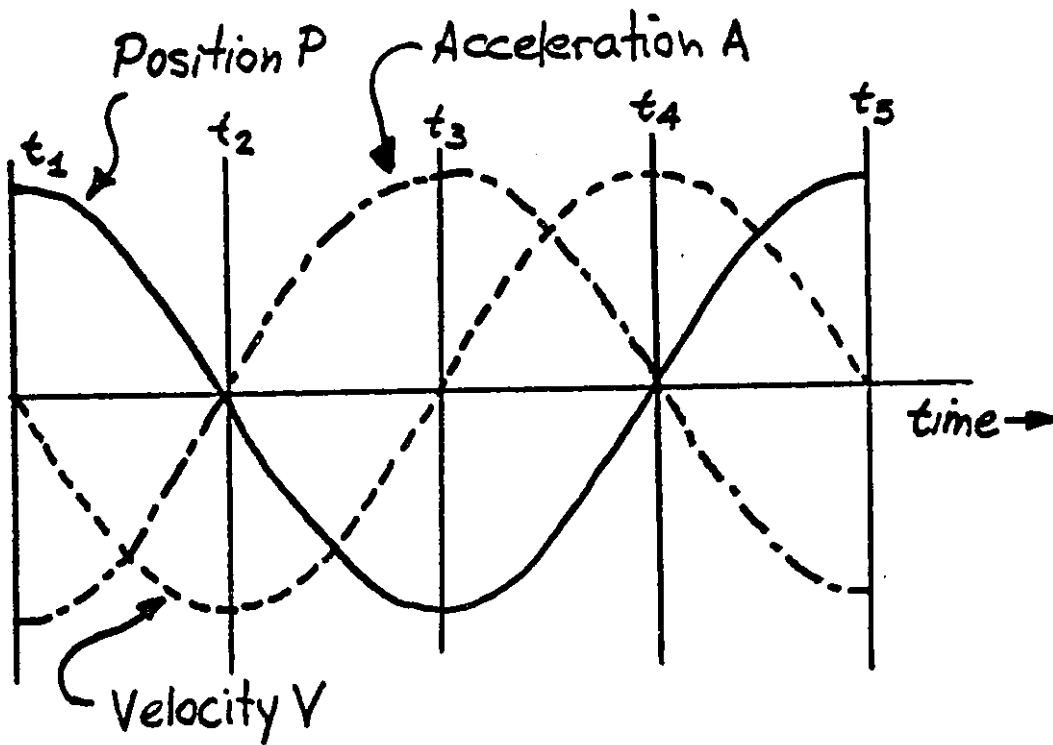


Figure 2-7.  
COMPLETE CYCLE OF SPRING-MASS OSCILLATION

The preceding explanation identifies the feature of a phase-lag subsystem essential to producing oscillation: if the input to such a subsystem goes to its equilibrium value, the output goes to its equilibrium value only later.\*5\* The preceding explanation also suggests the concept of a disturbance propagating around a loop: an exogenous input or initial condition disturbs a subsystem away from its equilibrium value, and even as that subsystem returns to its equilibrium value, the loop has transmitted the disturbance to leave another subsystem out of equilibrium. The disturbance propagates around and around the loop, moving successive subsystems out of equilibrium.

There are, of course, numerous questions left unanswered by the explanation in terms of disturbance from the equilibrium values. The explanation itself does not clearly identify underlying structures necessary for oscillations in general (even though the reflections above begin to): What happens if we change a parameter? Why is the period constant? Why is the period what it is? The following subsection uses another form of describing systems to begin to answer such questions.

---

\*5\*This is not to say that loops containing phase-lead subsystems cannot oscillate; I suspect that the other subsystems on such loops produce more than 360 degrees phase lag, so that aggregating the phase-lead subsystem with the surrounding phase-lag subsystems produces an aggregate phase-lag subsystem. Chapter 6 discusses the issue of phase-lead systems and forecasting as areas for future research.

Explanation in State Space. In order to see more clearly the relationships of each of the levels to their equilibrium values and to each other, we can plot the system behavior along two dimensions representing the two levels, and leave the time dimension implicit: a state-space diagram. (In contrast, the normal computer printout such as Figure 2-3 shows all levels plotted on the same vertical dimension, with time explicit along the horizontal dimension.) Figure 2-8 shows a state-space diagram for the spring-mass system, with one axis for the position  $P$  and one axis for the velocity  $V$ . The diagram gives rate-of-change vectors which indicate how the system will change for every state on the state plane. More specifically, the rate-of-change vectors are the vector sum of the rates of change of position  $P$  and velocity  $V$  that result from the state of the system at that point. Thus, the horizontal component of the rate-of-change vector gives the rate of change of velocity  $V$  (that is, the acceleration  $A$ ) and the vertical component of the rate-of-change vector gives the rate of change of the position  $P$ . The direction of the arrows describes how the system changes, and the magnitude of the arrows describes how fast it changes. For example, when velocity  $V$  is zero and the position  $P$  is greater than zero (as would occur when the spring is stretched to its farthest point), Figure 2-8 shows the vertical component of the arrow equal to zero; there is no rate of change of position  $P$ . The horizontal component of the arrow indicates that the rate of change of velocity  $V$  is negative; the acceleration  $A$  is reducing the velocity  $V$ , to reduce the position  $P$  back to zero.

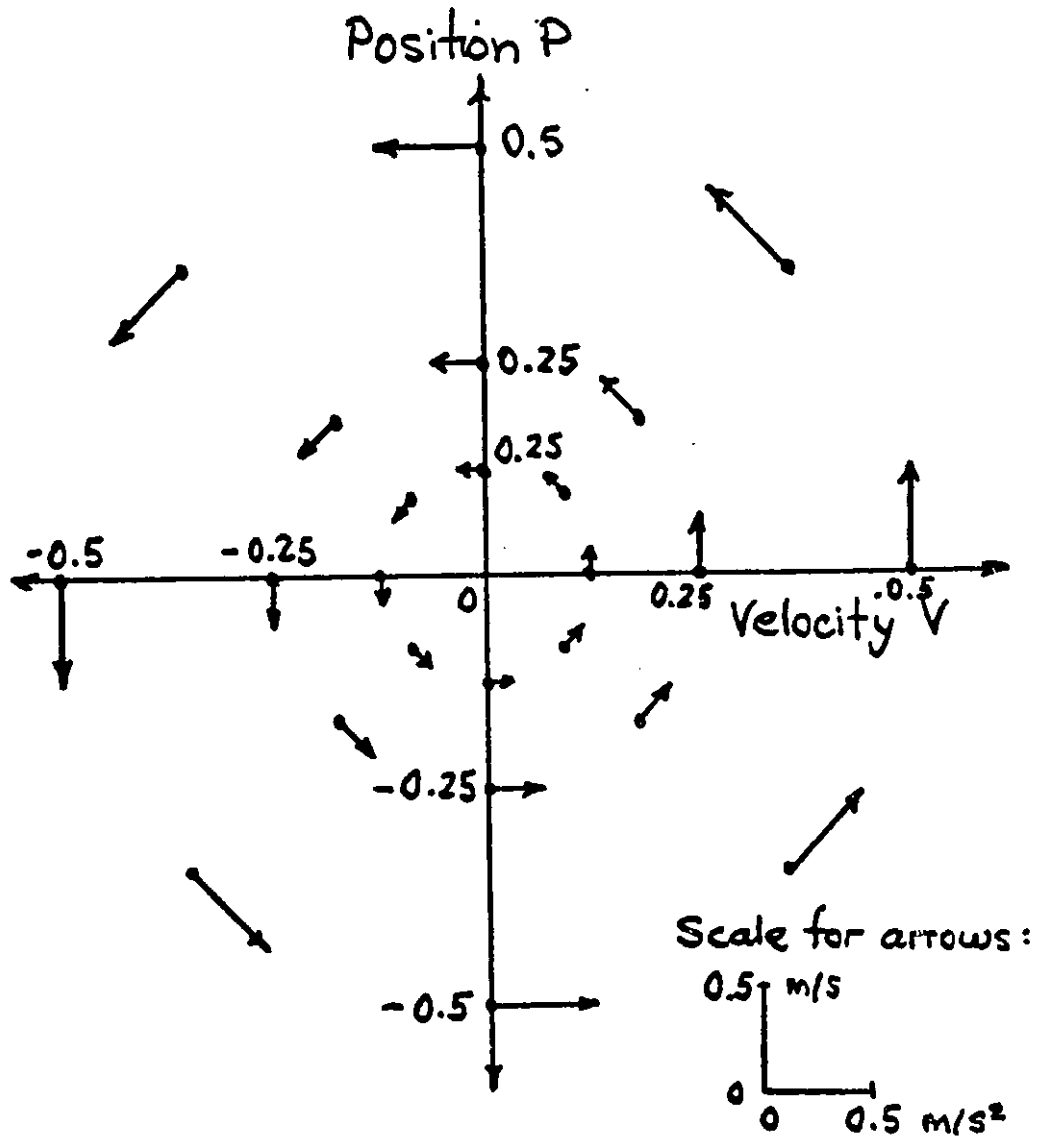


Figure 2-8.  
STATE-SPACE DIAGRAM FOR SPRING-MASS SYSTEM

Producing a diagram such as Figure 2-8 may seem tedious, but we can use the system equations to streamline the computation of the vectors for each point. For example, consider the vectors originating from the vertical axis, where the velocity  $V$  is always zero. The rate of change of position  $P$  is none other than the velocity  $V$ , so that the vertical component of these vectors is always zero. The horizontal

component, the rate of change of velocity  $V$ , equals the acceleration  $A$ . The system equations work out to make acceleration  $A$  equal to the negative of position  $P$ . Thus, the magnitude of the arrows originating from the vertical axis is always proportional to the distance from the equilibrium point. In fact, the derivation below shows that the magnitude of any rate-of-change vector on the state-space diagram is proportional to its distance from the equilibrium point, and that the direction of the vector is at right angles (90 degrees clockwise) from the direction from the equilibrium point to the specific state.

The magnitude of a rate-of-change vector (" $r$ ") is given by the normal Euclidean distance formula

$$\begin{aligned} |r| &= (((d/dt)P)**2 + ((d/dt)V)**2)**(1/2) \\ &= (V**2 + (-P)**2)**(1/2) \\ &= (V**2 + P**2)**(1/2) \end{aligned}$$

Thus, the magnitude of the rate-of-change vector is exactly equal to the magnitude of the state vector (i.e., the Euclidean distance from the equilibrium point to the point on the state plane). The direction of the rate-of-change vector can be specified by its slope, which equals the vertical component divided by the horizontal component, or

$$\text{slope}(r) = (d/dt)P/(d/dt)V = V/(-P)$$

From geometry, we know that if a line has a slope  $S$ , then the line at right angles to it has a slope  $-1/S$ . The state vector has a slope  $P/V$ , so the line at right angles to it has a slope of  $-V/P$ , which is exactly the slope of the rate-of-change vector. Therefore, the rate-of-change

vector is either 90 degrees clockwise or counterclockwise from the state vector. The counterclockwise direction of the rate-of-change vector follows from observing that when velocity  $V$  is positive, the rate of change of position  $P$  is also positive. Thus, in the right side of the state plane, the vertical (position  $P$ ) component of the rate-of-change vector must be positive: the vector goes upward and not downward. (The opposite happens in the left side of the state plane.) The upward orientation in the right side of the state plane corresponds to a counterclockwise rotation from the direction of the state vector.

Given a state-space diagram such as Figure 2-8 with a large number of arrows, the trajectory of the system can be approximated graphically by connecting the arrows. Figure 2-9 shows two such trajectories for the spring-mass system, each corresponding to a different set of initial conditions. Consider the trajectory beginning at point A, where the position  $P$  equals one-half foot, and the velocity  $V$  equals zero. (This corresponds to the simulation in Figure 2-3.) At point  $t_1$ , the position  $P$  is away from its equilibrium value.\*6\* The negative loop begins to build up velocity  $V$  to carry  $P$  back to equilibrium, which moves the trajectory directly left. The

-----

\*6\*"Equilibrium value" characterizes the system behavior more than it does the system structure. Although models of managerial systems equilibrate when levels are equal to managerially-determined desired levels, many systems equilibrate at values with no particular structural significance. For example, the structure of the Urban Dynamics model does not contain any targets at which the system should equilibrate. Indeed, until one has studied the behavior of the model, one cannot easily determine from the system structure even approximate equilibrium conditions.

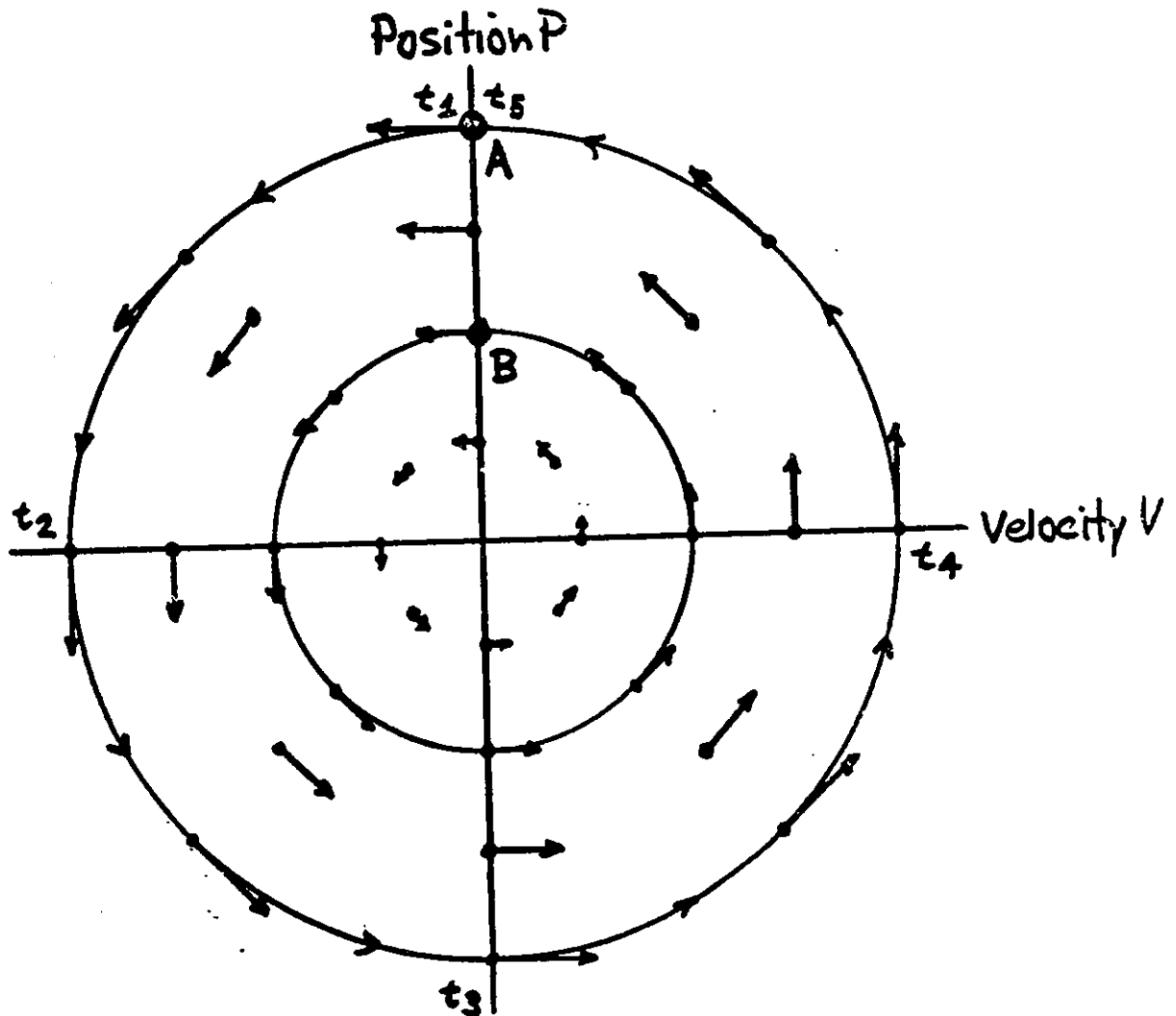


Figure 2-9.  
TRAJECTORY OF SPRING-MASS SYSTEM IN STATE SPACE

position P continues to exceed its equilibrium value, so velocity continues to build up until at time  $t_2$ , P is at its equilibrium value, but the disturbance has propagated to leave V away from its equilibrium value. The built-up velocity V continues to move the level P past its equilibrium value, until time  $t_3$ , when the velocity V becomes zero, but the position P is again away from its equilibrium value. The situation

at time  $t_3$  is in effect the mirror image of the situation at time  $t_1$ : the velocity  $V$  is at its equilibrium value, but the position  $P$  is disturbed away from its equilibrium value, this time below it. Thus, the same events will ensue in reverse, with velocity  $V$  building up and moving  $P$  past its equilibrium point. One cycle of oscillation is completed when the system returns the position  $P$  at time  $t_5$  to the same value it had at time  $t_1$ . The system will obviously continue to oscillate around the circular trajectory.\*7\*

The state-space diagram indicates that oscillation is an inescapable property of the structure of the spring-mass system. The negative loop shown in Figure 2-2 compensates for any of the system levels being out of equilibrium. But the compensation involves moving to the side in state space, so that a level travels to zero only when the other level moves away from zero, forming a quarter-circle in state space. The system always attempts to move toward the equilibrium point, but always fails because the system is always moving sideways as well. A physical analog to this trajectory is an orbiting satellite, which is always falling toward earth, but never reaches earth because it is also moving sideways with respect to the earth.

-----

\*7\*For the trajectory to form a perfect circle is fairly unusual; this is caused by the absence of minor loops and the values chosen for the parameters. One way of seeing that the trajectory moves in a circle is to remember that the rate-of-change vectors are always at right angles to the line from the vector to the origin, so that the distance from the system state to the origin can never change. Obviously, when this property is changed, the trajectory no longer necessarily forms a perfect circle. Trajectories which will be explored later include inward- and outward-spiraling trajectories.



The state-space form of describing the structure and behavior offers tentative explanations to many of the questions posed at the end of the last subsection.

Why is the period constant? Because the system returns after one cycle to values of the levels completely identical to the initial conditions, so that as far as a state-space description goes, the succeeding cycles of an oscillation are not similar behavior but the same behavior, which of course has the same period each time it manifests itself. We can formalize these ideas by saying that the initial disturbance propagates around the loop until it returns the system's levels to a comparable disequilibrium state.

Why is the period the same for all amplitudes of oscillation? We know that for the spring-mass system, the speed at which the state travels through state space (i.e., the magnitude of the rate-of-change vector) is proportional to the distance between the state and the equilibrium point. The distance through which the state must move to complete one cycle of oscillation is also proportional to the distance between the state and the equilibrium point. Thus, no matter what the amplitude of oscillation, the time it takes to circumnavigate one complete cycle remains constant. In any system where the rate of change is approximately proportional to the discrepancies between the actual levels and the equilibrium level values, increasing the amplitude of disturbance increases both the speed of change and the amount of changing needed to complete a cycle together. Thus, in such a system,

the period remains the same for any amplitude of disturbance.

Why is the period what it is? The period is the time it takes for the rate-of-change vectors to propel the system around one complete cycle. The magnitude of the rate-of-change vectors depends on how rapidly each subsystem attempts to move the next subsystem around the negative loop toward equilibrium. For example, the parameters that characterize "how rapidly" (i.e., the coupling time constants) for the spring-mass system are 1.0 ( $(d/dt)P = V*1.0$ ) and  $-SC/(W/G)$  ( $((d/dt)V = (-SC/(W/G))*P$ ). These two parameters connect the two subsystems of the spring-mass system together, and determine "how rapidly" disequilibrium in one system changes the other subsystem. The magnitude of the rate-of-change vector is

$$\begin{aligned} |r| &= ((d/dt)P^{**2} + (d/dt)V^{**2})^{**0.5} \\ &= (V^{**2} + ((-SC/(W/G))*P)^{**2})^{**0.5} \end{aligned}$$

When  $P = 0$ , the magnitude of the rate-of-change vector is 1.0 times the distance from the state to the equilibrium point. When  $V = 0$ , the magnitude of the rate-of-change vector is  $|(-SC/(W/G))|$  times the magnitude of the state vector. (For the parameters chosen,  $(-SC/(W/G))$  also equals 1.0, but this is not true in general.) For states not on the axes, the magnitude will be somewhere between 1.0 and  $|(-SC/(W/G))|$  times the magnitude of the state vector. The description in terms of magnitude of the rate-of-change vector jibes well with Forrester's Principle 10.3-1 in Principles of Systems that

The second-order negative loop with no minor loops oscillates as a sustained sinusoid with a period  $P = 2\pi(A_1 A_2)^{0.5}$  where the A's are the coupling time constants or the reciprocals of the gain multipliers that relate levels to succeeding rates.

So the rate-of-change propels the state around one cycle at an average speed proportional to the geometric mean of the gains.\*8\*

One way of deriving the period (at least approximately) is to assume that the average speed of movement through state space is:

$$(\text{geometric mean of gains})(\text{magnitude of state vector})$$

and that the distance to be moved in one complete cycle is

$$2\pi(\text{magnitude of state vector})$$

Then the time it takes to move that distance at that speed is distance divided by speed or

$$2\pi/(\text{geometric mean of gains})$$

which, assuming that the coupling time constants are defined as the reciprocals of the coupling gains, equals Forrester's result of

$$2\pi(\text{geometric mean of coupling time constants})$$

Thus far we have seen three explanations of oscillations: in terms of equation solution, in terms of departures from equilibrium

-----

\*8\*There is another explanation of why the period is what it is: not so long that there is very little phase shift around the loop, and not so short that the signal is attenuated. Further exploration of this question, or finding commonality between the two explanations, however, seems inappropriate to the scope of this thesis.

values, and in terms of movements in state space. From these explanations we have abstracted three concepts: a loop formed of phase-lag subsystems, a disturbance propagating around the loop from subsystem to subsystem, and the disturbance moving the state through one complete cycle to return to a comparable disequilibrium state. Before formally unifying these concepts into a principle, let us examine the oscillatory behavior of two more systems to clarify and refine the concepts.

Oscillation from a Purely Positive Loop. It is customary to think of oscillations as arising from negative loops, even though it is not hard for us to conjure up a counterexample where a series of delays (first-order negative loops) are interconnected to form a positive loop. Disturbances might propagate around such a positive loop to produce oscillatory behavior superimposed on exponentially-growing behavior. Indeed, that combination of behavior modes arises from such a structure in Forrester's "Market Growth as Influenced by Capital Investment."\*9\* Figure 2-10 shows a causal-loop diagram of the structure. (The levels are shown on the positive loop.) It is tempting to ascribe the oscillation in some fashion to the presence of minor negative loops on the major positive loop. This ascription is not correct, as can be seen from the following example.

-----  
\*9\*Forrester, Jay W., "Market Growth as Influenced by Capital Investment," Industrial Management Review [now Sloan Management Review] (Winter 1968).

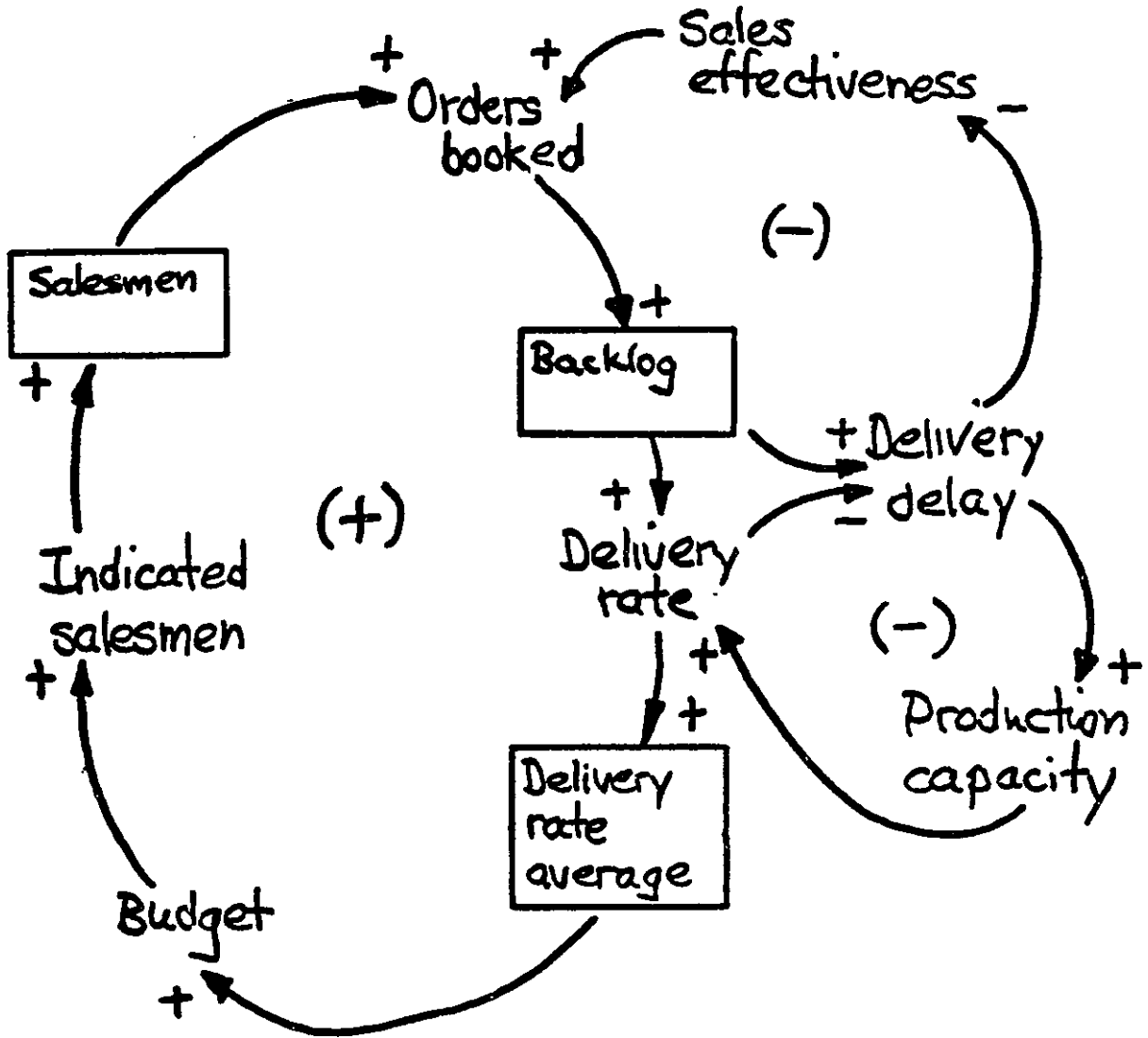


Figure 2-10.  
MARKET-GROWTH MODEL

Consider a system where four integrators are connected in a single positive loop, as shown in Figure 2-11. The equations for the system are:

$$(d/dt)L1 = L4$$

$$(d/dt)L2 = L1$$

$$(d/dt)L3 = L2$$

$$(d/dt)L4 = L3$$

Solving for a differential equation in terms of L1,

$$(d^4/dt^4)L1 = L1$$

or

$$(d^4/dt^4)L1 - L1 = 0$$

The characteristic equation is therefore

$$s^{**4} - 1 = 0$$

This equation factors as follows

$$s^{**4} - 1 = 0$$

$$(s^{**2} + 1)(s^{**2} - 1) = 0$$

$$(s+j)(s-j)(s+1)(s-1) = 0$$

so that the eigenvalues of the system are  $-j$ ,  $+j$ ,  $-1$ , and  $+1$ , where  $j$  is the square root of negative one.

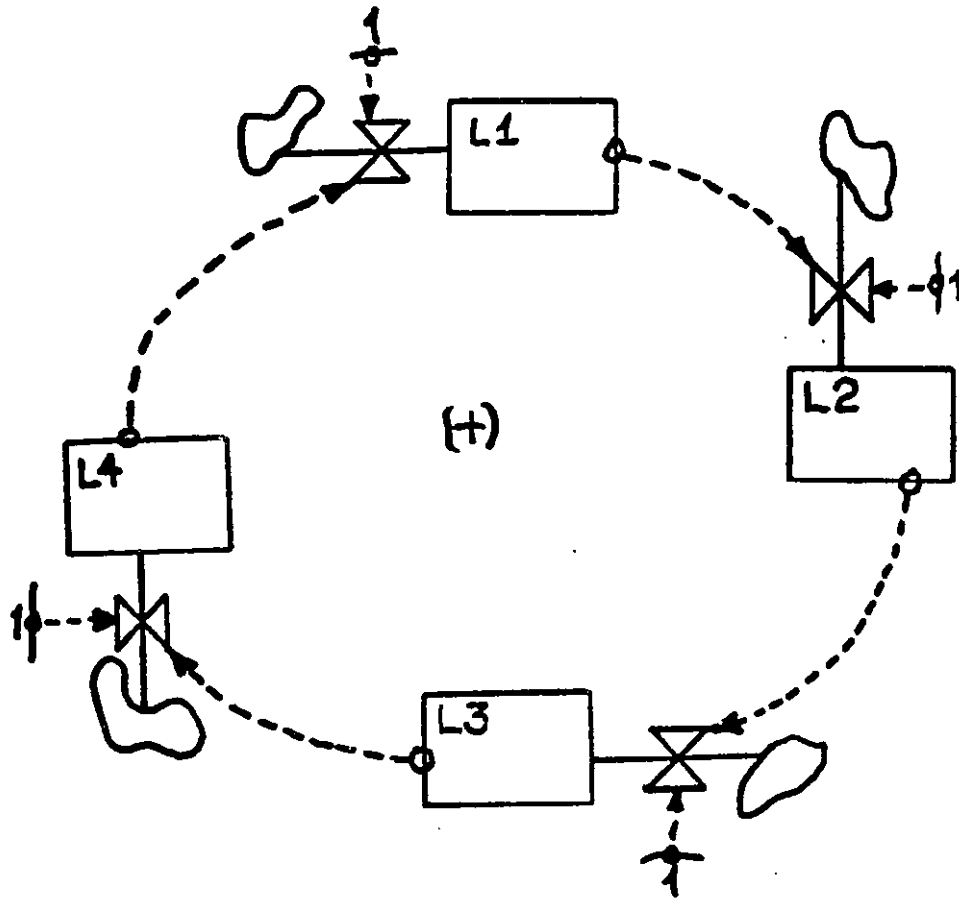


Figure 2-11.  
PUREL. POSITIVE LOOP

The eigenvalue of +1 is expected, as it corresponds to exponential growth. The eigenvalue of -1 might be thought of as characterizing the time it takes for all of the levels to enter a pure steady-state exponential growth mode. Both of these eigenvalues occur in a second-order positive loop with pure integrators and unity gains, as well as the fourth-order system discussed here.

The eigenvalues of +j and -j correspond to sustained sinusoidal oscillation. Four integrations produce four 90-degree phase shifts, so that a disturbance, once started, can propagate itself indefinitely. This oscillation is analogous to the oscillation in Figure 2-4, where two integrations produced a 90-degree phase shift each, and the negative sign in the negative loop in effect produce another 180-degree shift, so that the total 360-degree phase shift allows an initial disturbance to propagate itself indefinitely.

Just as structural changes to the second-order negative loop in Figure 2-4 can amplify or attenuate the propagating disturbance, so probably can modifications to the higher-order positive loop in Figure 2-11 likewise amplify or attenuate a propagating disturbance. Oscillation arises from disturbances propagating around loops.



Period.\*10\*

If oscillations can arise as an integral part of an exponentially-growing (i.e., non-equilibrium) behavior mode, it is clear that the previous explanations in terms of disturbance from equilibrium must be modified: the oscillatory behavior in the system in Figure 2-11 can be described in terms of departures from steady-state values. "Steady-state" here denotes all other behavior modes, and excludes the oscillation being examined (which in this case also could be considered a steady-state phenomenon, as the oscillations do go on indefinitely).

Let us further refine the concepts with one more example before integrating them into a principle.

Damped Oscillation in a Managerial System. The previous examples have had structures consisting of pure integrations and oscillatory behavior consisting of pure sinusoids. We can examine a slightly more complex structure that gives rise to another type of oscillatory behavior: damped

-----

\*10\*The traditional discussion of loop polarity and behavior connects negative loops with convergent, sometimes oscillatory behavior, and positive loops with divergent behavior. But there are certainly now too many practical counterexamples for these connections to remain useful. Negative loops can show explosively divergent oscillatory behavior. Positive loops with steady-state open-loop gains less than 1.0 can show convergent, nonoscillatory behavior (see Chapter 4). And finally, as the discussion above indicated, positive loops can generate convergent or divergent oscillations superimposed on other behavior modes. For further discussion as well as explicit simulations of the four-integrator positive loop, see Graham, Alan K., "Positive Loops and Divergent Behavior," System Dynamics Group Working Paper D-2751, Alfred P. Sloan School of Management (Cambridge, Mass.: MIT, 1977).

oscillations, where the magnitude of the oscillation diminishes with each cycle. The example (which will be reexamined many times in subsequent chapters) is an oscillatory managerial system. In brief, the system can be described as follows: a company receives orders for its products, which accumulate in an order backlog until the company fills the order by producing the required product. If the order backlog becomes too high, the company hires more people to produce its goods more rapidly and reduce the backlog. Figure 2-12 shows a DYNAMO flow diagram of the system. The major outside loop is negative. As the order rate OR increases the backlog B, this increases the correction for backlog CB, the desired employment DE, the correction for employment CE, and the net hiring rate NHR. As hiring increases the employment E, the output OUT increases, which decreases the backlog B. The goals for desired employment DE and desired backlog DB are based on the average level of activity in the company, represented by the expected average output EAO.\*11\*

---

\*11\*The expected average output EAO is one of several variables held constant in this section. These variables will be made active in later sections to show the effects on behavior of structural changes.

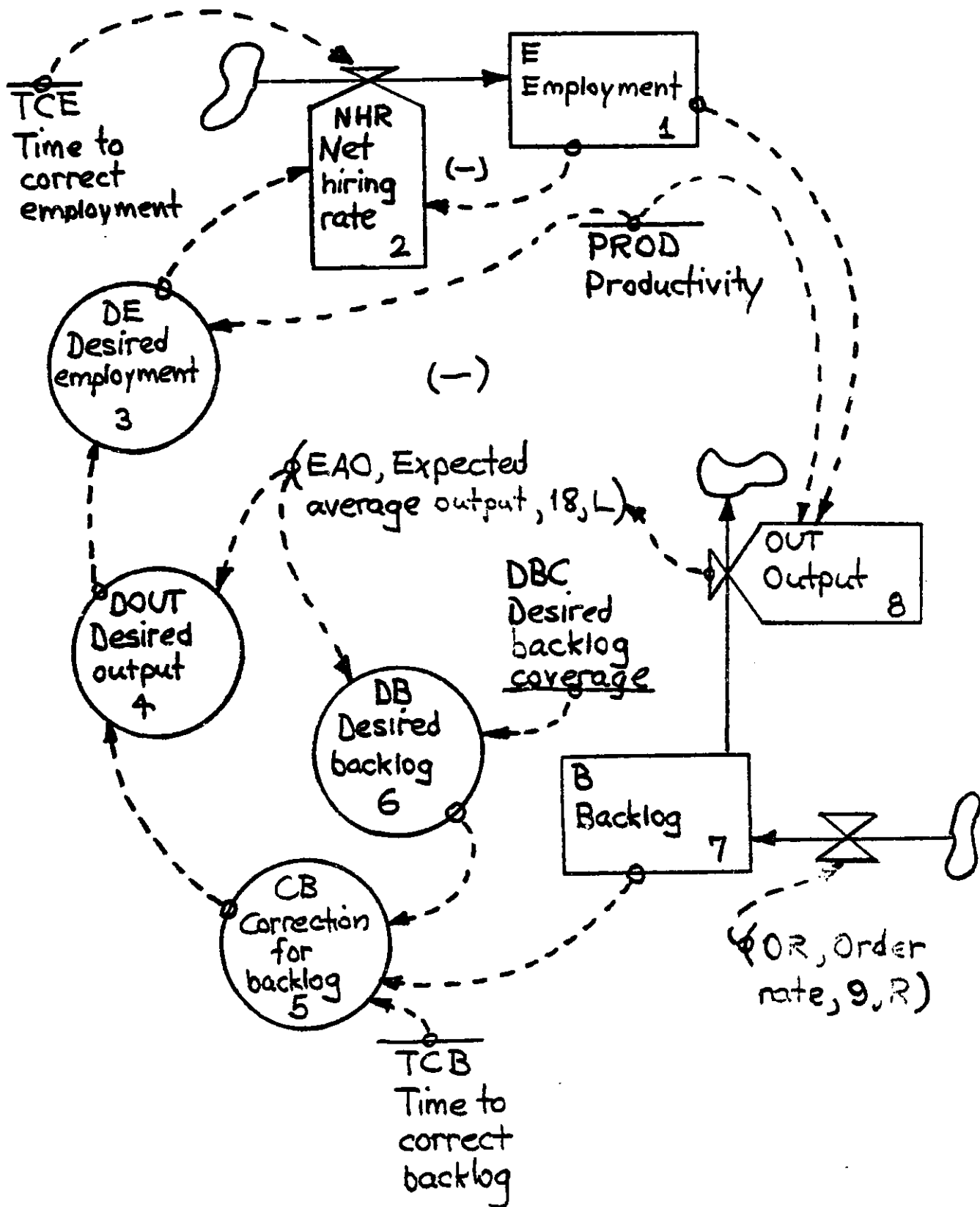


Figure 2-12.  
DYNAMO FLOW DIAGRAM OF EMPLOYMENT-BACKLOG SYSTEM

The employment-backlog system may at first seem much more complex than the spring-mass system whose flow diagram appears in Figure 2-2. However, much of the apparent complexity is due to the auxiliaries connecting the backlog B to the net hiring rate NHR. Those auxiliaries can be thought of as a disaggregation of the rate equation for NHR. If we think of the rate equation subsuming all of the auxiliaries, the structure appears much simpler: Figure 2-13 shows that the structure of the employment-backlog system is a simple second-order negative loop, like the spring-mass system. The only difference is that the employment-backlog system has an additional minor negative loop around one level.

Employment E is a level variable altered by the net hiring rate NHR.

$$\begin{aligned} E.K &= E.J + (DT)(NHR.JK) && 1, L \\ E &= EN && 1.1, N \\ EN &= 50 && 1.2, C \end{aligned}$$

E - EMPLOYMENT (MEN)  
NHR - NET HIRING RATE (MEN/YEAR)  
EN - EMPLOYMENT INITIAL (MEN)

The net hiring rate NHR represents a management policy, in this case a rather simple one. The discrepancy between the actual employment E and the desired employment DE is corrected with a time constant of TCE--the time to correct employment.

$$\begin{aligned} NHR.KL &= (DE.K - E.K) / TCE && 2, R \\ TCE &= .5 && 2.1, C \end{aligned}$$

NHR - NET HIRING RATE (MEN/YEAR)  
DE - DESIRED EMPLOYMENT (MEN)  
E - EMPLOYMENT (MEN)  
TCE - TIME TO CORRECT EMPLOYMENT (YEARS)

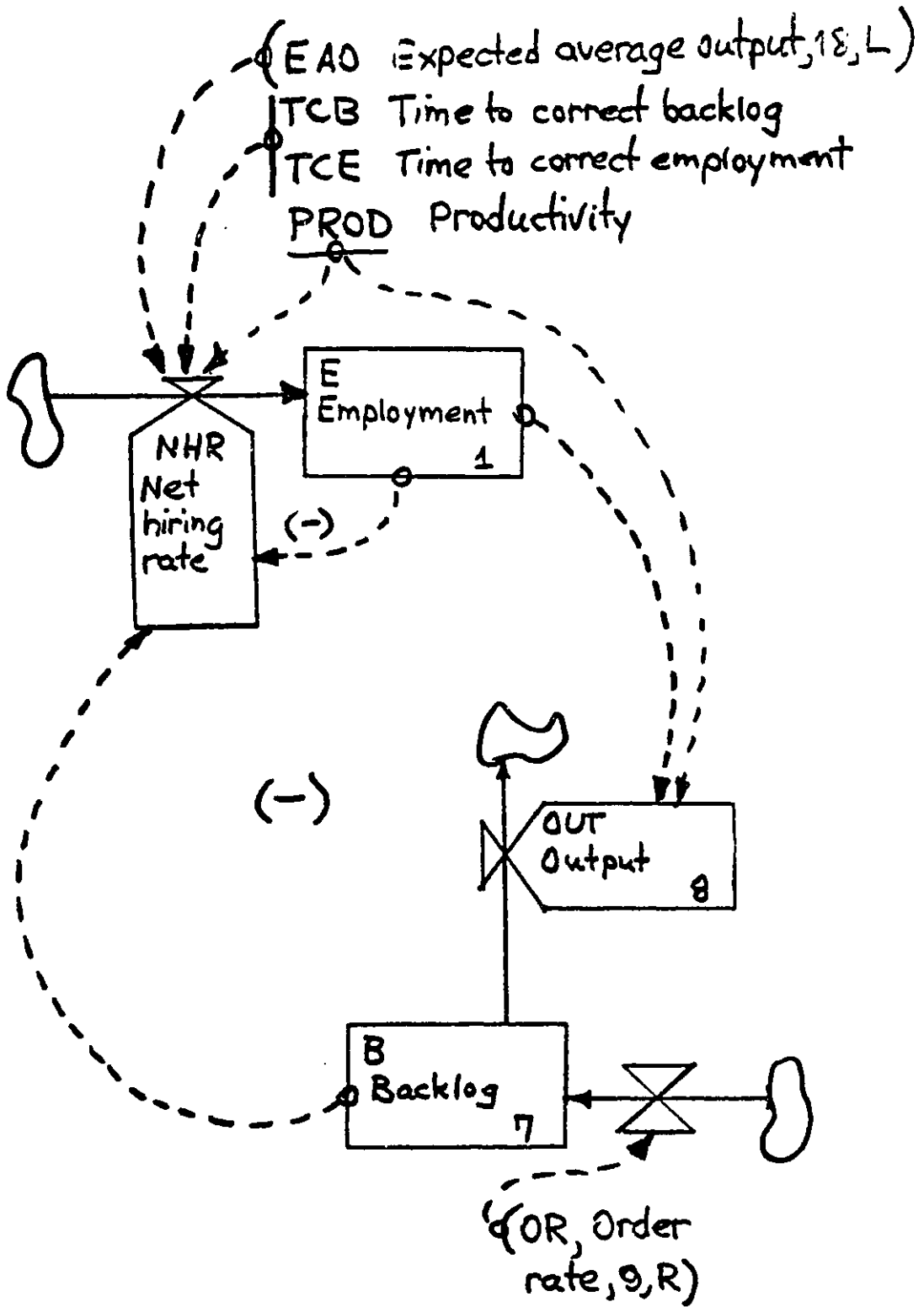


Figure 2-13.  
SIMPLIFIED FLOW DIAGRAM OF EMPLOYMENT-BACKLOG SYSTEM

The desired employment DE responds to production plans, as represented by the desired output DOUT. The desired employment DE is simply equal to the desired output DOUT divided by the productivity PROD. This constant represents the number of units produced by an employee in one year. It has been set at 30 units per year per man, representing a fairly large, complex unit to be manufactured.

$$DE.K = DOUT.K / PROD$$
$$PROD = 30$$

3, A  
3.1, C

DE - DESIRED EMPLOYMENT (MEN)  
DOUT - DESIRED OUTPUT (UNITS/YEAR)  
PROD - PRODUCTIVITY (UNITS/YEAR/MAN)

The desired output DOUT has two components. The first component, the expected average output EAO, provides a base figure around which production can be planned. The second component, the correction for backlog CB, adjusts this base figure by increasing desired production when the backlog of orders is high, or decreasing production when very few orders have been placed.

$$DOUT.K = EAO.K + CB.K$$

4, A

DOUT - DESIRED OUTPUT (UNITS/YEAR)  
EAO - EXPECTED AVERAGE OUTPUT (UNITS/YEAR)  
CB - CORRECTION FOR BACKLOG (UNITS/YEAR)

The correction for backlog CB responds to a discrepancy between the backlog B and the desired backlog DB, with a time constant equal to the time to correct backlog TCB.\*12\* Thus, when backlog B exceeds the desired backlog DB, the correction for backlog CB is greater

-----

\*12\*Technically, TCB is not a time constant but a coupling time constant. This distinction is discussed in Section 3.1.

than zero, which increases the desired output DOUT in the face of an excessive number of orders.

$CB.K = (B.K - DB.K) / TCB$  5, A  
 $TCB = .5$  5.1, C  
CB - CORRECTION FOR BACKLOG (UNITS/YEAR)  
B - BACKLOG (UNITS)  
DB - DESIRED BACKLOG (UNITS)  
TCB - TIME TO CORRECT BACKLOG (YEARS)

The desired backlog DB sets a goal of a constant backlog coverage (the desired backlog coverage DBC), based on the expected level of activity (as measured by the expected average output EAO). Thus, if the company is manufacturing 1200 units per year, and the desired backlog coverage DBC equals 0.5 years, the company wants to have  $1200 \times 0.5 = 600$  orders in the backlog.

$DB.K = EAO.K * DBC$  6, A  
 $DBC = .5$  6.1, C  
DB - DESIRED BACKLOG (UNITS)  
EAO - EXPECTED AVERAGE OUTPUT (UNITS/YEAR)  
DBC - DESIRED BACKLOG COVERAGE (YEARS)

The backlog B of orders is a level increased by the incoming order rate OR, and decreased by the company's output OUT (which by assumption is shipped immediately and thus depletes the backlog B.) The backlog B is initialized at its desired level, the desired backlog DB.

$B.K = B.J + (DT)(OR.JK - OUT.JK)$  7, L  
 $B = BN$  7.1, N  
 $BN = 660$  7.2, C  
B - BACKLOG (UNITS)  
OR - ORDER RATE (UNITS/YEAR)  
OUT - OUTPUT (UNITS/YEAR)  
BN - BACKLOG INITIAL (UNITS)

The actual output OUT, as mentioned before, is simply equal to the current employment E times the productivity PROD per man. (In this simple model, we will assume that the company uses no overtime or fractional workdays to meet production plans.)

$$\text{OUT.KL} = \text{E.K} * \text{PROD} \qquad \qquad \qquad 8, R$$

OUT	-	OUTPUT (UNITS/YEAR)
E	-	EMPLOYMENT (MEN)
PROD	-	PRODUCTIVITY (UNITS/YEAR/MAN)

Figure 2-14 shows the response of the employment-backlog system when employment E is initialized out of equilibrium. At first, the overshoot and slight undershoot of most variables may seem to bear little resemblance to the persistent sinusoidal behavior shown in the spring-mass system. However, closer examination reveals the same qualitative pattern of behavior. The initial employment E exceeds the steady-state value of employment necessary to produce output OUT at a rate equal to the order rate OR. As a result of the initial disturbance in employment E, OUT exceeds the order rate OR, and the backlog B begins to fall. The falling backlog B reduces the desired employment DE and thus the actual employment E falls. At about time 0.6 years, employment E has been reduced to its steady-state value, and the disturbance has propagated to the backlog B, which is well below its steady-state value. But employment E still exceeds the desired employment DE (which is below its steady-state value because the backlog B is), so employment E and output OUT continue to fall. After



about time 0.6 years, the order rate OR exceeds the output OUT, so that the backlog B rises, soon to equal its steady-state value of 660 units at time 1.6 years. But the phase-lag subsystem that determines output OUT is still below its steady-state value at time 1.6 years; the disturbance has propagated back to a disturbance in the employment E, this time of the opposite sign as the initial disturbance.

Subsequently, the same pattern ensues to return the system to a comparable disequilibrium state at about time 3.2 years, where the backlog B is at its steady-state value, and the employment E is disturbed above its steady-state value.

The major difference between Figure 2-14 and the spring-mass oscillation in Figure 2-3 is that the disturbance attenuates as it propagates around the employment-backlog loop. Thus, after one cycle, the disturbance in the employment E is reduced to about a twelfth of its initial magnitude. (Section 2.2 explains how minor negative loops such as that around employment E can attenuate disturbances and cause damped oscillation.)

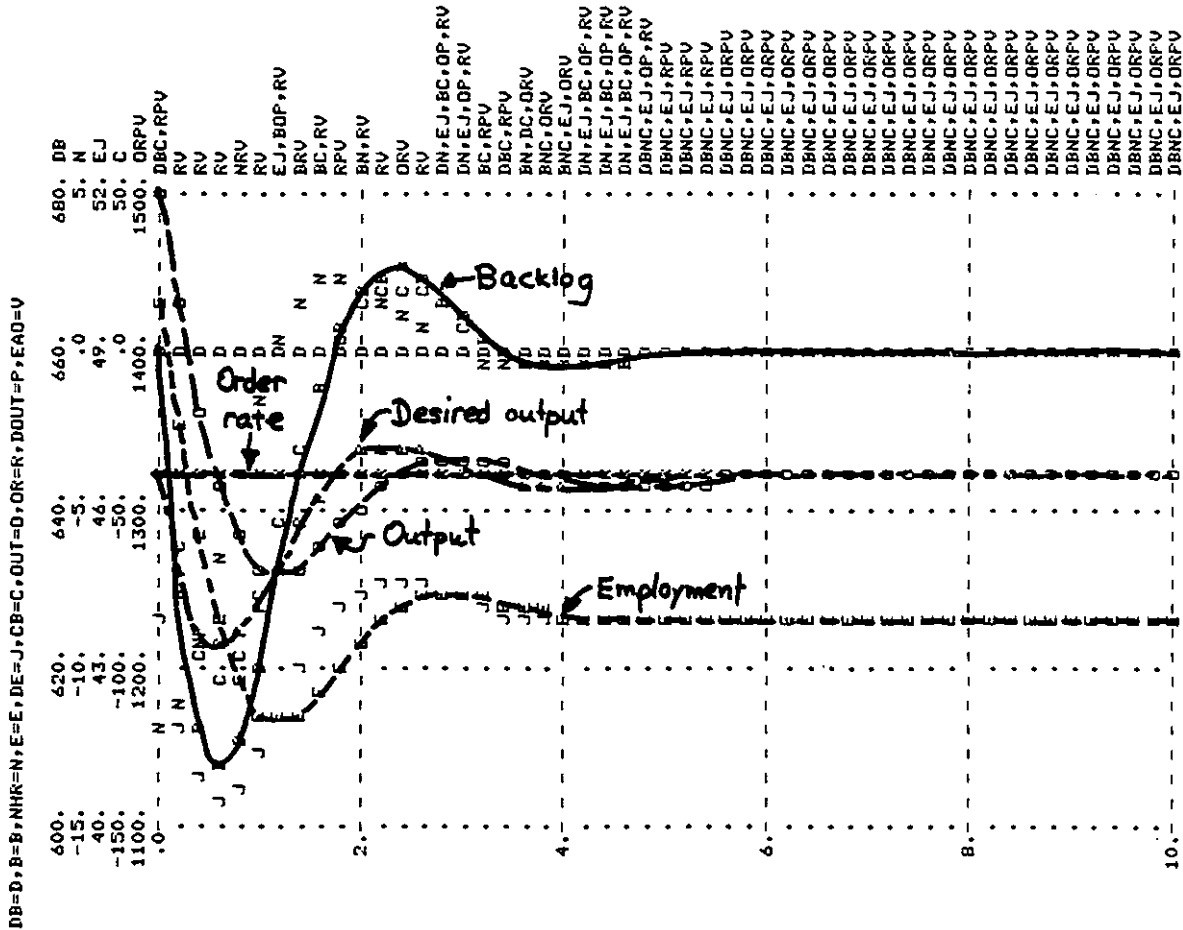


Figure 2-14.  
BEHAVIOR OF EMPLOYMENT-BACKLOG SYSTEM

Figure 2-15 shows the state-space trajectory of employment  $E$  and backlog  $B$ . Instead of the circular trajectory of the spring-mass system (Figure 2-9), the trajectory spirals inward, indicating a constant decrease in the disturbance. The disturbance does produce a

comparable disequilibrium state after one cycle of 3.3 years, however. Figure 2-15 shows the trajectory of E and B after one cycle on an enlarged scale, with a dotted line. The trajectory after year 3.3 is quite similar in shape and period to the initial trajectory; only the magnitude is different.\*13\* (Again, the analogy of the satellite is apt, for a satellite whose motion is restricted by atmospheric friction shows such inward spirals, until it either hits the earth or burns from friction heating.)

The example of the employment-backlog system indicates that we must extend the concept of comparable state to include not only equal states, as it did before, but also states whose disturbances from their steady-state values are proportional to one another. Thus, the initial state is comparable to any state on the upper half of the vertical axis. In general comparable states will lie on the same ray in state space

---

\*13\*The similarity of trajectories comes about because of the linearity of the system around the equilibrium point. As long as the system is approximately linear, multiplying the initial conditions by some constant will multiply the response by the same constant, by the principle of superposition. Thus, a ray in state space from the equilibrium point specifies a set of states whose ensuing trajectories differ from one another only by multiplication by a constant. (This linearity also explained why the spring-mass system shows the same period for any magnitude of oscillation.)

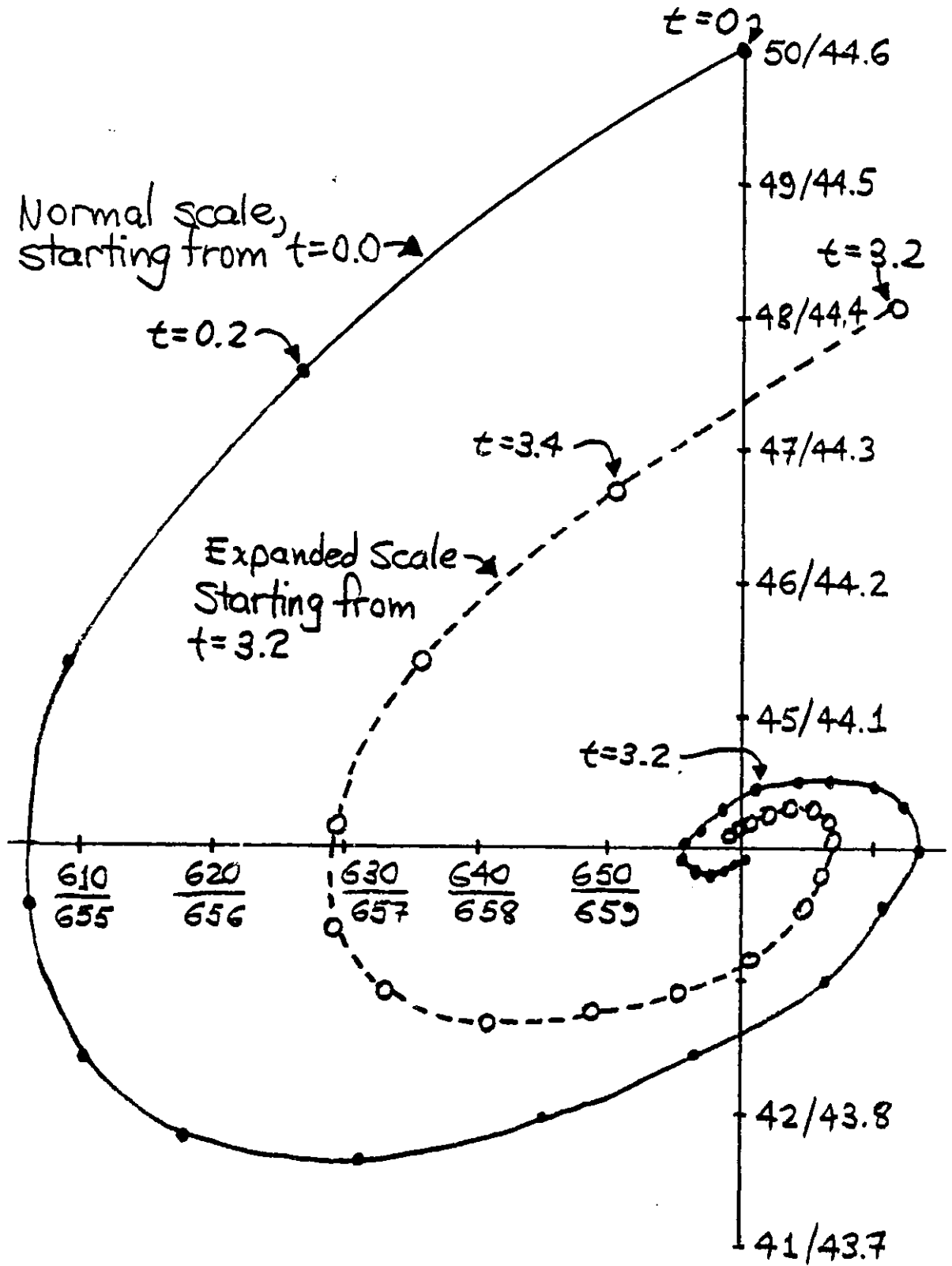


Figure 2-15.  
STATE-SPACE PLOT OF EMPLOYMENT-BACKLOG SYSTEM

emanating from the equilibrium point.\*14\* \*15\* Thus, for the damped oscillation of the employment-backlog system, we can characterize each cycle as showing comparable trajectories, and the explanation for the constant period of the spring-mass system applies to the present oscillations as well.

-----

\*14\*Technically, we could define any states on a line (instead of a ray) as comparable, with the disturbances away from steady-state values of some states being the negative of the disturbances of other states. Restricting comparable states to a ray, however, allows us to visualize comparable trajectories with just scale changes (instead of mirror-image reflections).

\*15\*Most socioeconomic systems contain relative smooth nonlinearities, so that the system is approximately linear about the equilibrium point. Thus, the description given above often holds, but only approximately.

A Principle. As remarked earlier, the function of a principle is to state a relationship among concepts that summarizes, explains, and makes accessible a large number of more specific experiences. The concepts developed thus far--the "raw material" for a principle on oscillations--are:

### Subsystem

A collection of rates, levels, and auxiliaries within a system with one input and one output, which, for the purpose of analyzing the behavior of the overall system, is characterized only by its input-output characteristics; internal structure per se is not of major concern in the analysis.

### Phase-lag subsystem

A subsystem (possessing one or more levels) that produces a phase lag between its input and its output, so that when the input reaches its steady-state value, the output reaches its steady-state value only later.

This concept describes both integrators and first order delays, as well as more complex structures. The concept of phase-lag subsystems allows us to aggregate above the level of individual levels and rates, and thus simplify our thinking about oscillation. The concept also allows us to begin to make the coupling between individual parameters and overall behavior, by asking how a parameter change affects the propagation of a disturbance through a particular subsystem.

### Loop formed of phase-lag subsystems

Several phase-lag subsystems, each of whose inputs is the output of another phase-lag subsystem, with each output being used as input only once.

The concept of a loop formed of subsystems underlies the explanation of oscillation in terms of departures from equilibrium: If one phase-lag subsystem is out of equilibrium, it will cause the next subsystem around the loop to move away from its own equilibrium, so that even if the first subsystem returns to equilibrium, the next subsystem will be out of equilibrium. In turn, the next subsystem after that will begin to move out of equilibrium, and the condition of disequilibrium moves around the loop. This chain of events is formalized in the next two concepts.

### Disturbance

A condition where one or more levels in a system are different from their respective steady-state values. The steady-state values can be either constant, or varying as a function of another behavior mode, such as exponential growth, or oscillations of a longer period.

### Disturbance propagating around a loop formed of phase-lag subsystems

Behavior in which a disturbance away from steady-state values in one subsystem disturbs the next subsystem around the loop away from its steady-state values, and so on, so that a pattern of disturbances moves around the loop.\*16\*

The concept of disturbances propagating ties together the explanation in terms of departures from equilibrium with the mathematical procedures where one first analyzes the open loop propagation of a sinusoid, and then somehow deduces the closed-loop behavior. One instance of this

-----

\*16\*Obviously, two or more subsystems are necessary for a disturbance to have some place to propagate to; a first-order system cannot generate endogenous oscillations. If ever the level reaches its equilibrium value, no further movement is possible, because the whole system (of one level) is at equilibrium. For further discussion, see Mass and Senge, op. cit.

procedure is shown in Figure 2-4; more general methods are available.\*17\*

### Comparable disequilibrium states

Two total states of a system (possibly at different points in time) where the disturbances of each of the levels from their steady-state values are proportional to one another. For systems that go to equilibrium in steady-state, comparable disequilibrium states in state space lie on the same ray from the equilibrium point.\*18\*

The concept of comparable disequilibrium states allows one to utilize the powerful geometric intuition inherent in visualizing a state space. Thus, for example, the concept of comparable states puts damped, expanding, and constant-amplitude oscillations into a uniform framework.

The concepts discussed above seem to embody most of the oscillatory phenomena exhibited thus far. It remains only to tie them all together by stating a principle which gives the relationships among

-----

\*17\*One method is a Nichols chart which converts a graph of open-loop phase and gain at various frequencies to a graph of closed-loop phase and gain. It is thus a simple and elegant way of arriving at closed-loop characteristics of a system. Chapter 6 discusses the Nichols chart as a possible adjunct to the materials presented here. Another means of deducing closed-loop response from open-loop response is with transfer-functions, which can be used to perform the same operation as the Nichols chart, but mathematically instead of graphically.

\*18\*For systems with a dynamic steady-state, the geometry is slightly more complicated. Two total states are comparable disequilibrium states if the vectors representing the disturbance away from steady state (i.e. the total state minus the nominal steady-state state for that time) are parallel.



the concepts:

Principle on the origin of oscillations

Oscillations occur when a disturbance propagates around a loop formed of two or more phase-lag subsystems to return the system to a state comparable to the initial disequilibrium state.

There are two types of oscillatory system that do not fit neatly into the descriptive framework established by the principle. One such class of systems contains approximate differentiators or other non-phase-lag subsystems that play an important role in the oscillation. Section 6.3 discusses such systems as the subject for further research. The other class of systems are agglomerations of individual systems, each of which has some tendency to oscillate. There are many possible loops around which disturbances can propagate, each with a characteristic natural period. For example, some electric power networks, the head of a drum, and a waterbed all fall into this category. Such systems (unless disturbed in highly special ways) will show complex and irregular fluctuations, corresponding to a multiplicity of oscillatory behavior modes.

What are the implications of the principle on the origin of oscillations? How can one use it as a practical modeling tool? First and foremost, the principle tells us that when we are looking for the cause of regular fluctuations, a loop with two or more phase-lag subsystems with response times on the same order of magnitude may be the cause of them.

Additionally, the principle gives us some insight into a very common property of complex oscillatory systems: one or two feedback

loops often dominate the behavior despite the presence of thousands of others. (Here, a loop dominates the behavior in the sense that if the loop is disconnected or substantially altered, the behavior mode also changes substantially.) Thinking in terms of the propagation of a disturbance, the reason is fairly clear: We know that a disturbance will propagate longest around the loop that propagates the disturbance most strongly (and remember that "strongly" has to do with both the phase and gain with which a sinusoid is propagated around a loop.) It seems unlikely that there would be several separate loops around which a disturbance could propagate, each with about the same time constant of delay as the others. More likely, one or two loops propagate the disturbance markedly better than the other loops present in the system. By analogy, we can think of a large number of rooms or cubicles, some connected to each other by a door, just as the various subsystems of a system are interconnected by material flows and information links. Within that "system" of cubicles, there are probably only one or two shortest paths around which one can rapidly and easily run in a circle. The other paths tend to be roundabout, and not as fast for running, just as many loops in a feedback system do not propagate disturbances very well. One example of this type of loop dominance occurs when one shouts into a valley or canyon whose shape is somewhat contorted and irregular. Often an echo results, which is a form of oscillation where a distinctive disturbance travels back and forth, to give periodic but very nonsinusoidal oscillatory behavior, which is damped: HELLO...Hello...hello.... One path along which the sound travels dominates the behavior, to result in oscillations of a single frequency, simply because that

two-way path is the loop best able to propagate the sound. Sometimes, one can hear echoes not at the dominant frequency, but they usually fade much more quickly than the main echo. These quickly-fading oscillations arise from paths or loops that cannot propagate the sound as well as the dominant loop.

The description above of loop dominance in oscillation gives one a guideline for considering parametric and structural changes intended to alter the loop dominance. One instance where one would want to know how to change loop dominance is in designing policies to stabilize an oscillatory system. The description above implies that a very efficient way to stabilize a system's oscillations is to identify the dominant loops and reduce their ability to propagate disturbances, through either parameteric or structural changes.

Prior to policy testing, one often investigates the sensitivity of the behavior to parameter changes--in other words, are there alternative sets of realistic parameters capable of yielding behavior that results from a different set of dominant loops? Again, the principle and the discussion that follows it implies that one should answer such questions in terms of changing a loop's ability to propagate disturbances, which is certainly a much more focused beginning point than picking parameters at random to test for sensitivity.

Further Examples: Wire Memory and Glucose Regulation. In the early days of computers before the development of magnetic core memory, one form of computer memory in effect used echoes to preserve and "remember" information: instead of having an echo that preserves a HELLO...Hello...hello..., the echo preserves the value of a binary number: 11001001...11001001...11001001.... Figure 2-16 shows schematically the physical equipment used to preserve this complex oscillation. (The actual device was rendered more compact by using a coil of wire instead of a loop; the principle of operation, however, is exactly the same.) The figure shows a loop of wire, to which an electromechanical device can apply a sharp twist. The twist does not twist the whole wire around and then back as a unit; the twist must propagate around the wire, just like a sound wave, so a very short section of wire twists, twists back, and is ready for the next twist to come through. One can start a very complex echo around the loop to represent a binary number: twist for a 1, and no twist for a zero. When the twist reaches the terminal end of the wire, a sensor detects the twists, and causes the electromechanical twister to do the same twists at the beginning of the wire as are being detected at the end. The effect of this recirculation is to cause the same pattern of disturbances to propagate endlessly around the loop. (Naturally, in a working computer system, there are provisions for temporarily blocking the recirculation in order to allow the electromechanical twister to put new information into the memory.)

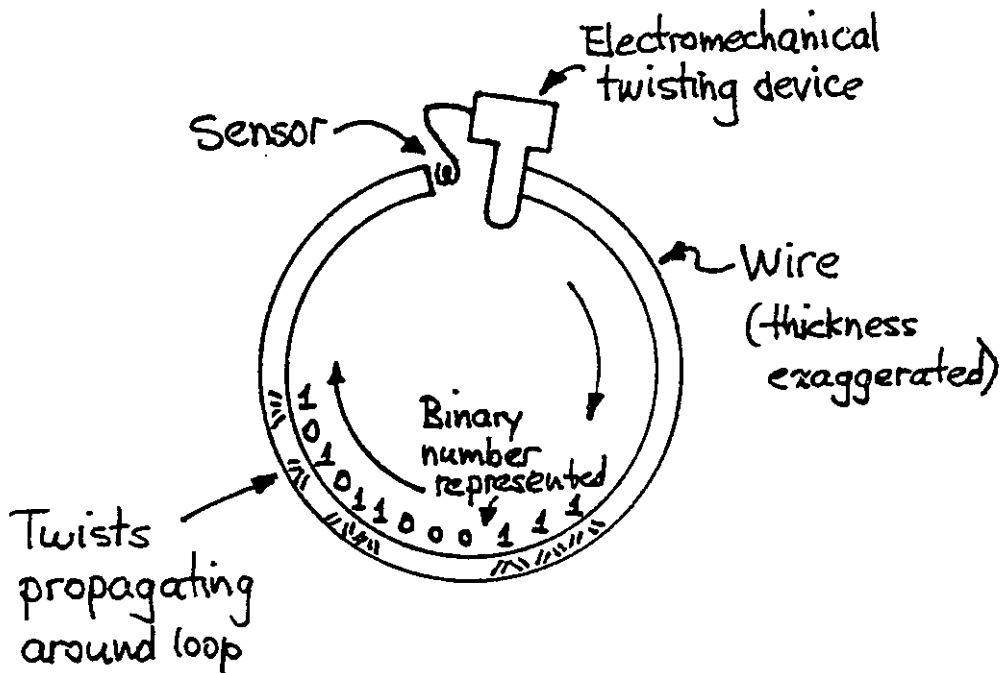


Figure 2-16.  
WIRE MEMORY

As the final example, we will consider the body's system for controlling the sugar content of the blood.\*19\* There is only one type of sugar that circulates in the blood, called glucose. Figure 2-17 shows a simplified flow diagram of the system that controls the level of glucose in the blood. The amount of glucose  $G$  in the blood is a level altered by the rate of change in glucose  $CG$ , which aggregates the release of glucose from the liver and uptake of glucose by most organs

---

\*19\*This example is a simplified version of the system described in Foster, Richard O., "The Dynamics of Blood-Sugar Regulation," unpublished M.S.E.E. thesis (Cambridge, Mass.: MIT, 1970).

in the body. The change in glucose CG is determined by the effects of a hormone that circulates within the blood called insulin.\*20\* Insulin causes both an increase in glucose uptake by organs and a decrease in glucose release by the liver, so that insulin causes glucose to disappear from the bloodstream. The insulin concentration IC does not influence CG instantaneously; there is a delay, represented in the level of physiologically effective insulin concentration PEIC, which influences change in glucose CG with a negative polarity. The insulin concentration IC itself depends on the amount of insulin I in the blood and the volume of blood in which the insulin circulates. Insulin is broken down in the liver and other places, so it has a characteristic half-life. Thus, the level of insulin I is depleted by the rate of insulin degradation ID that depends on the amount of insulin I and the time constant of insulin removal. Insulin enters the bloodstream by being released from the pancreas, a small organ near the stomach and intestines. This insulin release IR responds to high glucose concentration GC.\*21\* Thus, insulin release IR closes a negative loop running through the levels of insulin I, physiologically effective insulin concentration PEIC, and glucose G.

-----

\*20\*Insulin is the medication taken by diabetics when their bodies are no longer able to secrete it internally. For further discussion, see Foster, ibid.

\*21\*The actual system is more complex than indicated by the flow diagram. One of the omitted complexities concerns the stock of insulin available in the pancreas for secretion. If this stock is exhausted by a sudden influx of glucose, the response of insulin release IR to glucose concentration GC changes.

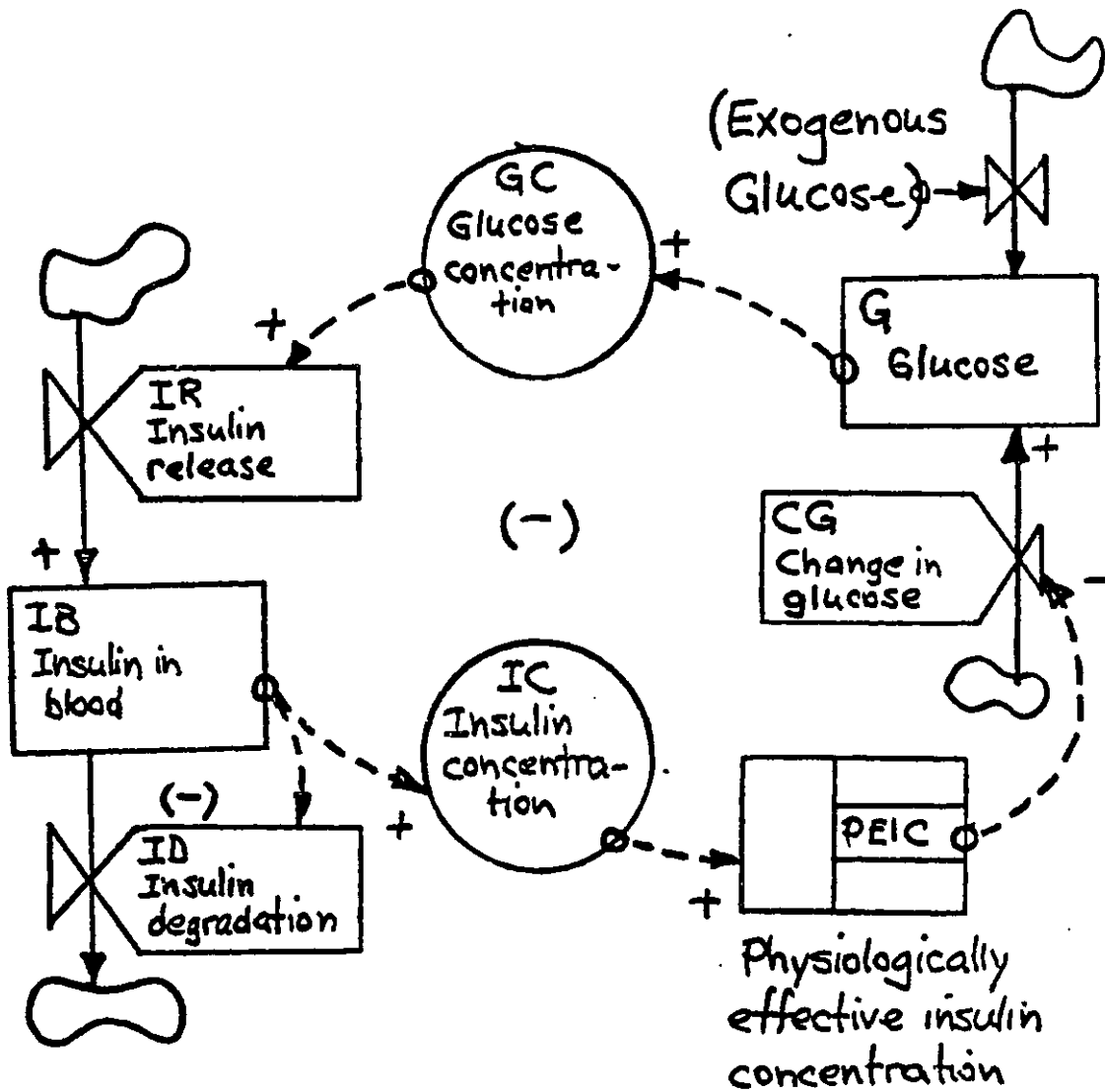


Figure 2-17.  
GLUCOSE-INSULIN SYSTEM

One can influence the amount of glucose in the blood by either injecting glucose directly, or ingesting sugary foods such as candy bars. If sugary foods are eaten on an empty stomach, they digest quickly and provide almost an impulse of glucose into the blood. Figure 2-18 shows a sketch of the response of the glucose-insulin system to such a disturbance. As the reader no doubt expected, the glucose concentration  $GC$  rises rapidly at first due to the influx of glucose from outside the system. This disturbance causes the insulin concentration  $IC$  to build up and removes glucose from the bloodstream, until glucose returns to its steady-state value, but insulin is still disturbed above its steady-state value. The disturbance continues to propagate around the loop, until glucose undershoots (i.e., is disturbed below) its steady-state value, although the magnitude of the disturbance is much smaller than the initial disturbance. The undershoot in glucose is noticeable in real life: if a person is hungry, eating a candy bar makes him or her feel much better, due to the more-than-adequate supply of sugar in the bloodstream. However, after about an hour, the blood-sugar concentration dips lower than ever, leading to a feeling of weakness, dizziness, and apathy. (This overshoot can be eliminated by vigorous exercise, which prevents glucose from building up in the blood, or by eating food that is less digestible than the candy bar, so that the glucose is released into the blood in smaller amounts over a longer period of time.)



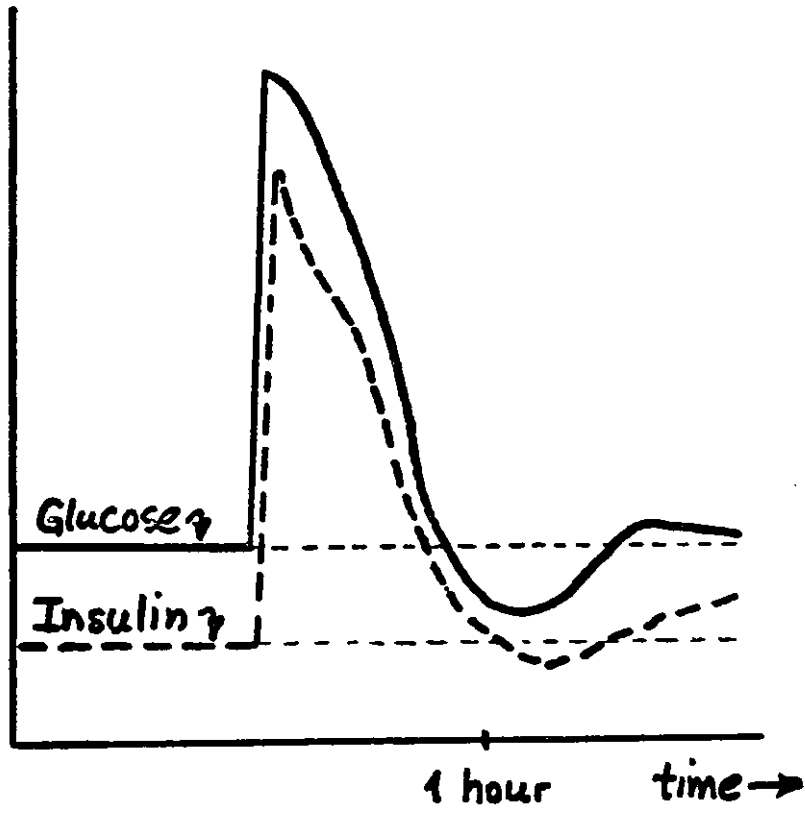


Figure 2-18.  
RESPONSE OF GLUCOSE-INSULIN SYSTEM

## 2.2. DAMPING

The word "damping" usually denotes both an action or device intended to reduce the magnitude of an oscillation, and also the resulting oscillation with diminishing magnitude. In one fell swoop of semantics, the distinction between structure and behavior is erased. For the remainder of this thesis, I would like to let "damping" denote only a type of behavior or behavior change: oscillations whose magnitudes are diminished over time. A first-order negative loop around a level (often called a "damping loop" because it so often damps the oscillatory behavior) I will call a minor negative loop. To explore the structural features that result in damped behavior, let us begin with a simple example.

An Example: A Damped Spring-Mass Oscillation. The spring-mass system as described in the previous section oscillates forever with constant magnitude. In real life, no such perpetual-motion systems have been discovered to date. Mechanical resistance (friction), electrical resistance, and fluid resistance all dissipate the energy of the oscillating system, and return it to an equilibrium eventually. One source of energy loss in the spring-mass system is friction between the mass and the support upon which it slides. In terms of the equations described in the previous section, friction exerts a force on the mass:

$A.KL=(FS.K+FF.K+FD.K)/(W/G)$  3, R  
M=10 3.1, C  
A - ACCELERATION (FEET/SECOND/SECOND)  
FS - FORCE FROM SPRING (POUNDS)  
FF - FORCE FROM FRICTION (POUNDS)  
FD - FORCE FROM DRIVING (POUNDS)  
W - WEIGHT (POUNDS)  
G - GRAVITATIONAL ACCELERATION (FEET/SECOND/  
SECOND)

The force from friction FF is proportional to the velocity V. The constant of proportionality, the friction coefficient FC, is an empirical result that indicates how much force is required to keep the mass moving at one foot per second on its supporting surface. FC was set equal to zero in the previous section; for the simulation below it is set at 2.5 pounds per foot per second. Because the force from friction FF always works in the opposite direction of the velocity, the equation for FF contains a negative sign:

$FF.K=-FC*V.K$  5, A  
FC=0 5.1, C  
FD - FORCE FROM FRICTION (POUNDS)  
DC - FRICTION COEFFICIENT (POUNDS/FOOT/SECOND)  
V - VELOCITY (FEET/SECOND)

The force of friction FF closes a minor negative loop around velocity V, which decreases FF, which decreases the rate of acceleration A, which decreases the level of velocity V.

Figure 2-19 shows the behavior of the spring-mass system with damping superimposed on the undamped simulation of Figure 2-3. As before, the position P declines to its equilibrium value, overshoots, and reaches a minimum value. But unlike the previous simulation of the spring-mass system, the minimum value is only about half the magnitude of the initial displacement. The maximum excursion of P seems to

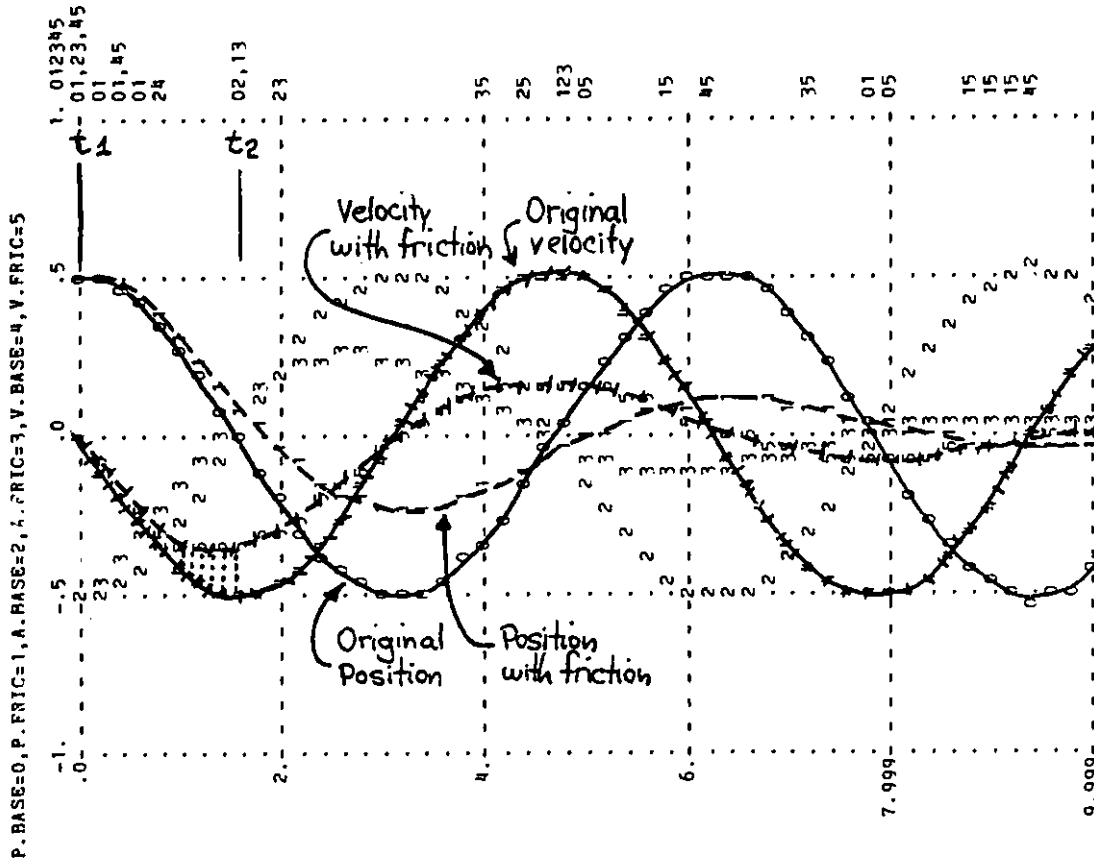


Figure 2-19.  
DAMPED BEHAVIOR OF SPRING-MASS SYSTEM

diminish by half each half-cycle. The other variables show similar diminution. How does the added minor loop cause this change?

Explanation in Terms of Diminution of Disturbances. In the light of the preceding section, one might be tempted to say that adding a minor negative loop damps oscillations by attenuating disturbances.

Figure 2-19 illustrates: the solid lines show how a disturbance in position  $P$  propagates to become a disturbance in velocity  $V$  at time  $t_2$  if there is no minor loop (i.e., when  $V$  is simply minus the integral of  $P$ ). The dotted line between the two curves indicates the diminution in the buildup of velocity  $V$  caused by the minor loop. (In effect, the phase-lag subsystem around the level of velocity  $V$  has been made less of a phase-lag subsystem, because the minor loop reduces the velocity  $V$  to its equilibrium value more quickly.) Thus at the time when position  $P$  equals zero, the disturbance (which will continue to propagate around the loop) is smaller than it would have been without the minor negative loop. Each succeeding half-cycle will further attenuate the disturbance, so the amplitude of the oscillation diminishes exponentially toward zero.

The discussion above shows that the concepts developed in the previous section can be used to explain one kind of damping. Still, the concepts thus far do not provide much basis for detecting exceptions: can one add minor loops that destabilize a system? Also, minor negative loops are not the only structural changes that can cause damping. What characterizes the other, more complex changes? To begin to formulate a more general characterization of structures that cause damping, let us describe oscillation and damping in state space.

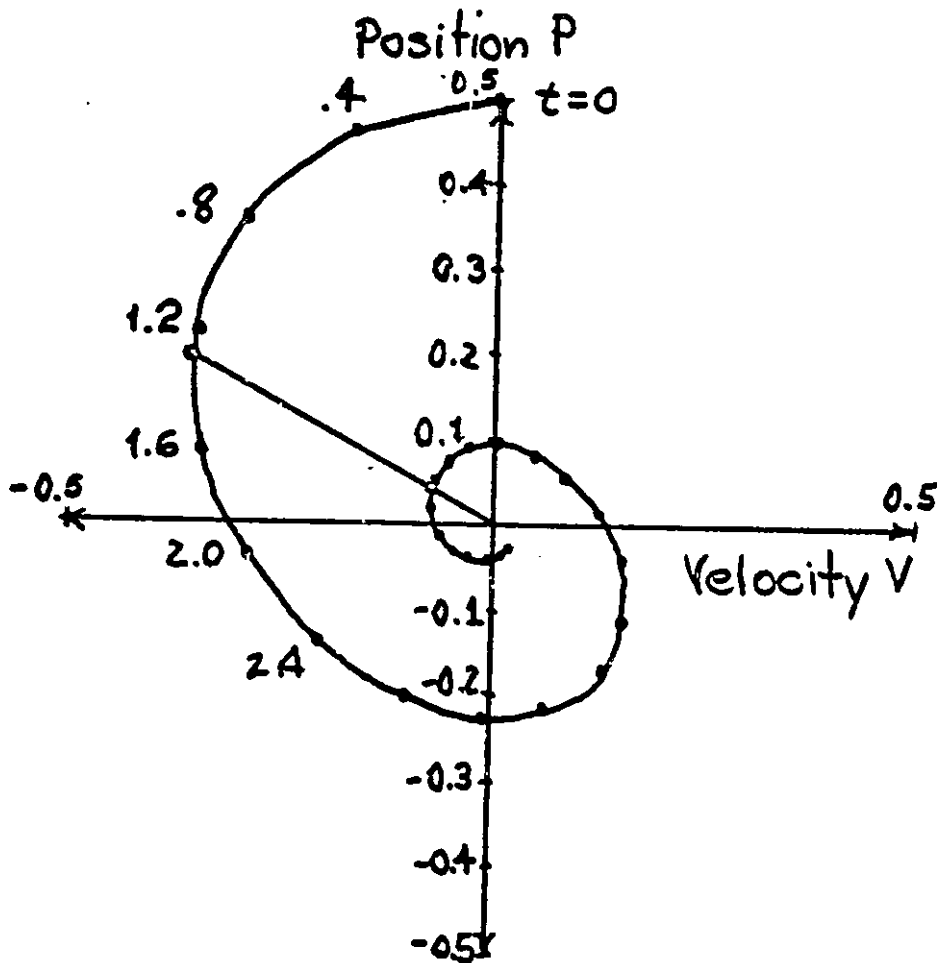


Figure 2-20.  
STATE-SPACE TRAJECTORY OF SPRING-MASS SYSTEM WITH FRICTION

Explanation in State-Space. Figure 2-20 shows a state-space trajectory of the pendulum. The addition of a minor negative loop changes the former circular trajectory into an inward-spiraling trajectory; no longer does the system return to its exact initial conditions on each oscillation. However, each oscillation results in similar subsequent oscillations: any ray from the equilibrium point intersects the

trajectory at several points, each representing a comparable state in relation to the equilibrium state and each causing comparable subsequent trajectories. For example, each half-cycle, the maximum excursion of the position  $P$ , is diminished by about half.

Figure 2-21 shows how the addition of the minor negative loop representing friction changes the trajectory of the system. The solid arrows show the direction and magnitude of changes in the state (as in Figure 2-8). Friction exerts a force proportional to the velocity  $V$ , and in the opposite direction as velocity  $V$ , which changes the acceleration  $A$  (the rate of change of velocity  $V$ ). Thus, friction has the effect of moving the arrows toward the vertical axis. The dotted arrows represent the direction and magnitude of the change in the states when friction is present. When velocity  $V$  is zero, friction exerts no force, so that the arrows emerging from the vertical axis are unchanged. All other arrows move the system in a direction that corresponds to oscillations of smaller magnitude, as indicated by the concentric circular arcs on Figure 2-21.

It is tempting to conclude that a structural change produces damping whenever it tends to diminish the magnitude of the state vector (i.e., the distance from the state to the equilibrium point in state space). Unfortunately, this is roughly equivalent to saying "The fastest way to get somewhere is to head directly at it." As anyone who has driven a car in a large city knows, this is not always the case. Sometimes, the roundabout way is faster. Section 4.2 describes one structural change that both produces damping and tends to increase the

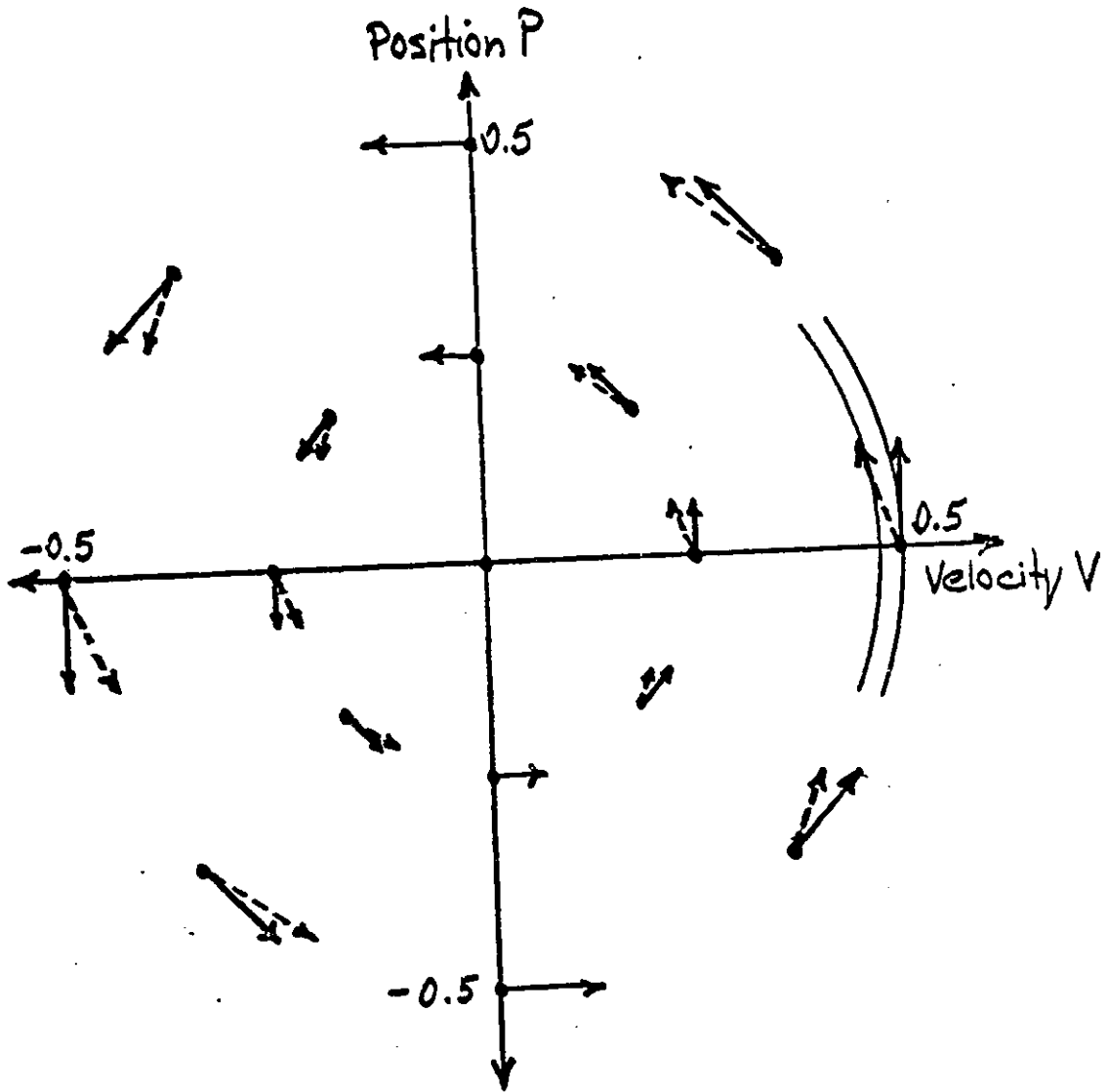


Figure 2-21.  
STATE-SPACE DIAGRAM OF SPRING-MASS SYSTEM WITH FRICTION

magnitude of the state vector at first (the distance between the state and the equilibrium point). Figure 2-22 shows how this can happen.

The solid line in Figure 2-22 represents the trajectory of two levels oscillating without a structural addition that involves movement of a third level. The trajectory therefore lies on the plane specified by the two horizontal axes. The heavy dotted line shows the trajectory



when the structural change is made, so the third level is allowed to vary. The third level increases at first until after a quarter-cycle, it is well above the plane of the two original state variables. The structure of the system is assumed to be such that when the third state variable is not equal to zero, it alters the trajectory of the first two state variables. Figure 2-22 shows the trajectory bending inward, crossing over the original trajectory without the structural change. When the trajectory of the structural change intersects the plane of the first two state variables, the state is closer to the equilibrium point than it would have been had it followed the original trajectory. Thus, although the structural change increased the magnitude of the state vector at first, by the time the two states were comparable (i.e., lying on the same ray in state space) the state resulting from the structural change had a smaller magnitude.

Figure 2-22 suggests one class of structural changes where adding a minor negative loop around a level can decrease the damping. Suppose a minor negative loop were added around the third level, plotted on the vertical axes. This change would tend to drive the third state variable toward the horizontal plane, which would nullify the damping effect of activating the third level. In general, if a minor negative loop diminishes the magnitude of oscillation in a loop that itself contributes to damping, then the added minor loop probably decreases damping.

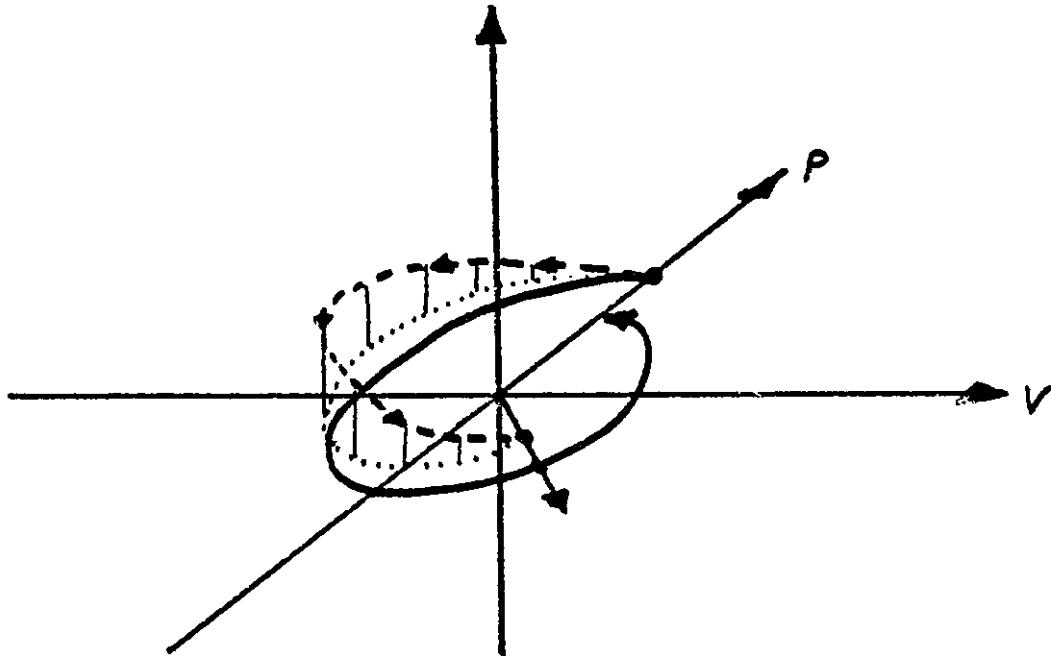


Figure 2-22.  
ROUNABOUT TRAJECTORY IN STATE SPACE

A Principle. The outset of this section provided a very simple definition for damping:

Damping

A type of behavior where the magnitude of oscillations diminishes over time.

By examining oscillations in state space, we saw that a minor negative loop can produce this behavior by directly moving the state toward a state corresponding to oscillations of smaller magnitude. Figure 2-22 indicated how a structural change might produce damping by moving the state in a direction which eventually resulted in a state comparable to the original trajectory, but with smaller magnitude. We can characterize these results in a principle as follows:

Principle on damping

A structural or parametric change produces more damping when it realigns the trajectory toward a future state, comparable to a state on the original trajectory, that corresponds to oscillations of smaller magnitude. The simplest structural change that potentially can produce such a realignment is the addition of a minor negative loop. More complex changes can produce damping as well.

In effect, this principle expands the definition of damping to include both cause and effect. The effect is a diminished tendency to continue oscillating, and the cause is a realignment of the system's trajectory in state space. Unfortunately, since a very large variety of types of parametric and structural changes can increase damping, one can characterize them only with the very general concept of "realigning the trajectory." Even this very general principle can be of assistance in identifying structural changes that will increase damping.

One type of structural change that can increase damping is where the change increases the rate of movement of a single state variable toward its steady-state value, which can diminish the propagation of a disturbance through that level, so that the magnitude of future oscillations decreases. The old standby, the minor negative loop is one such structural change. But one can also search for changes that move two or more levels toward their equilibrium values. The rabbit-coyote example in the next subsection gives a change of this type, where a single parameter change (corresponding to a single policy in the real system) alters the rate-of-change for both the level of rabbits and the level of coyotes. Perhaps identifying such multiple-effect policies is very

difficult, but the effects on several levels at once can have a dramatic effect on the overall oscillation, as the rabbit-coyote example will illustrate.

The principle on damping suggests a final type of structural change that will produce damping: changes whose effect on the movements of the levels is not immediate, but is delayed, so that only some future state shows a disturbance of smaller magnitude. Sections 3.3 and 4.2 will give examples where the effect of adding a minor loop with a delay depends entirely on the delay time. Further discussion on the implications of delays for damping should be deferred until those sections develop specific examples.

It must be admitted that the principle on damping is not a directly-applicable rule of thumb that says "if you change this parameter, that behavior will result." This material in Chapter 2, however, is the foundation for the very specific rules of thumb in Chapters 3 and 4. Perhaps the most important implication of the principle on damping is that one cannot pass over damping lightly, for the phenomenon in general is more complex than is commonly acknowledged, as the following examples illustrate.

Another Example: Influence of Backlog on Ordering. Adding one minor negative loop can increase damping. Adding a second or third or whatever can again increase the damping. We can locate another minor negative loop in real employment-backlog systems as follows: The employment-backlog system in Section 2.1 oscillates due to out-of-phase movements of the level of order backlog B and the level of employment E, as shown in Figure 2-14. The negative loop involving employment E is one means by which the magnitude of the backlog B is controlled. In real life, many factors compensate for high or low backlogs. Figure 2-26 gives a representation of one such factor: an attempt by the individual firms placing the orders to avoid having placed an excessive or insufficient number of unfilled orders. If someone places too many orders, they will be filled later, and the person will wind up with too many units. Therefore, the individuals or firms placing the orders attempt to maintain some desired backlog for ordering DBO, by making a correction for backlog on ordering CBO, which results in an adjusted order rate AOR. (The order rate OR itself is modified by the multiplier from delay on ordering MDO, which is set at 1.0 in this chapter.) These relationships form a minor negative loop around the backlog B.

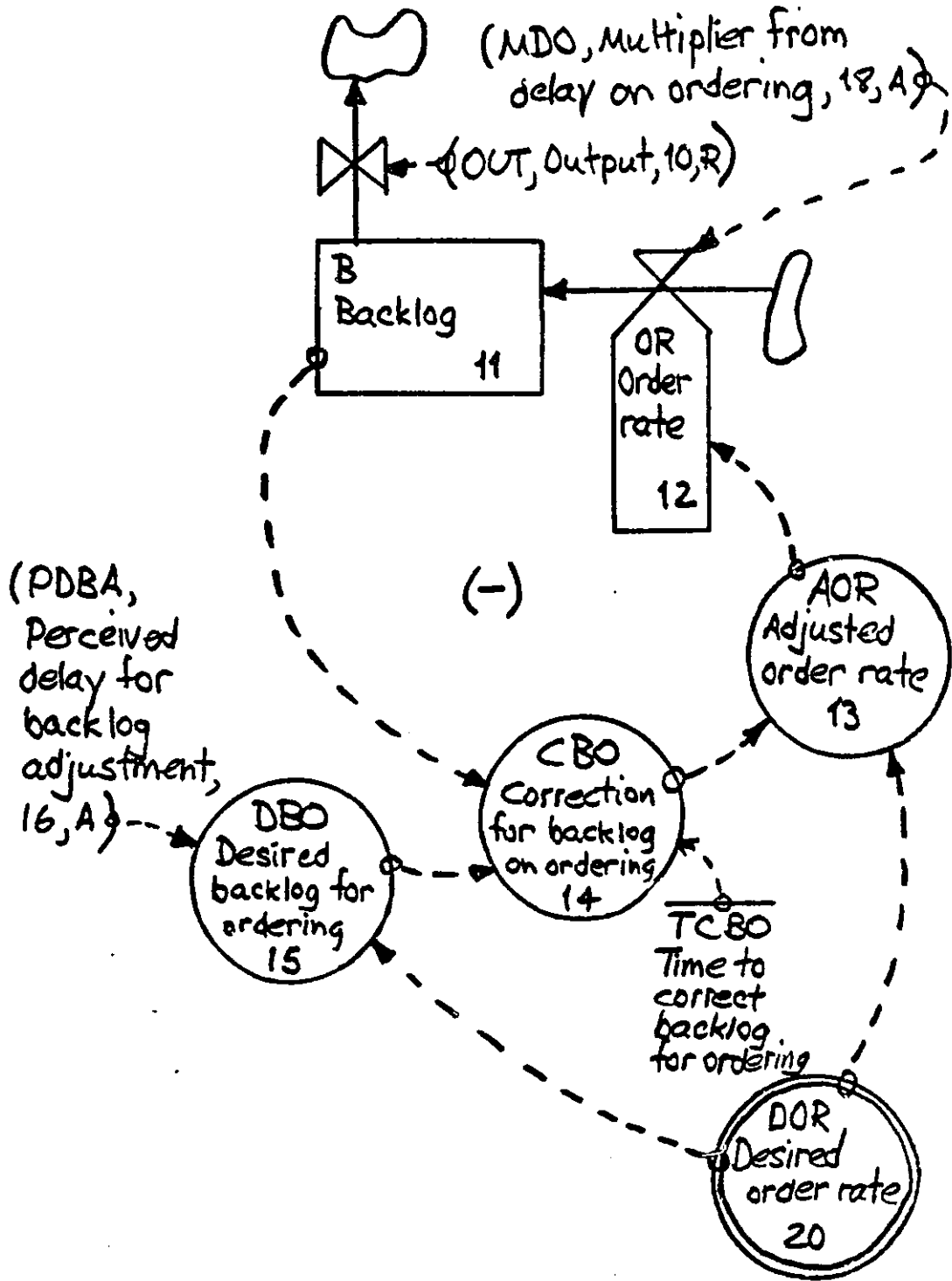


Figure 2-23.  
INFLUENCE OF BACKLOG ON ORDERING

The order rate OR is based on the adjusted order rate AOR, and modulated by the multiplier from delay on ordering MDO, which represents the effect of long-term unavailability of the firm's product on orders for that product. MDO is always set equal to 1.0 in this chapter.

$$\text{OR.KL} = (\text{AOR.K})(\text{MDO.K}) \quad 9, R$$

OR - ORDER RATE (UNITS/YEAR)  
AOR - ADJUSTED ORDER RATE (UNITS/YEAR)  
MDO - MULTIPLIER FROM DELAY ON ORDERING (DIMENSIONLESS)

The adjusted order rate AOR modifies the desired order rate DOR with the correction for backlog on ordering CBO. DOR is constant at 1320 units per year.

$$\text{AOR.K} = \text{DOR.K} + \text{CBO.K} \quad 10, A$$

AOR - ADJUSTED ORDER RATE (UNITS/YEAR)  
DOR - DESIRED ORDER RATE (UNITS/YEAR)  
CBO - CORRECTION FOR BACKLOG ON ORDERING (UNITS/YEAR)

The correction for backlog on ordering CBO represents the effect of a discrepancy between the desired and actual backlog B on ordering. In this simple formulation, the discrepancy alters the stream of orders through a time constant, the time to correct backlog for ordering TCBO.

$$\text{CBO.K} = (\text{DBO.K} - \text{B.K}) / \text{TCBO} \quad 11, A$$
$$\text{TCBO} = 1E11 \quad 11.1, C$$

CBO - CORRECTION FOR BACKLOG ON ORDERING (UNITS/YEAR)  
DBO - DESIRED BACKLOG FOR ORDERING (UNITS)  
B - BACKLOG (UNITS)  
TCBO - TIME TO CORRECT BACKLOG FOR ORDERING (YEARS)

The desired backlog for ordering DBO is the product of the desired order rate DOR and the perceived delay for backlog adjustment PDBA. Thus, DBO represents the size of the order backlog B necessary to keep units arriving at the desired order rate DOR, assuming that the only information about production of the units available to the orderers is the delay between placing an order and having it filled. This delay is smoothed into the perceived delay for backlog adjustment PDBA. In this chapter, PDBA is held constant.

$$\text{DBO.K} = \text{PDBA.K} * \text{DOR.K} \quad 12, A$$

DBO - DESIRED BACKLOG FOR ORDERING (UNITS)  
PDBA - PERCEIVED DELAY FOR BACKLOG ADJUSTMENT (YEARS)  
DOR - DESIRED ORDER RATE (UNITS/YEAR)

Figure 2-24 plots the state-space trajectory of the employment-backlog system when backlog influences ordering. In the previous simulation of the employment-backlog system, the time to correct backlog for ordering TCBO was set at a very large magnitude, so that the correction for backlog on ordering CBO was effectively zero at all times. In Figure 2-24 (and the plot versus time in Figure 2-25), the time to correct backlog for ordering TCBO is set at 0.3 years.

Figure 2-24 indicates that the employment-backlog system still possesses a slight tendency to oscillate, even with a minor loop added around the backlog B. The oscillation begins with the initially-excessive employment E causing the firm's output OUT to exceed the order rate OR, so that backlog B begins to fall. With a smaller backlog B, desired employment DE and employment E also fall. The trajectory with the time to correct backlog for ordering TCBO equal to



0.3 begins to diverge from the original trajectory: when the backlog B falls below the desired backlog for ordering DBO (constant at 660 units in this simulation), the correction for backlog on ordering CBO increases the order rate OR, so that backlog B does not drop as fast as in the original simulation. In other words, the minor negative loop around backlog B tends to reduce the disturbance in backlog B, so that the state is always moving toward a state corresponding to a smaller magnitude of oscillation. The effect of the minor loop is quite apparent by time 0.6, when backlog B in the original trajectory has just reached a minimum value at 605 units (55 units below the steady-state value), and in the trajectory with TCBO = 0.3 backlog B has already gone through its minimum value (of only about 629 units, or 31 units below its steady-state value) and begun to increase again toward equilibrium, having already reached 640 units. The minor loop has cut the magnitude of the disturbance in half over one quarter-cycle.

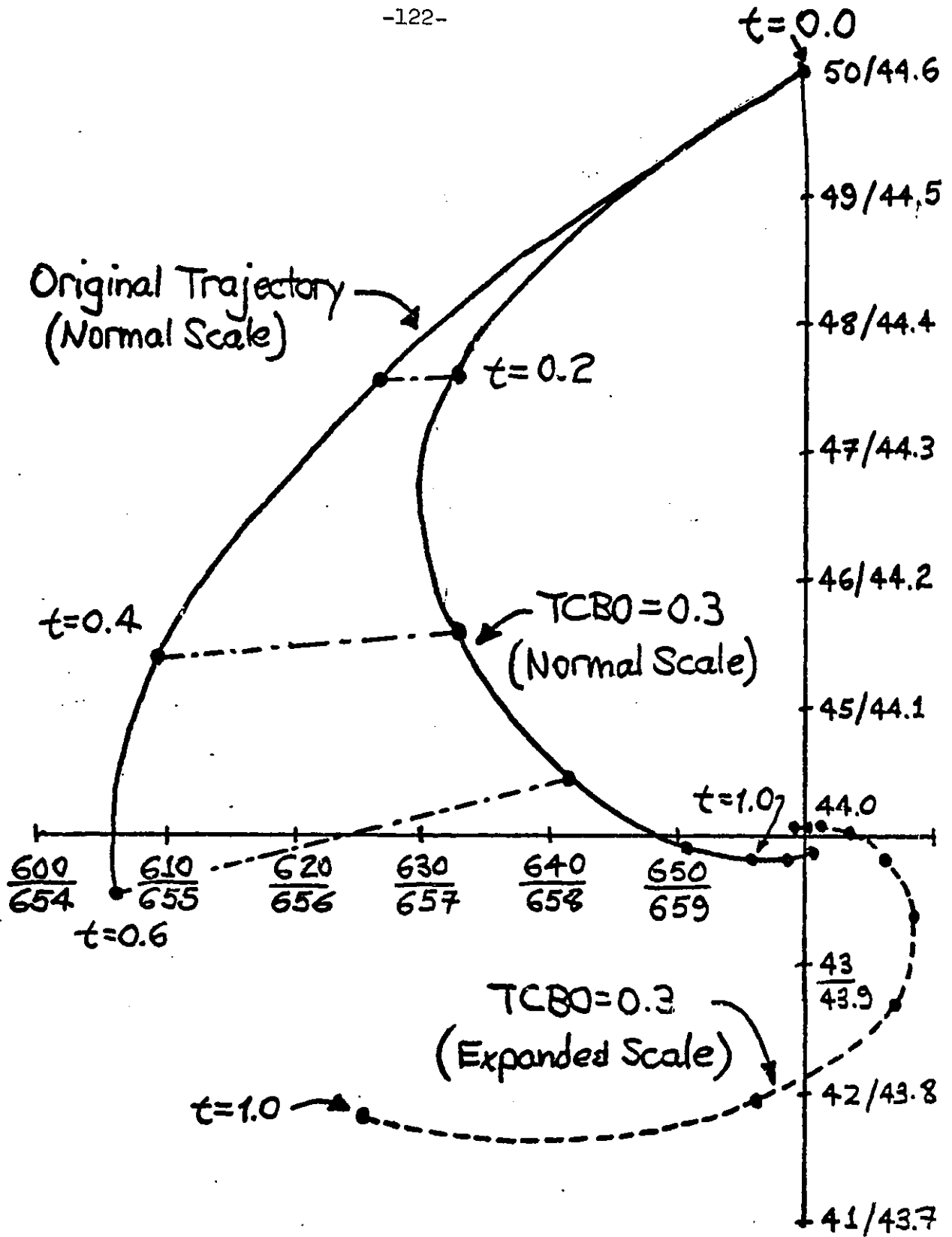


Figure 2-24  
STATE-SPACE TRAJECTORY WITH INFLUENCE OF BACKLOG ON ORDERING

Some continued oscillation is apparent on the state-space plot in Figure 2-24, even though the oscillation cannot be seen on the time plot in Figure 2-25. However, the magnitude of the oscillation is so reduced after a half-cycle that an expanded scale (expanded by a factor of ten) is necessary to see that the trajectory does in fact return to a disequilibrium state comparable to the initial state, i.e., lying on the vertical axis, where backlog B is at its steady-state value, and employment E exceeds its steady-state value.

So again the addition of a minor negative loop increased the damping of a system, by moving the state toward states corresponding to smaller magnitudes of oscillation. What have the added negative loops done to the periods of oscillation? For the spring-mass system, comparing Figures 2-3 and 2-20 reveals that the period increases very slightly from about 6.2 to 6.4 with the addition of a minor loop. For the employment-backlog system, comparing Figures 2-15 and 2-24 reveals that the period decreased from 3.5 years to 2.7 years with the addition of a minor loop. There is no consistent pattern as yet with respect to changing the period; Section 4.2 will provide further discussion when more examples of adding both positive and negative loops have been examined.

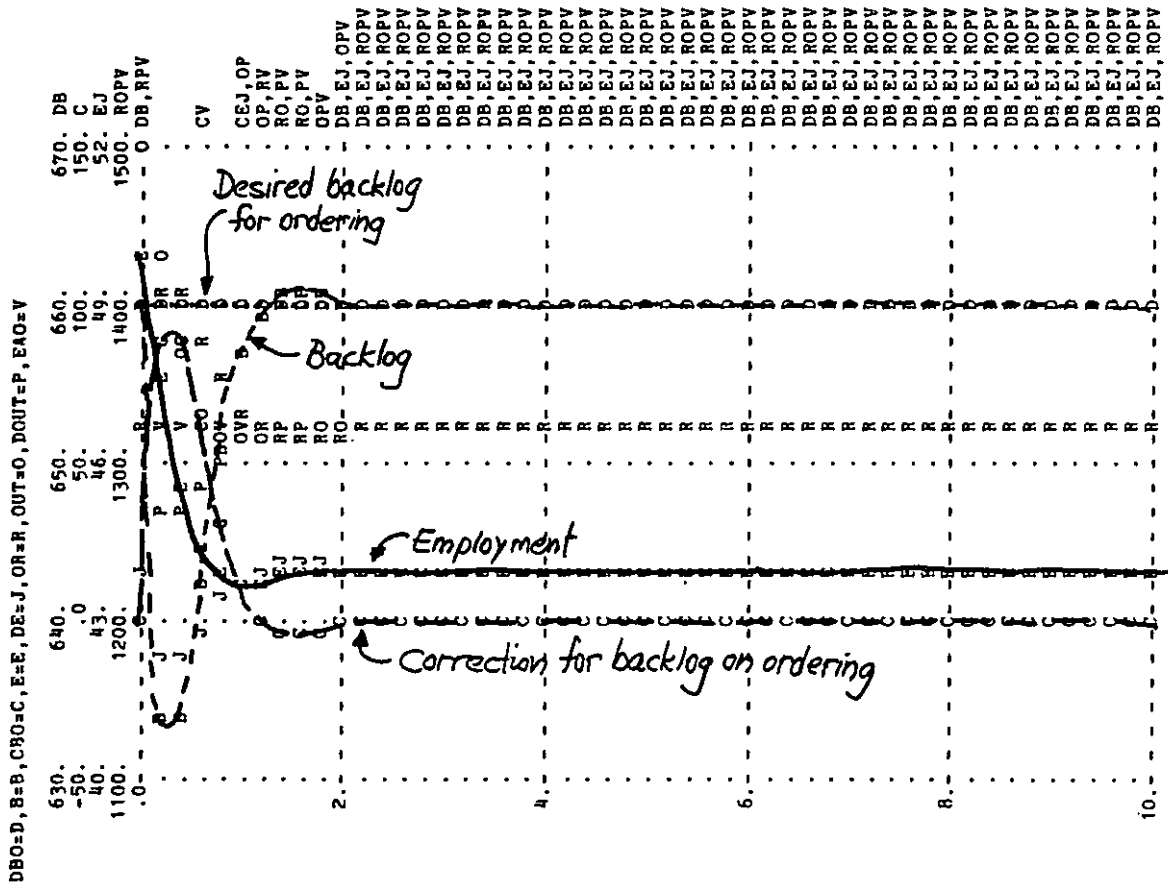


Figure 2-25.  
BEHAVIOR WITH INFLUENCE OF BACKLOG ON ORDERING

A Further Example: Rabbit-Coyote System. Thus far the examples of damping have been produced by adding minor negative loops. Many other less-common types of parametric or structural changes can increase damping as well. One more complex instance of damping occurs in Nathan Forrester's model of an ecological system.\*22\* Figure 2-26

\*22\*Forrester, Nathan B., "A Computer Approach to Environmental System Design--Dynamics of a Predator-Prey Relationship." Privately published by the author, 1969.

shows a flow diagram of the interaction between a coyote population and a rabbit population. Rabbit births  $RB$  are proportional to the population of rabbits  $R$ , but they are diminished when high density of rabbits causes disease, semi-starvation, and low fertility. Similarly, the rate of rabbit fatalities  $RF$  is proportional to the population of rabbits  $R$  but is increased when high rabbit densities cause disease and starvation. High rabbit densities also allow rabbits to be consumed more readily by coyotes. The health of the coyote population largely depends upon how well they eat, so that both coyote fatalities  $CF$  and coyote births  $CB$  depend on the average coyote consumption ratio  $ACCR$ , which measures the average rabbit consumption per coyote.

The structure shown in Figure 2-26 causes periodic population explosions of rabbits  $R$ , as shown in Figure 2-27. When the large rabbit  $R$  population begins to stimulate the growth of the coyote  $C$  population, the number of rabbits  $R$  begins to decline. Finally, a large number of coyotes  $C$  are hunting a vanishingly small population of rabbits  $R$ , so the population of coyotes  $C$  declines to a level low enough to trigger another explosion in the population of rabbits  $R$ .

Interestingly, the oscillation is highly resistant to the addition of a minor negative loop around the level of rabbits  $R$ . Such a loop might correspond to a policy of exterminating rabbits in proportion to the rabbit density. The maximum rate of rabbit extermination with such a policy would occur at the peaks of rabbit population, when the

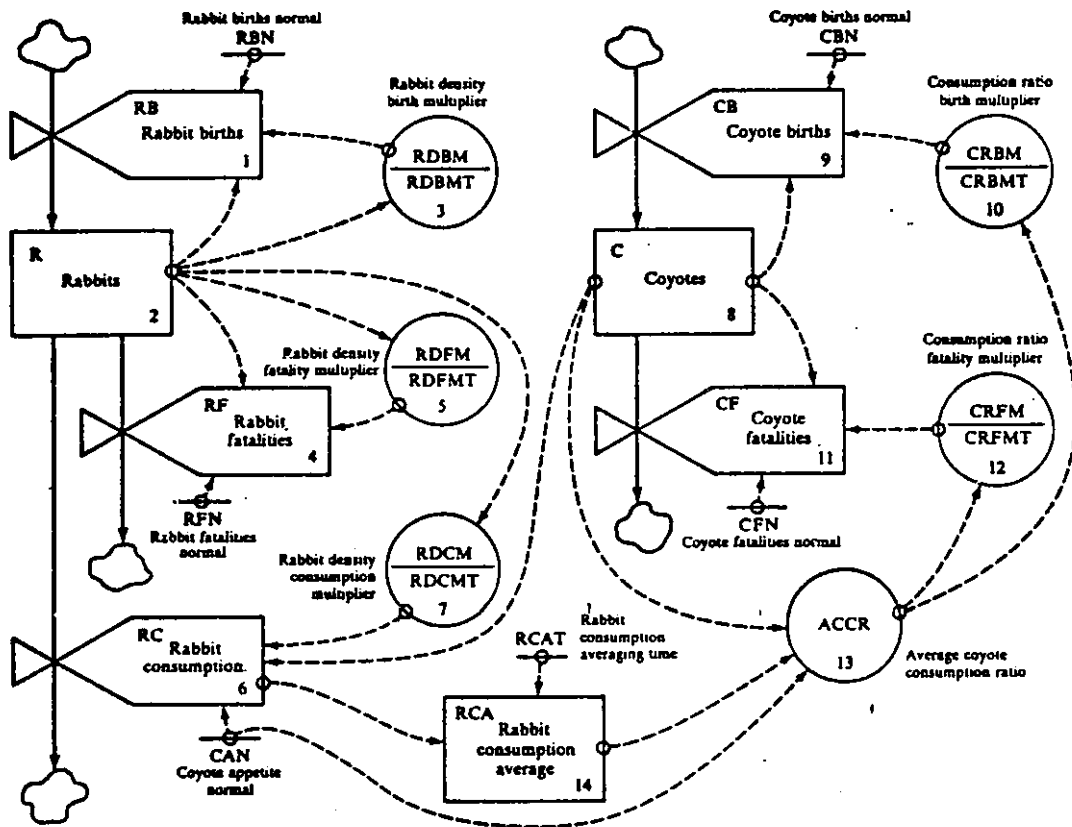


Figure 2-26.  
RABBIT-COYOTE SYSTEM

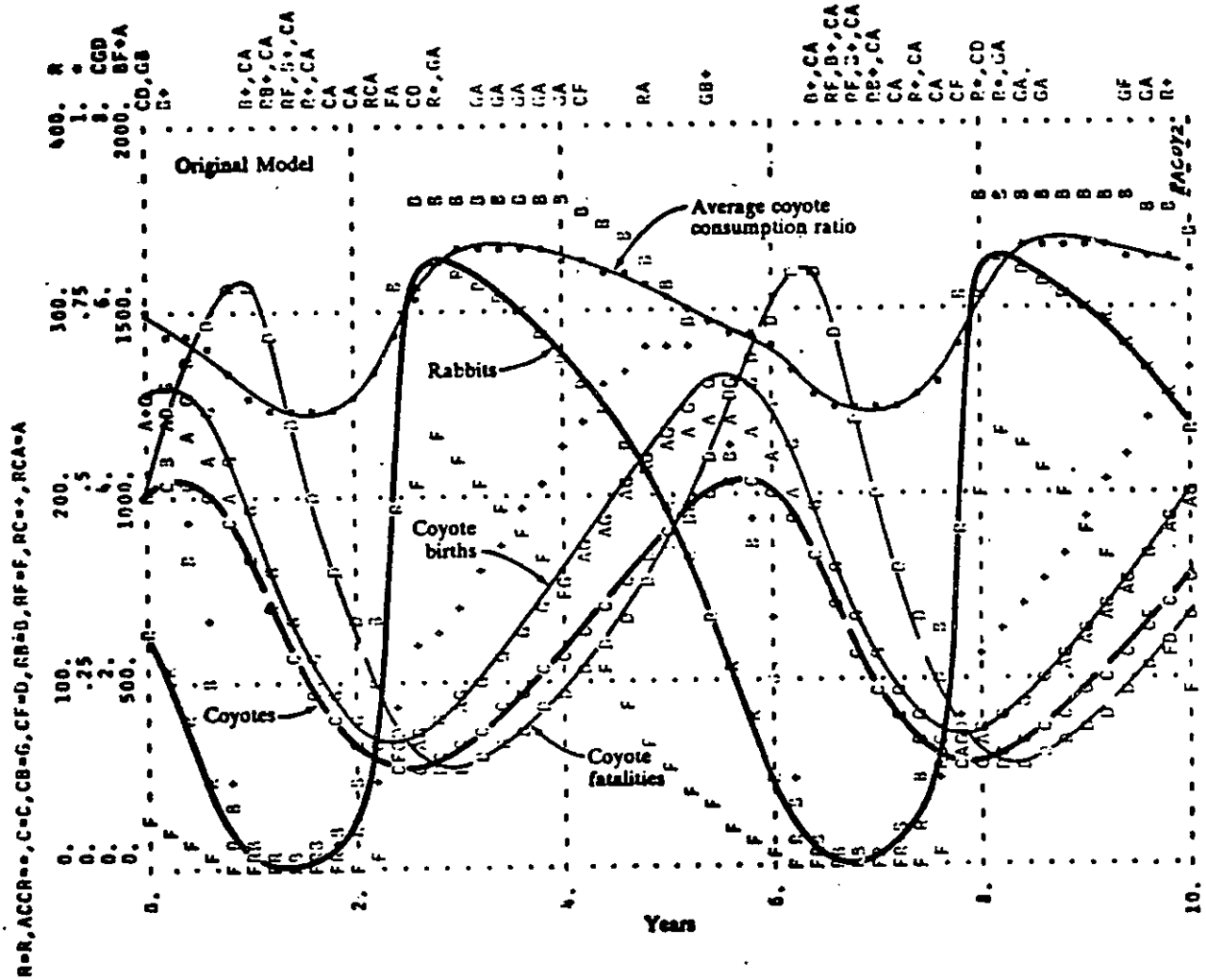


Figure 2-27.  
BEHAVIOR OF RABBIT-COYOTE SYSTEM

positive loop of rabbit population growth is controlled mostly by the negative loops representing disease and starvation of the rabbit population. Thus, the rabbit extermination program abates the maximum rabbit population only slightly.

The other side of the rabbit population cycle--an extremely sparse population--proves much more susceptible to manipulation.

A policy of protecting a small number of rabbits from coyote predation can be represented by shifting the rabbit density consumption multiplier RDCM, so that for a given number of rabbits, coyotes can consume fewer of them. (See the flow diagram in Figure 2-26.) This change affects the behavior primarily during conditions of low rabbit density. The change has two effects: first, the rabbit population is prevented from declining quite so precipitously when there are many coyotes about. Second, because the rabbit consumption per coyote is diminished for any rabbit density, the population of coyotes begins to decline sooner and more gradually when there are few rabbits to be eaten. The effects of this policy change are sketched in the state-space diagram in Figure 2-28. The original trajectory forms a closed loop, with the same cycle repeated again and again.\*23\* The rabbit protection policy moves the rate-of-change vectors inward. Even during non-extreme conditions, rabbit protection has some effect on the rate-of-change vectors. As it turns out, this effect suffices to change the limit-cycle behavior shown

-----

\*23\*In fact, although the sketch does not show it, the cycles are limit cycles, where any disturbance away from equilibrium causes oscillations that increase in magnitude to the amplitude shown in the simulation.



to the damped behavior shown in Figure 2-29.\*24\*

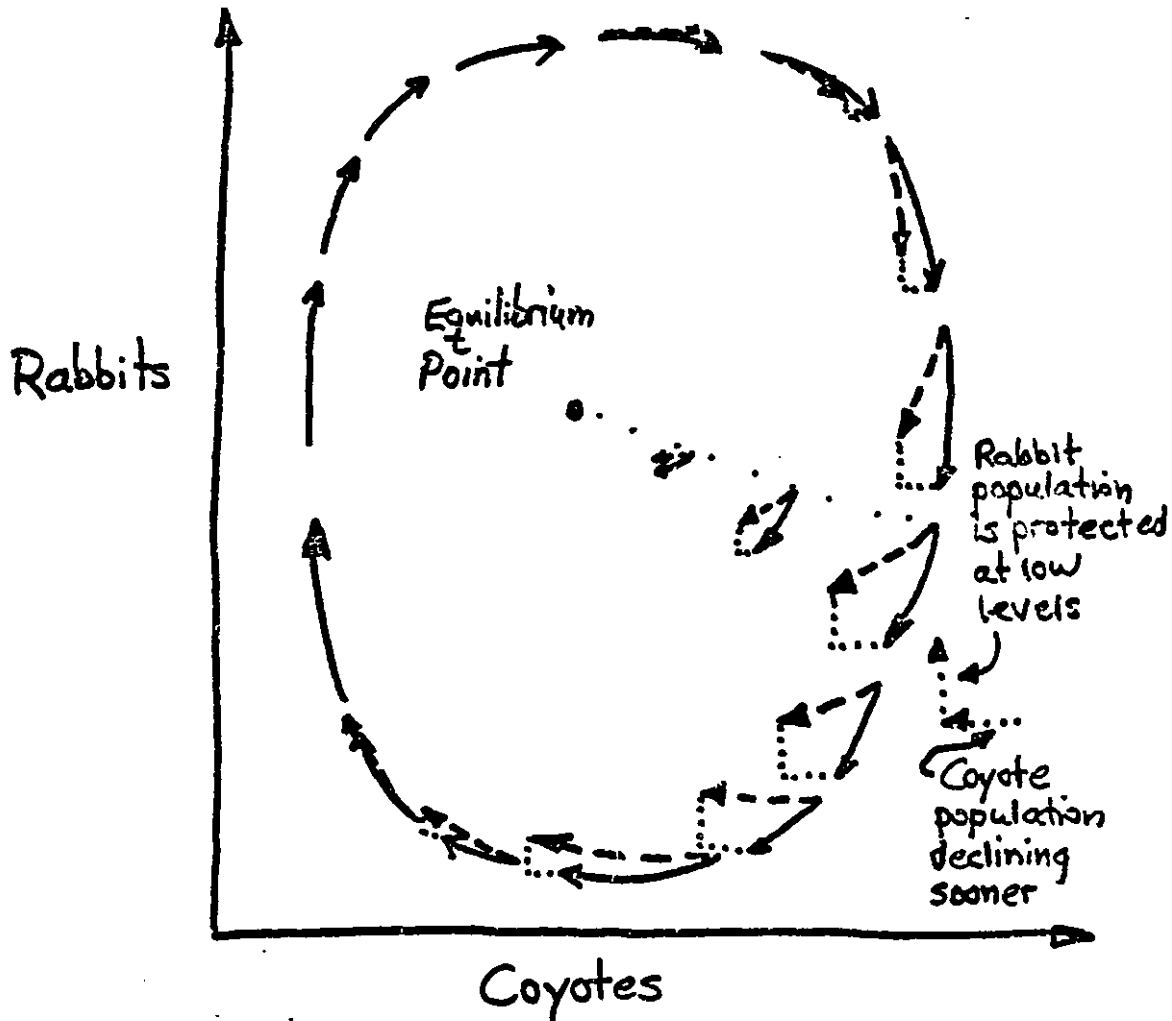


Figure 2-28.

EFFECT OF RABBIT PROTECTION ON STATE-SPACE TRAJECTORY

-----

\*24\* One possible lesson from the rabbit-coyote system is that damping can be altered by changing the manner in which disturbances propagate from one phase-lag system to another. Indeed, as Section 3.1 shows, in the employment-backlog system the parameter TCB (time to correct backlog) controls propagation of disturbances from backlog B to employment E, and increasing TCB (decreasing the gain) increases damping. However, there is a glaring exception to this possible lesson: in the spring-mass system, decreasing the gain from one phase-lag subsystem to another does absolutely nothing to the damping. Chapter 6 discusses this matter of gains between subsystems as an area for future research.

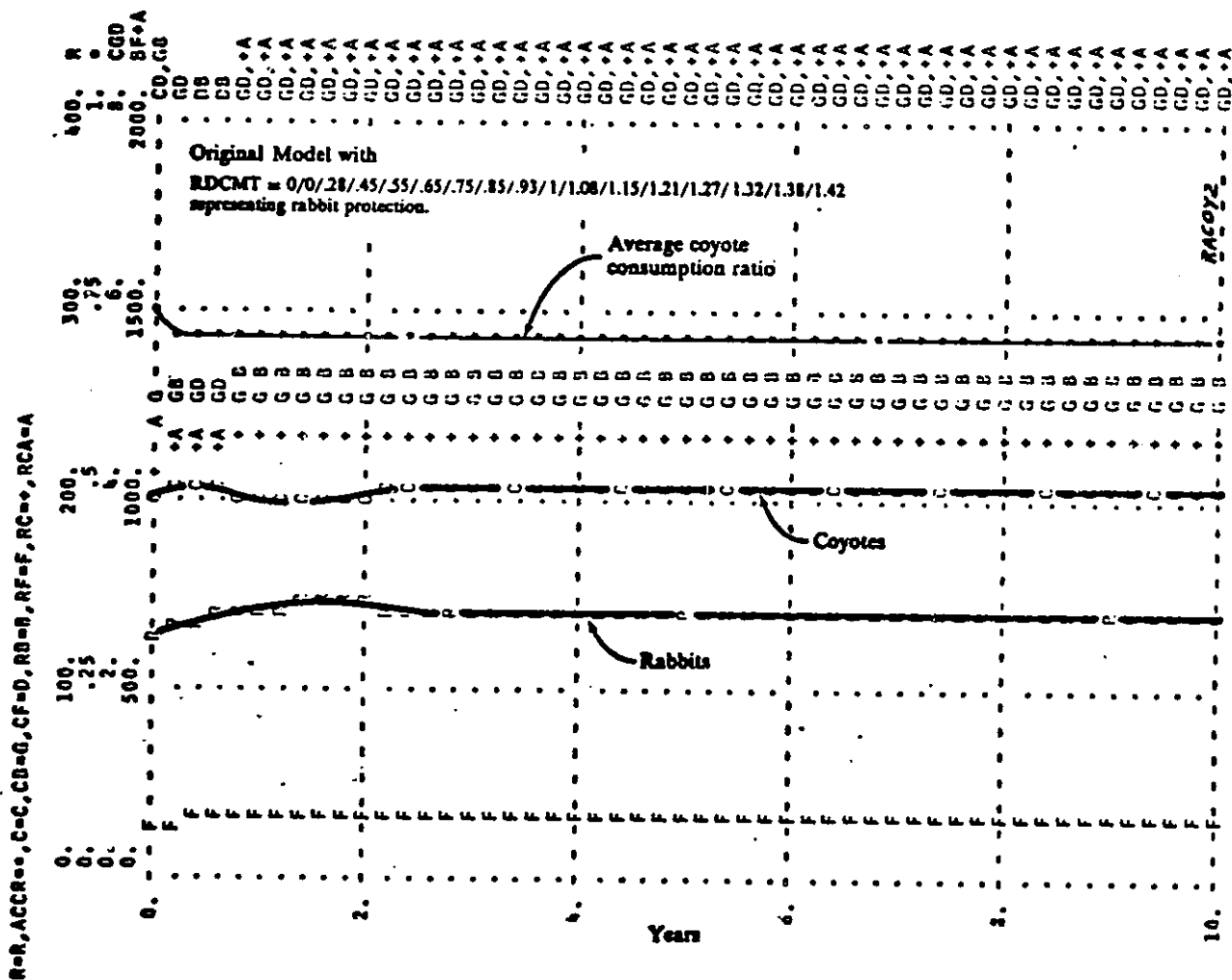


Figure 2-29.  
BEHAVIOR OF RABBIT-COYOTE SYSTEM WITH RABBIT PROTECTION

The rabbit-protection policy was represented in Forrester's model by manipulating a structural feature that prevents coyotes from eating rabbits when there are no rabbits to eat. Such loops--loops that prevent a level from traveling to or below zero--occur frequently in systems, and are frequently omitted from models of those systems. For example, in the employment-backlog system, there is the possibility of

filling orders when no orders are in the backlog to be filled. In real life, a firm even approaching such a situation would probably put its workers on half days, so that the output OUT would decline when the backlog B approached zero. If this phenomenon was modeled, it would further decrease the tendency of the employment-backlog system to oscillate.\*25\*

---

\*25\*Of course added structure does not always reduce a system's tendency to oscillate. As one adds more delays in information links and intervening accumulations in material flows, the stability of the system usually tends to decrease. Also, just as adding minor negative loops to increase model realism often stabilizes the system's oscillations, adding minor positive loops to increase model realism often destabilizes the system's oscillations.

## CHAPTER 3

### FURTHER PRINCIPLES ON OSCILLATIONS

Chapter 2 explained oscillation in terms of a disturbance propagating around a loop of two or more phase-lag subsystems, eventually returning the system to a state comparable to the initial disequilibrium state. Chapter 3 utilizes these concepts (disturbance, phase-lag subsystem, and comparable states) to characterize three specific types of structural change: reduction to a first-order system, addition of cross-links between subsystems, and addition of a minor loop containing a delay.

CONTENTS

3.1. Reduction to a First-Order System	134
Characterizing the Effect of Parameter Changes: Time Constants	134
A Principle	147
Another Application: Orders in a Pipeline	151
Further Examples: The Spring-Mass System and Time to Correct Backlog for Orders TCBO	155
3.2. Cross-Links between Subsystems	161
A Descriptive Example: Vehicle Control	162
Another Descriptive Example: A Production-Distribution Chain	165
A Principle	167
An Optimal Control Viewpoint	174
3.3 Adding a Minor Negative Loop with a Delay	181
An Example: Influence of Availability on Ordering	181
Explanation in Terms of Phase Shift	189
A Principle	198

### 3.1. REDUCTION TO A FIRST-ORDER SYSTEM

Characterizing the Effect of Parameter Changes: Time Constants. There are two parameters in the employment-backlog system identified as time constants: the time to correct employment TCE, and the time to correct backlog TCB. In the basic employment-backlog model, both parameters have a value of 0.5. Figures 3-1 and 3-2 show the effect of reducing each parameter respectively to 0.1. Reducing TCE in Figure 3-1 appears to completely eliminate the tendency of the employment-backlog system to oscillate. In contrast, reducing TCB in Figure 3-2 produces sustained, lightly-damped oscillations with a period of about 1.4 years. Despite TCE and TCB being identified as "time constants," they differ dramatically in their effect on system behavior. Seemingly, there should be a way of characterizing the various parameters in a model that is more closely aligned with the effect of those parameters on the model behavior. What, then, is the difference between the time to correct employment TCE and the time to correct backlog TCB?

The effect of reducing the time to correct employment TCE can be seen in the state-space diagram in Figure 3-3. The dotted line rising from left to right represents the locus of states where the employment  $E$  equals the desired employment  $DE$ . The equation for desired employment  $DE$  can be derived from the model equations as follows:

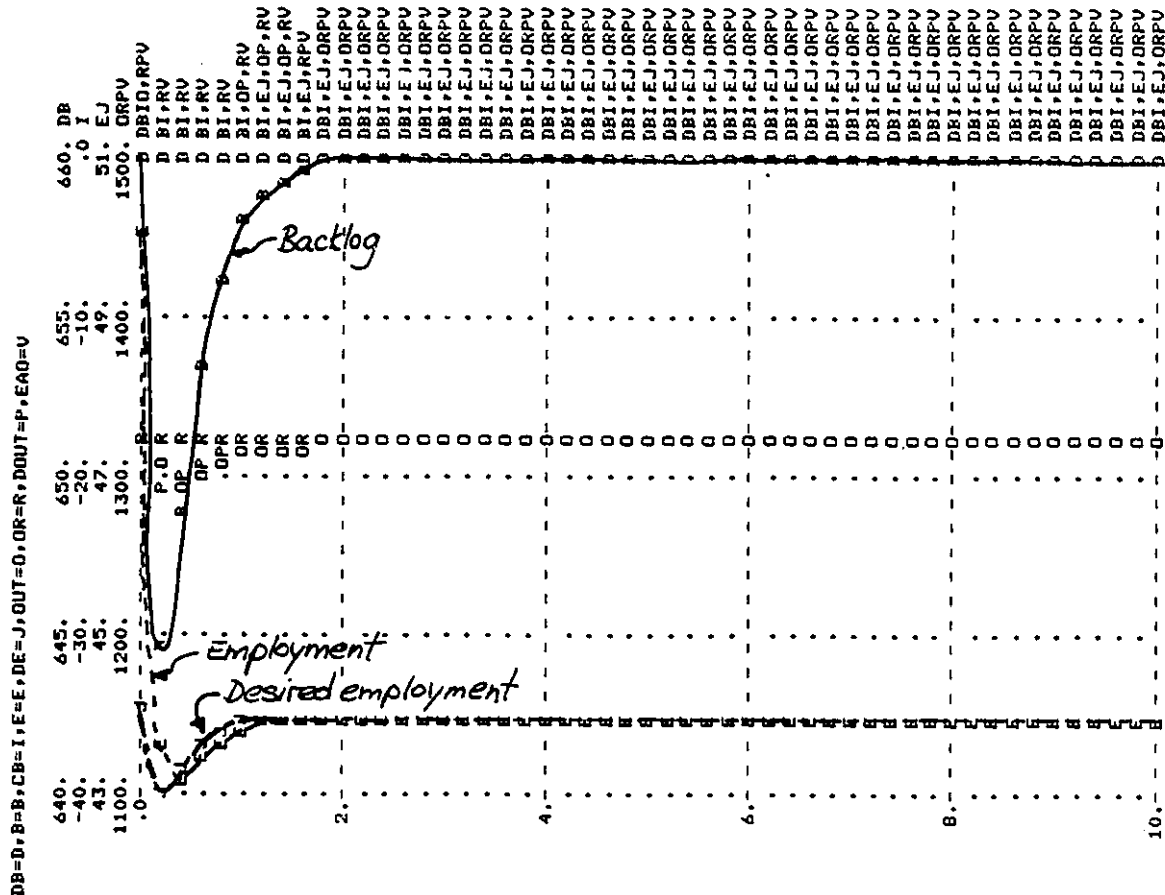


Figure 3-1.  
REDUCING TIME TO CORRECT EMPLOYMENT TCE FROM 0.5 TO 0.1

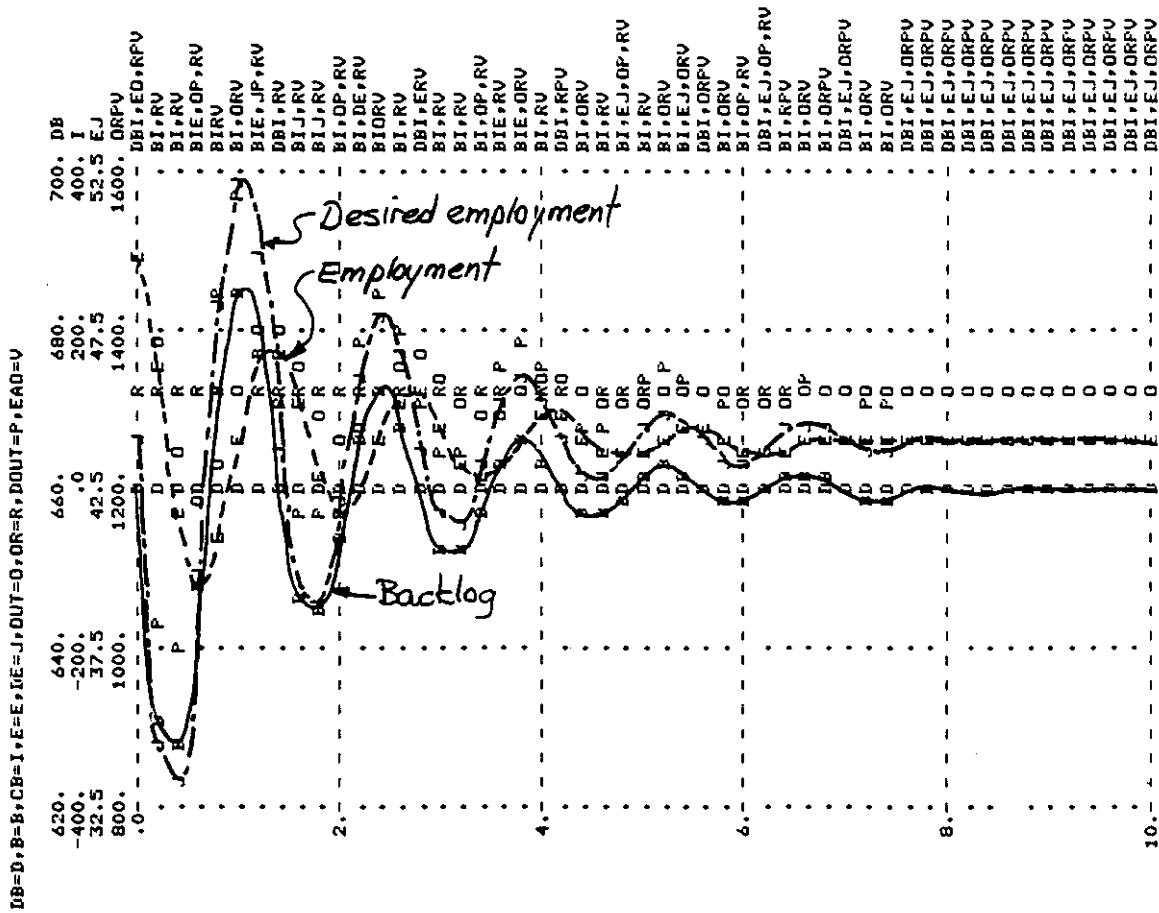


Figure 3-2.  
REDUCING TIME TO CORRECT BACKLOG TCB FROM 0.5 TO 0.1



$$DE = DOUT/PROD$$

$$DE = (EAO+CB)/PROD$$

$$DE = (EAO+(DB-B)/TCB)/PROD$$

$$DE = (EAO+(EAO*DBC-B)/TCB)/PROD$$

$$DE = 1320+(1320*0.5-B)/TCE/30$$

$$DE = 44+(660-B)/(TCE*30)$$

$$DE = 44+(660-B)/15$$

Thus, the desired employment DE is a decreasing function of the backlog B, so that a value of backlog B above the steady-state value calls for employment E larger than its steady-state value in order to produce units and lower the backlog. The actual employment E moves toward the desired employment DE. In state space, the rate of change of employment E is given by the vertical component of the rate-of-change vectors. The rate of change of employment E is given by

$$\begin{aligned}(d/dt)E &= NHR \\ &= (DE-E)/TCE\end{aligned}$$

Thus, the employment E will always move toward the dotted line: down when above the line, and upward when below the line. The farther employment E is from desired employment DE, the larger will be the vertical component of the rate-of-change vector. Over the entire state plane, the magnitude of the rate of change of employment E (the vertical component of the rate-of-change vector) is inversely proportional to the time to correct employment TCE. Thus, at every point in state space,

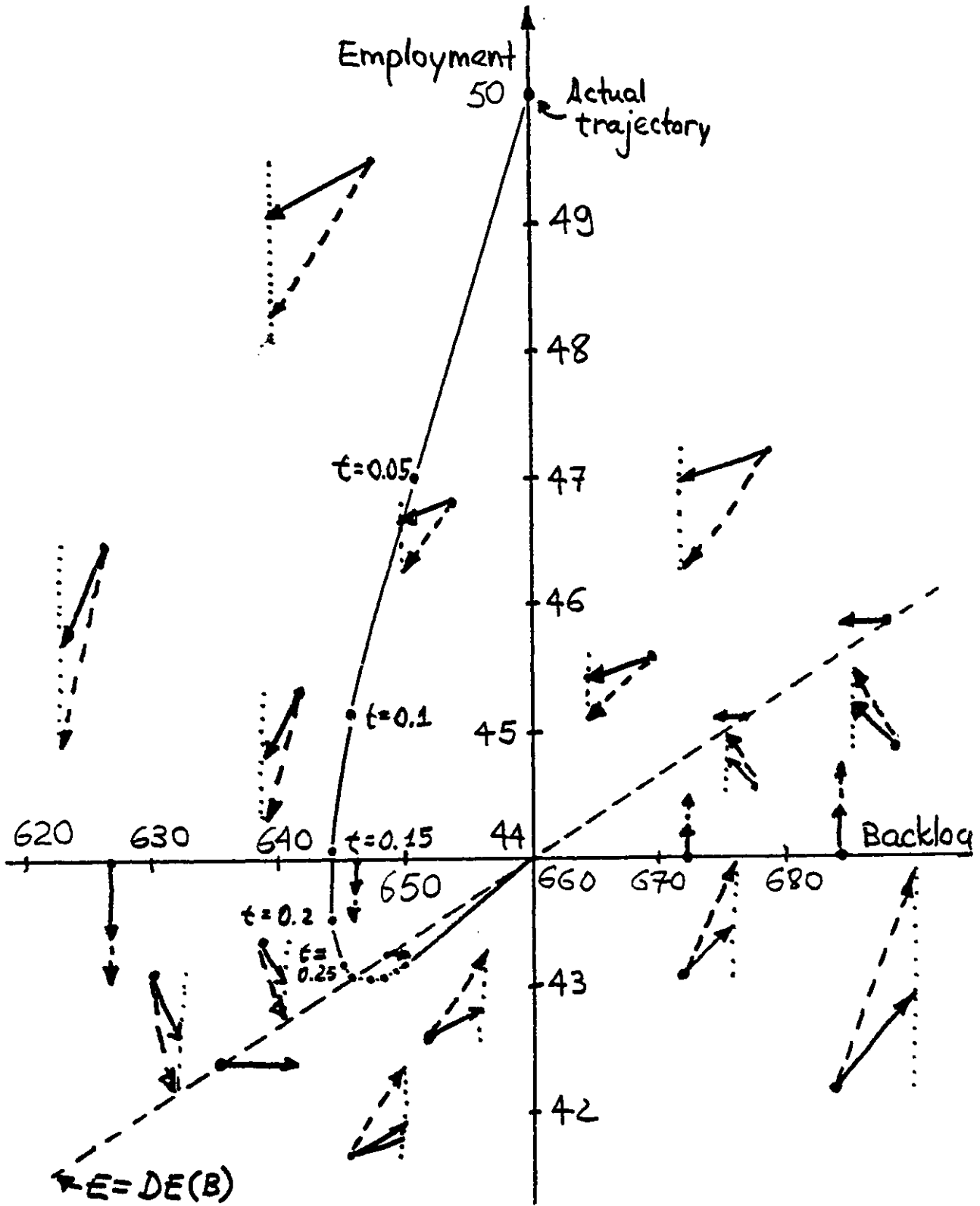


Figure 3-3.  
EFFECT OF REDUCING TCE IN STATE SPACE

reducing TCE from 0.5 to 0.1 magnifies the vertical component of the rate-of-change vector by a factor of five. This is shown schematically in Figure 3-3. (On the diagram, the magnitudes are only multiplied by a factor of about 2.5, because of the unwieldy length of the rate-of-change vectors.)

Describing Figure 3-3 in very simple terms, the response has two very distinct components: a very fast initial transient, and a much more leisurely, asymptotic approach to equilibrium. The effect of the initial transient is to bring the initially-excessive employment  $E$  down to the value of the desired employment  $DE$ . Because employment  $E$  was initially higher than its equilibrium value, the backlog  $B$  drops slightly during the initial transient. Afterwards, when  $E$  essentially equals  $DE$ , the backlog  $B$  shows a first-order, asymptotic approach to equilibrium.

The actual trajectory does not differ very much from the simple overall description just given. The initial employment  $E$  is high and drops rapidly at first. The backlog  $B$  decreases as long as employment  $E$  exceeds its steady-state value. When employment  $E$  crosses the horizontal axis and equals 44 men, the rate of output  $OUT$  exactly equals the order rate  $OR$ , so the rate of change of backlog  $B$  is zero, and the trajectory is exactly vertical at that point. When employment  $E$  passes through 44 men (the number of employees necessary to produce at a rate of output  $OUT$  equal to the order rate  $OR$ ), the backlog  $B$  begins to decrease slightly. The overall movement, however, is predominantly downward toward the dotted line. As employment  $E$  approaches the dotted line,  $E$  changes more slowly, and the backlog  $B$  maintains about the same

rate of change, so the trajectory curves right. When the trajectory crosses the dotted line, the rate of change of employment E is zero (since employment E equals desired employment DE), so the trajectory is horizontal at that point.

After the trajectory crosses the dotted line, both backlog B and employment E move smoothly toward the equilibrium point. As they move, employment E remains slightly below the desired employment DE, since as B increases, the desired employment DE increases, and actual employment E will lag DE by the time to correct employment TCE. The discrepancy between E and DE causes the net hiring rate NHR to exceed zero, which allows employment E to rise. Overall, the trajectory in Figure 3-3 can be characterized by first, rapid movement of employment E toward desired employment DE, and second, much slower movement of both backlog B and employment E along the desired employment DE line toward the equilibrium point. For most of the time, then, the employment E can be considered equal to desired employment DE, which is in turn a function of the backlog B. In effect, employment E becomes an auxiliary variable in the system. Figure 3-4 shows employment E as an auxiliary variable in the employment-backlog system. The resulting system effectively contains only one level, so that no endogenous oscillation can occur. Oscillation would require the backlog B to pass through the equilibrium point, but if the backlog B is in equilibrium, the whole

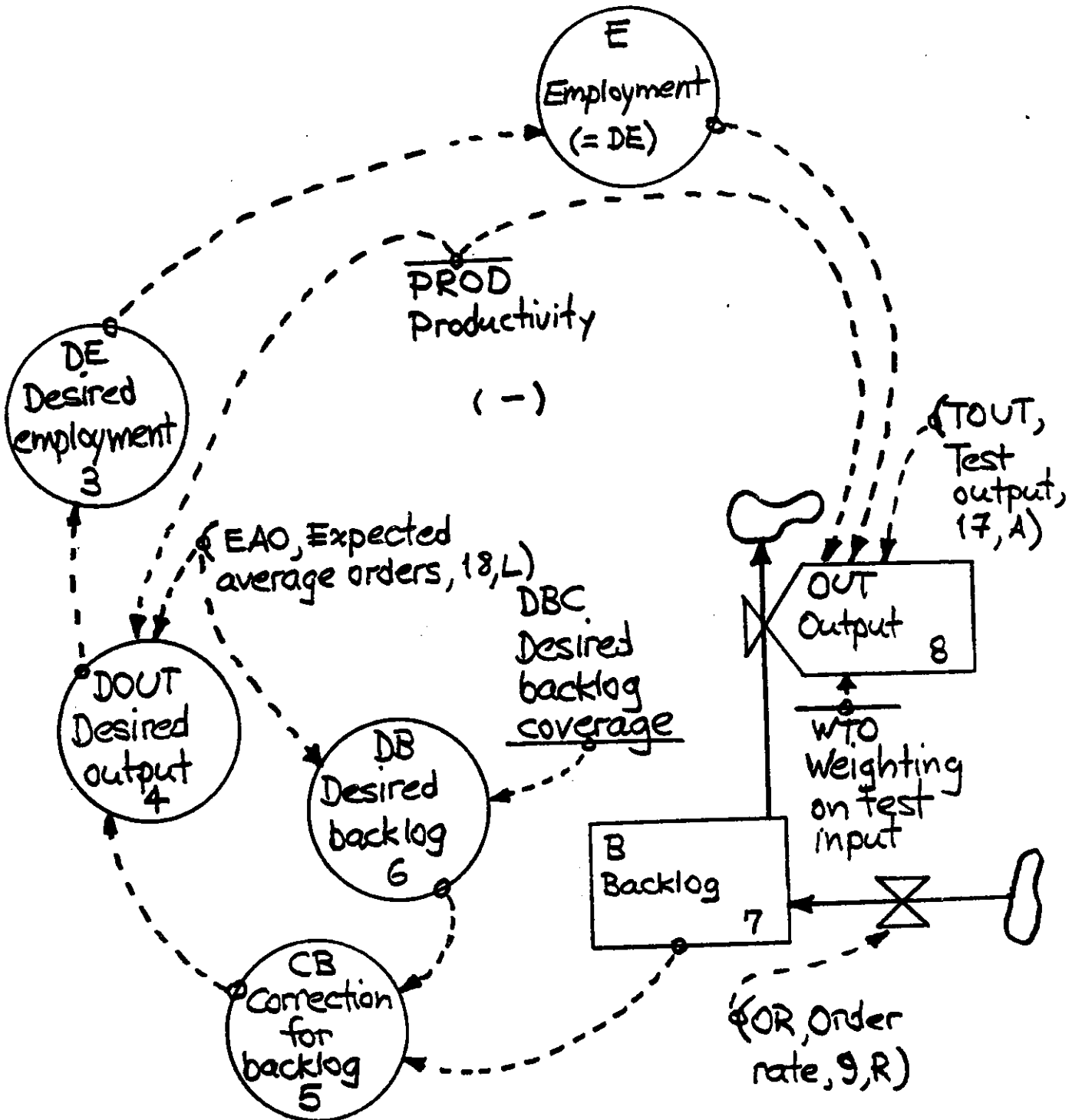


Figure 3-4.  
EMPLOYMENT E EFFECTIVELY CHANGED TO AN AUXILIARY VARIABLE

system is in equilibrium, so that no further change can occur.\*1\*

Reducing the time to correct employment TCE eliminates oscillations in the employment-backlog system. Before generalizing this result into a principle, let us first examine the effect of reducing the other explicit time constant in the system, the time to correct backlog TCB, and find the characteristics that distinguish TCB from TCE. Recall that the desired employment DE is a function of the backlog B and the time to correct backlog TCB:

$$DE = 44 + (660 - B) / (TCB * 30)$$

Thus, reducing the time to correct backlog TCB increases the slope of the dotted line showing the locus where employment E equals desired employment DE, as shown in Figure 3-5. As discussed above, the magnitude of the vertical component of the rate-of-change vectors (the rate of change of employment E) is proportional to the discrepancy between employment E and desired employment DE. Thus, increasing the slope of the dotted line changes the rate-of-change vectors from the original solid arrows to the dotted arrows in Figure 3-5. In the second and fourth quadrants (shaded), reducing the time to correct backlog TCB causes the rate-of-change vectors to move the state toward states corresponding to oscillations of smaller magnitude than the original rate-of-change vectors. In the first and third quadrants (unshaded), the effect of reducing TCB is to move the state toward states corresponding to oscillations of larger magnitude. Over the entire

-----  
\*1\*For further discussion, see Mass, N.J., and P.M. Senge, "Understanding Oscillations in Simple Systems," System Dynamics Group Working Paper D-2045-2, Alfred P. Sloan School of Management, E40-253 (Cambridge, Mass.: MIT, 1976).

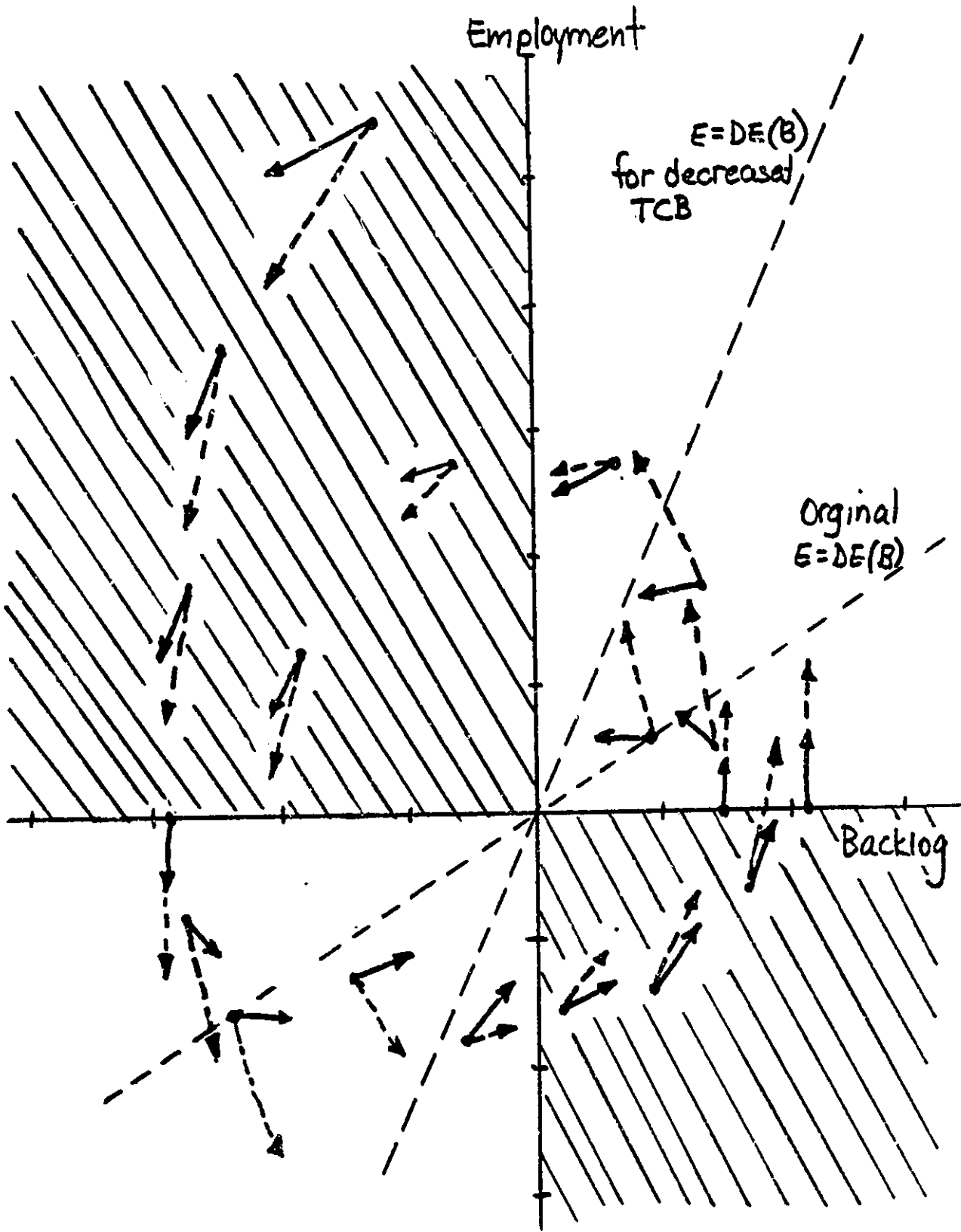


Figure 3-5.  
EFFECT OF REDUCING TCB IN STATE SPACE

spiral trajectory, then, the effect of reducing TCB appears to be mixed, tending to increase damping in some states, and tending to decrease damping in other states. However, closer inspection of Figure 3-5 reveals that in the shaded quadrants, the effect of reducing TCB is mostly to propel the state around virtually the same trajectory more rapidly. In contrast, reducing TCB in the unshaded quadrants causes the rate-of-change vectors to propel the system definitely outward, away from the equilibrium point. Thus, Figure 3-5 indicates that reducing TCB from 0.5 to 0.1 should markedly reduce the damping of the oscillations. Indeed, Figure 3-2 shows very lightly-damped oscillations, instead of the original moderately-damped oscillations.

Thus far, we have reduced two parameters from 0.5 to 0.1. For the time to correct employment TCE, this reduction eliminated oscillations in the employment-backlog system. For the other parameter, the time to correct backlog TCB, this reduction increased the tendency to oscillate. The distinction between the two parameters is shown in Figure 3-6, which plots the open-loop response of employment E to an increase in the backlog B, for the original and reduced parameters. (In other words, backlog B is treated as exogenous, not affected by employment E.) The trajectory for the original parameters shows the employment E approaching desired employment DE with a time constant equal to the time to correct employment TCE. This parameter is actually a time constant, whose value in fact characterizes the behavior of part of a system. Reducing TCE reduces the time constant of the approach to



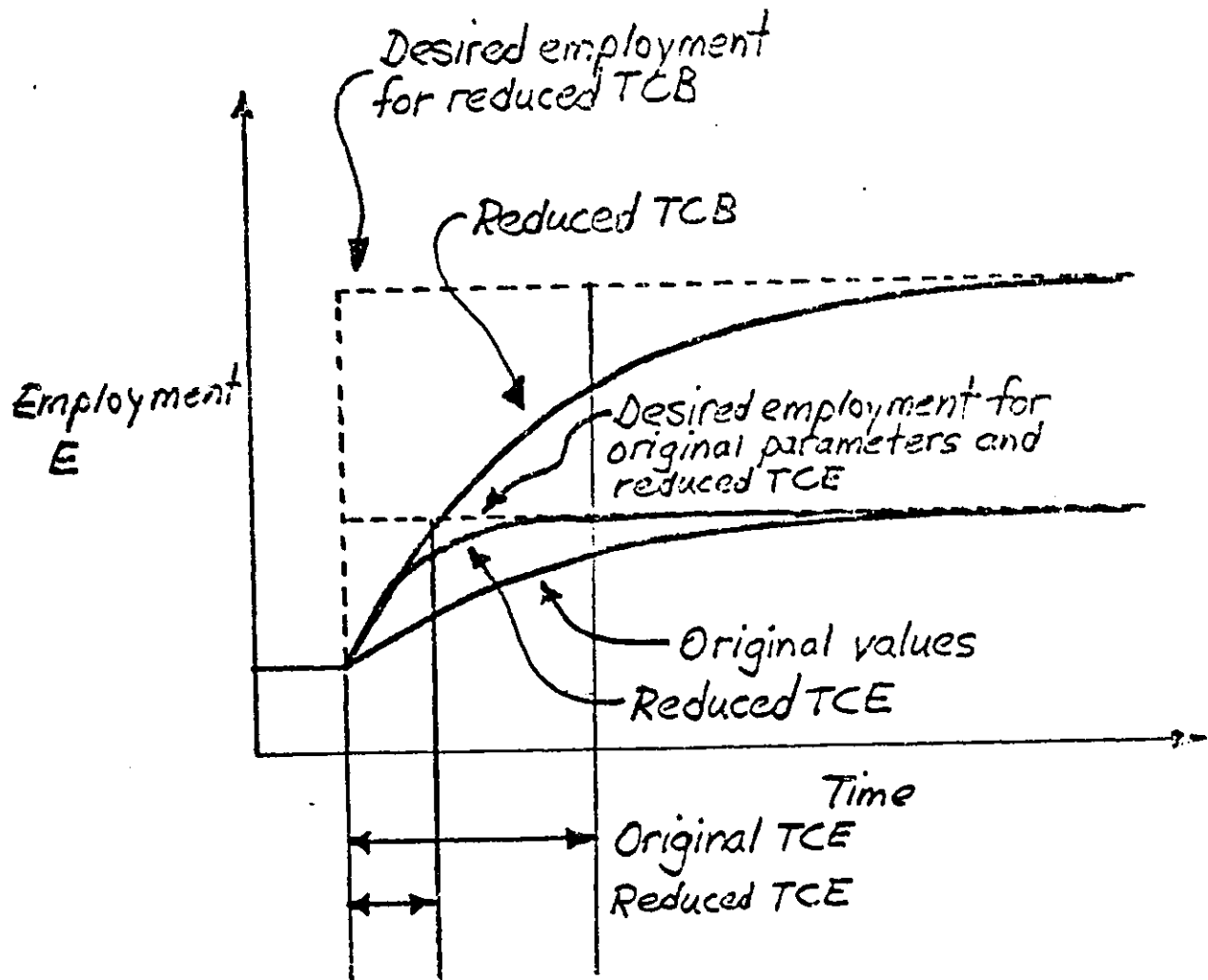


Figure 3-6.  
DISTINCTION BETWEEN TCE AND TCB

desired employment  $DE$ . In contrast, the time to correct backlog  $TCB$  seems to function as a gain: reducing  $TCB$  increases the magnitude of the response to the negative step in backlog  $B$ , without altering the time form. Employment  $E$  still approaches desired employment  $DE$  with a time constant of  $TCE$  even though the change is made larger when  $TCB$  is reduced.

The time to correct employment  $TCE$  and the time to correct

backlog TCB perform quite different functions within the system structure. If we wish to be accurate in thinking about the effects of such parameters on system behavior, TCE and TCB probably ought not to be classified as "time constants," with both parameters subsumed under the same concept of "time constant." The time to correct employment TCE is in fact a true time constant, and its value characterizes the behavior of a subsystem. But the time to correct backlog TCB does not so characterize any particular aspect of the system behavior. Its function is that of a gain element--a system element whose output at each successive moment of time depends only on its input at that time. TCB thus characterizes an instantaneous input-output relation, not behavior over time, so TCB ought not to be called a time constant.\*2\* Because gain elements couple levels to rates, their dimensions often work out to be (1/time units). Since the inverse of gains often has the dimension of time units, it is sometimes convenient to think of gains as equivalent to a correction time. I suggest that such time-related quantities be called "coupling time constants," and that the name "time constant" be reserved exclusively to describe the time constant of a first-order loop.

-----

\*2\*In real life, of course, no input-output relationship is truly instantaneous. Gain elements are a simple way of modeling dynamic processes that equilibrate so rapidly that the dynamics of equilibration have a negligible effect on the behavior of the rest of the system.

A Principle. The principle on the origin of oscillations in Section 2.1 states that "oscillations occur when a disturbance propagates around a loop formed by two or more phase-lag subsystems to return the system to a state comparable to the initial disequilibrium state." The employment-backlog system has two phase-lag subsystems, one consisting of the level of backlog B and its associated rates, and the other consisting of the level of employment E and its associated rates and auxiliaries. Reducing the time to correct employment TCE to a very small value in effect causes the phase-lag property of the subsystem around employment E to vanish; In effect, employment E becomes an auxiliary variable. We can formalize this property by saying that the phase-lag subsystem containing the employment E has become a gain element, which can be defined as

Gain element

A subsystem whose output at any given time depends only on its input at that moment of time.\*3\*

In this case, with TCE very small, the employment subsystem just becomes a linear gain element connecting backlog B to the rate of output OUT,

---

\*3\*The gain of the gain element is defined as the partial derivative of the magnitude of the input with respect to the magnitude of the output. For linear gain elements, the gain of the element is therefore constant, regardless of the magnitude or time-behavior of the input.

whose gain is

$$\begin{aligned} (dOUT/dB) &= (d/dB)(PROD*E) \\ &= (d/dB)(PROD*DE) \\ &= PROD*(d/dB)(44+(660-B)/(TCB*PROD)) \\ &= PROD*(-1)/(TCB*PROD) \\ &= -1/TCB \end{aligned}$$

The concept of gain element provides a clear distinction between time constants and coupling time constants. Reducing the time to correct employment TCE (a true time constant) to near zero turns a phase-lag subsystem effectively into a pure gain element. Reducing TCE effectively eliminates the ability of the loop to propagate a disturbance from one part of the system to another. In contrast, reducing the time to correct backlog TCB (a coupling time constant) to near zero amplifies disturbances in the backlog B as they are transmitted to desired employment DE and employment E.\*4\* Altering a coupling time constant may alter the oscillatory characteristics of the system as a whole, but the alteration does not turn a phase-lag subsystem into a pure gain element (as is the case for time constants).

The coupling time constant from B to OUT to B is  $-TCB$  (which in fact has become a true time constant when B effectively becomes the only

-----

\*4\*Again, the reader is cautioned that decreasing the coupling time constant between one phase-lag subsystem and another does not necessarily destabilize oscillations, even though this is frequently the case. Section 6.3 discusses coupling time constants as a subject for future research.

level in a first-order system). We can formalize this property as follows:

Effectively-first-order system

A system in which the response time of one level significantly exceeds (perhaps by a factor of ten) those of other phase-lag subsystems in the system, which thus effectively become gain elements with respect to the movements of the remaining level.

The response time is a characteristic of a system's or a subsystem's behavior. The parametric change required to decrease a response time may be either an increase or decrease in the parameter value. The principle assumes that the parameter changes do not cause any of the subsystems to generate their own endogenous oscillations. (With complex, higher-order subsystems, it could be possible for a parameter change to render the subsystem oscillatory in its own right, independent of the larger system in which it functions.) There are two ways to make parameter changes that render a system effectively first-order. One way is to decrease the response times in all subsystems but one. The other way is to increase one response time well beyond the response times of the other subsystems in the loop.

The concept of gain element allows one to define the concept of effectively-first-order system, which in turn allows one to state a principle:

Principle on reduction to an effectively-first-order system

An oscillatory system can be made not to oscillate by changing it to an effectively-first-order system, so that when the remaining effective level passes through its equilibrium value, the entire system does so, and no further movement occurs.

The principle above may not seem particularly useful, in that one rarely has the opportunity to implement policies that correspond to reducing response times to near zero. However, it is often true that changing response times in the direction that would produce an effectively-first-order system can suffice to reduce the system's tendency to oscillate, even if the changes are not extreme enough to actually render the system effectively-first-order.

Another way of expressing the principle above is to say that at least two of the longest response times of phase-lag subsystems around a loop must be comparable for that loop to generate oscillations. If all subsystems but one reach equilibrium before that one subsystem begins to approach equilibrium, the approach of the one subsystem to equilibrium will be smooth, first-order-like behavior. This formulation of the principle explains why reducing the gain on an integral controller usually diminishes or eliminates oscillations generated by a loop going through the integral controller. If the rest of the system comes close to equilibrating before the level in the integral controller can overshoot its equilibrium value, then the level in the integral controller can approach equilibrium smoothly, as if it were a first-order system, and the rest of the system were auxiliary

variables.\*5\*

The notion of requiring comparable response times for oscillation can be useful in formulating a model of an oscillatory phenomenon. During model formulation one is always in need of guidelines both first selecting which real cause-and-effect relationships should be explicitly included in the model, and second, deciding which relationships should be modelled as instantaneous (by auxilliary variables, or pure gain elements), and which relationships should be modelled as dynamic (involving level and rate structure between input and output).

The principle above in some instances also provides a conceptually correct way to think about the effect of minor loops on oscillation: if adding a minor loop or decreasing the time constant of an existing minor loop tends to eliminate a phase-lag subsystem, that change will stabilize the system's oscillations.

Another Application: Orders in a Pipeline. The preceding discussion shows how a system effectively can be made into a first-order system by changing a time constant. There are structural changes that have the same effect. Section 3.2 discusses one type of structural change, cross-links between subsystems, that accomplishes this. The following

-----

\*5\*Strictly speaking, "integral controller" is usually a misnomer, since only very rarely do systems contain a pure integration. Much more often, integrations are embedded in feedback loops. However, we can think of an "integral controller" as a local structural feature, so that an integration without minor loops, that is, an integration embedded in only second-order or higher-order loops, can be considered an integral formulation.

example is another instance.

Figure 3-7 shows a very simple system designed to maintain an inventory  $I$  at its desired level, the desired inventory  $DI$ , despite the varying usage rate  $UR$ . The order rate  $OR$  simply replaces the units removed from inventory by the usage rate  $UR$ , and adjusts the inventory  $I$  toward the desired inventory  $DI$ , through a coupling time constant, the time to correct inventory  $TCI$ . The order rate  $OR$  feeds into a third-order delay, so on the average, after the time to deliver orders  $TDO$ , the orders are filled and flow into the inventory  $I$ . If the time to correct inventory  $TCI$  is sufficiently much shorter than the time to deliver orders  $TDO$ , the system shown in Figure 3-7 can exhibit damped, steady-state, or even expanding oscillations.



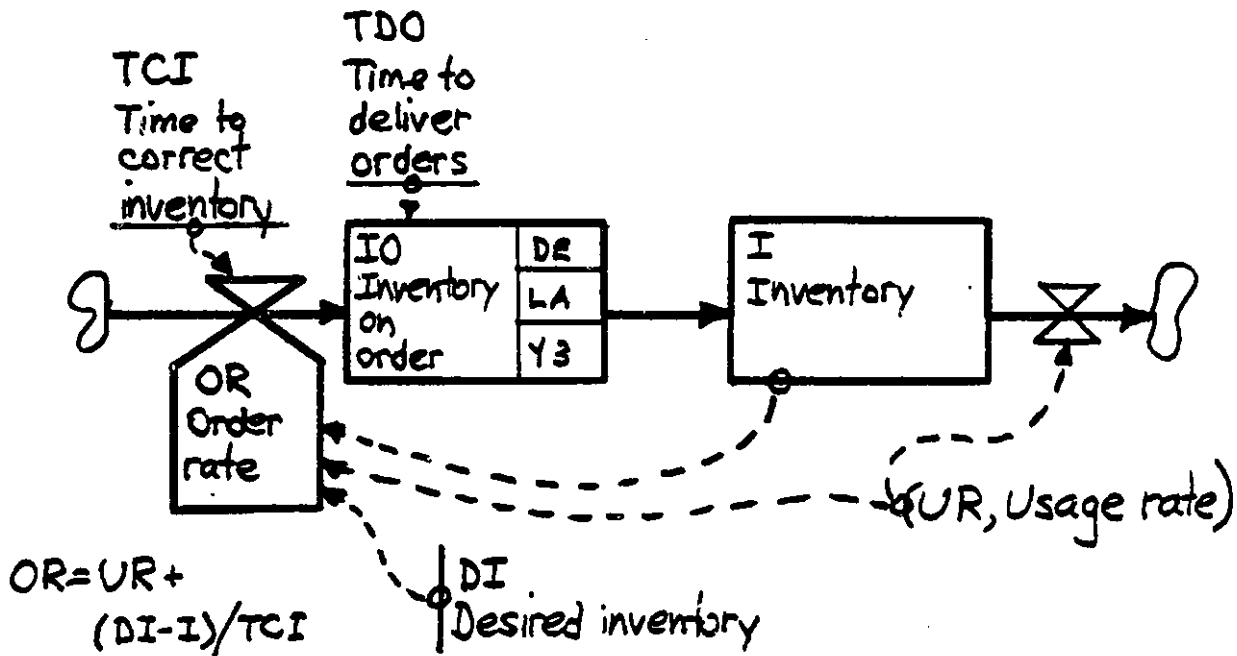


Figure 3-7.  
PIPELINE SYSTEM

Figure 3-8 shows one structural change that will stabilize the pipeline system shown in Figure 3-7. The order rate OR now incorporates a desired inventory on order DIO and the actual inventory on order IO. Thus, when the inventory on order IO exceeds the desired inventory on order DIO, the order rate OR will be reduced, even before excessive orders have raised the inventory I. Notice that the order rate OR in effect regulates the total number of units in all four of the system levels:

$$\begin{aligned}
 OR &= UR + (DI - I + DIO - IO) / TCI \\
 &= UR + ((DI + DIO) - (I + IO)) / TCI \\
 &= UR + (\text{Desired total inventory} - \text{Total inventory}) / TCI
 \end{aligned}$$

The term "total inventory" denotes the sum of the actual inventory I on hand plus an inventory already on order IO. Figure 3-9 shows the first-order system that is the exact mathematical equivalent of the pipeline system shown in Figure 3-8. The feedback link from inventory on order IO to the order rate OR has rendered the pipeline system an effectively-first-order system. We can therefore predict, a priori, that this seemingly-complex system will show no oscillations but instead just simple first-order asymptotic behavior, with a time constant of TCI.

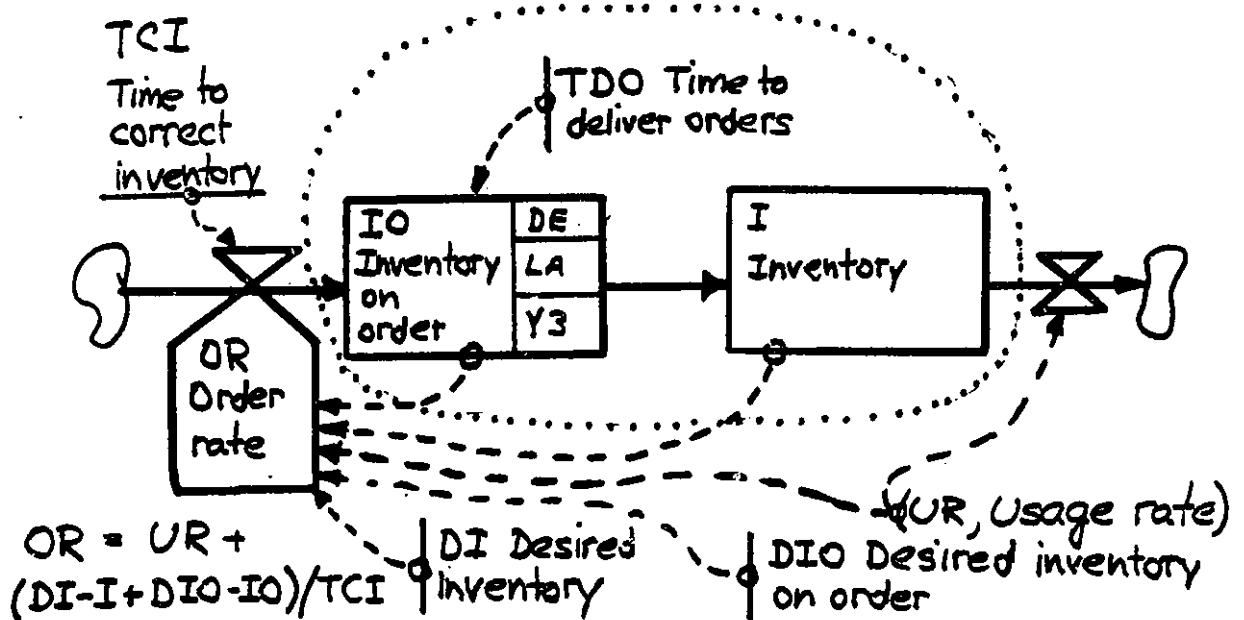


Figure 3-8.  
PIPELINE SYSTEM WITH FEEDBACK FROM PIPELINE

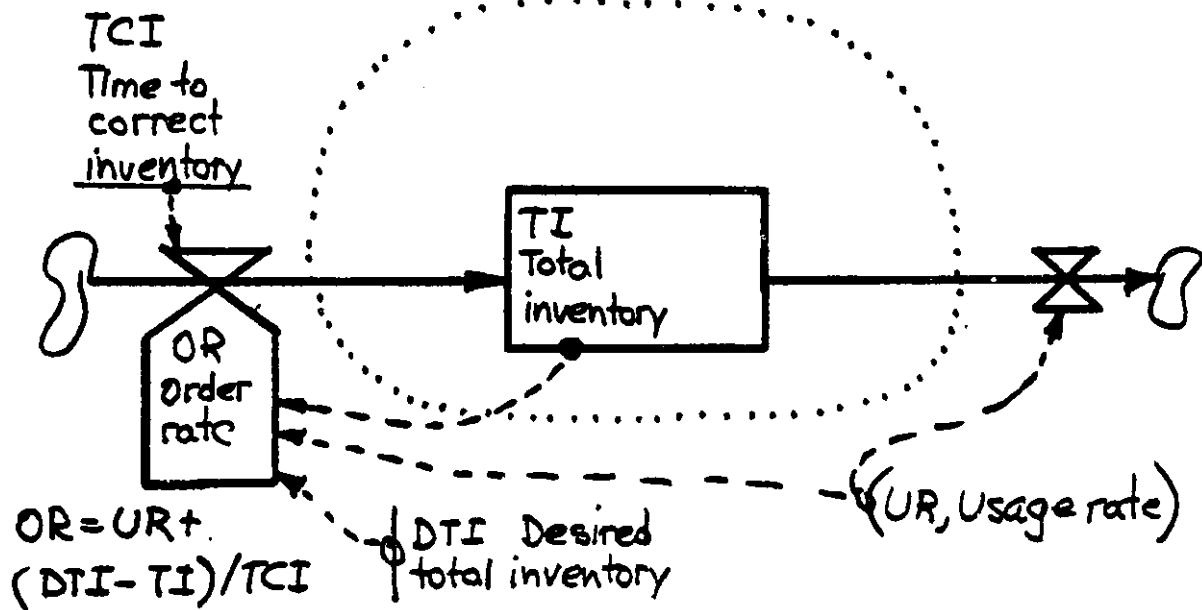


Figure 3-9.  
FIRST-ORDER EQUIVALENT OF PIPELINE SYSTEM

Further Examples: The Spring-Mass System and Time to Correct Backlog for Orders. A variety of physical, biological, and social systems are capable of exhibiting very rapid initial transients, but for many purposes these systems can be regarded as having a much lower order (number of levels) if only the longer responses are of interest. The following two examples give situations where a somewhat-oscillatory second-order system can be regarded as an effectively-first-order system. A door with a closing spring is quite similar in structure to the spring-mass system analyzed earlier. It has two state variables, position P and velocity V. Although many doors do in fact swing and exhibit oscillation, many doors exhibit virtually no oscillation, because

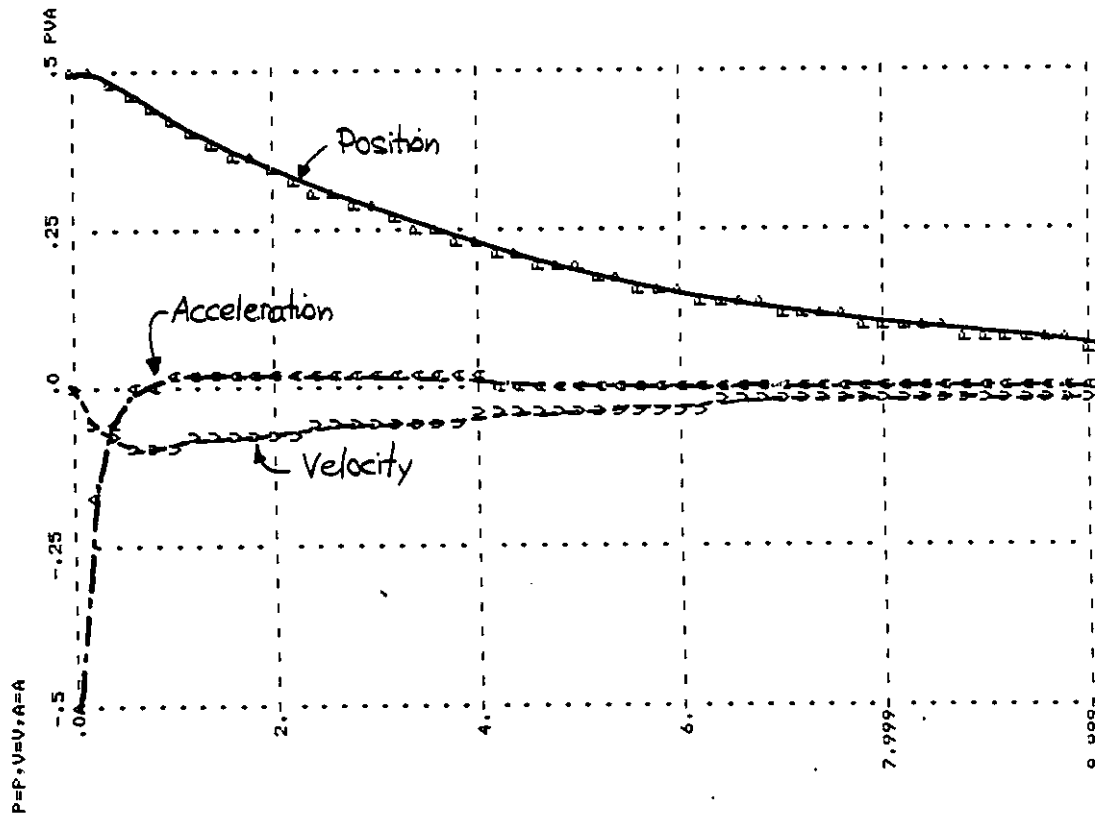


Figure 3-10.  
SPRING-MASS SYSTEM WITH HIGH FRICTION

they are heavily damped. We can use the spring-mass system to see how this behavior comes about. Figure 3-10 shows the behavior of the spring-mass system when the friction coefficient FC is set at 25. (In previous simulations, FC has been set to 0.0 or 2.5.) The velocity V drops relatively quickly to a small value, so that the position P

declines very slowly.

One way to think of the behavior in Figure 3-10 is to notice that the velocity  $V$  equilibrates fairly quickly to a value determined by the position  $P$ . We can derive this quasi-steady-state value of velocity  $V$  as follows: Imagine for a moment that the position  $P$  is fixed (which is close to the actual short-term dynamics. To what value  $V(P)$  would the velocity  $V$  equilibrate? Substituting the system equations into the equation for the rate of change of velocity  $V$ ,

$$(d/dt)V = -P - (FC/5)V = 0$$

$$V(P) = -(5/FC)P$$

The quasi-steady-state value of velocity  $V$ ,  $V(P)$ , is thus a function of the position  $P$ . The expression for the velocity  $V$  can be rewritten to show that in fact the actual velocity  $V$  is an exponential smoothing of the quasi-steady-state value.

$$\begin{aligned}(d/dt)V &= -P - (FC/5)V \\ &= (FC/5)(-5/FC)P - (FC/5)V \\ &= (FC/5)(V(P) - V)\end{aligned}$$

When the friction coefficient  $FC$  is 25, the velocity  $V$  approaches the quasi-steady-state velocity  $V(P)$  with a time constant of

$$\begin{aligned} &1/(FC/5) \\ &= 5/FC \\ &= 5/25 \\ &= 0.2\end{aligned}$$

With the position P declining with a time constant of about 7 seconds, the delay in velocity V adjusting to V(P) of 0.2 seconds is negligible. In effect, the velocity V has become an auxiliary variable, and the spring-mass system with such heavy damping is effectively a first-order system.

As the discussion of the principle in this section indicated, there is more than one way to make a system into an effectively-first-order system. In fact, we have already seen two ways of making the employment-backlog system into an effectively-first-order system. One way was to reduce the time to correct employment TCE, which made the employment E virtually a function of the backlog B. We saw a second way earlier, where reducing the time to correct backlog for orders TCBO also strongly reduced the tendency of the employment-backlog system to oscillate. In fact, lowering TCBO effectively makes the backlog B a function of employment E, as can be seen from the following derivation.

Assume for the moment that the employment E remains unchanging, so that the backlog B will seek an equilibrium value B(E), which can be computed from the equation for the rate of change of backlog B as follows:

$$\begin{aligned} (d/dt)B &= OR-OUT = 0 \\ &= DOR+(DBO-B)/TCBO-E*PROD \end{aligned}$$

implying that

$$\begin{aligned} (B-DBO)/TCBO &= DOR-E*PROD \\ B(E) &= DBO+(DOR-E*PROD)*TCBO \end{aligned}$$

So the backlog B seeks a steady-state value B(E), with a time constant we can see from rewriting the equation for the rate of change of backlog B:

$$\begin{aligned} (d/dt)B &= OR-OUT \\ &= DOR+(DBO-B)/TCBO-E*PROD \\ &= -B/TCBO+(DBO+(DOR-E*PROD)*TCBO)/TCBO \\ &= -B/TCBO+B(E)/TCBO \\ &= (B(E)-B)/TCBO \end{aligned}$$

Thus, when the time to correct backlog for ordering TCBO is made small, the backlog B for all practical purposes becomes a function of the level of employment E, and the system is effectively a first-order system. (Figure 3-11 illustrates.) There will be little if any oscillation.

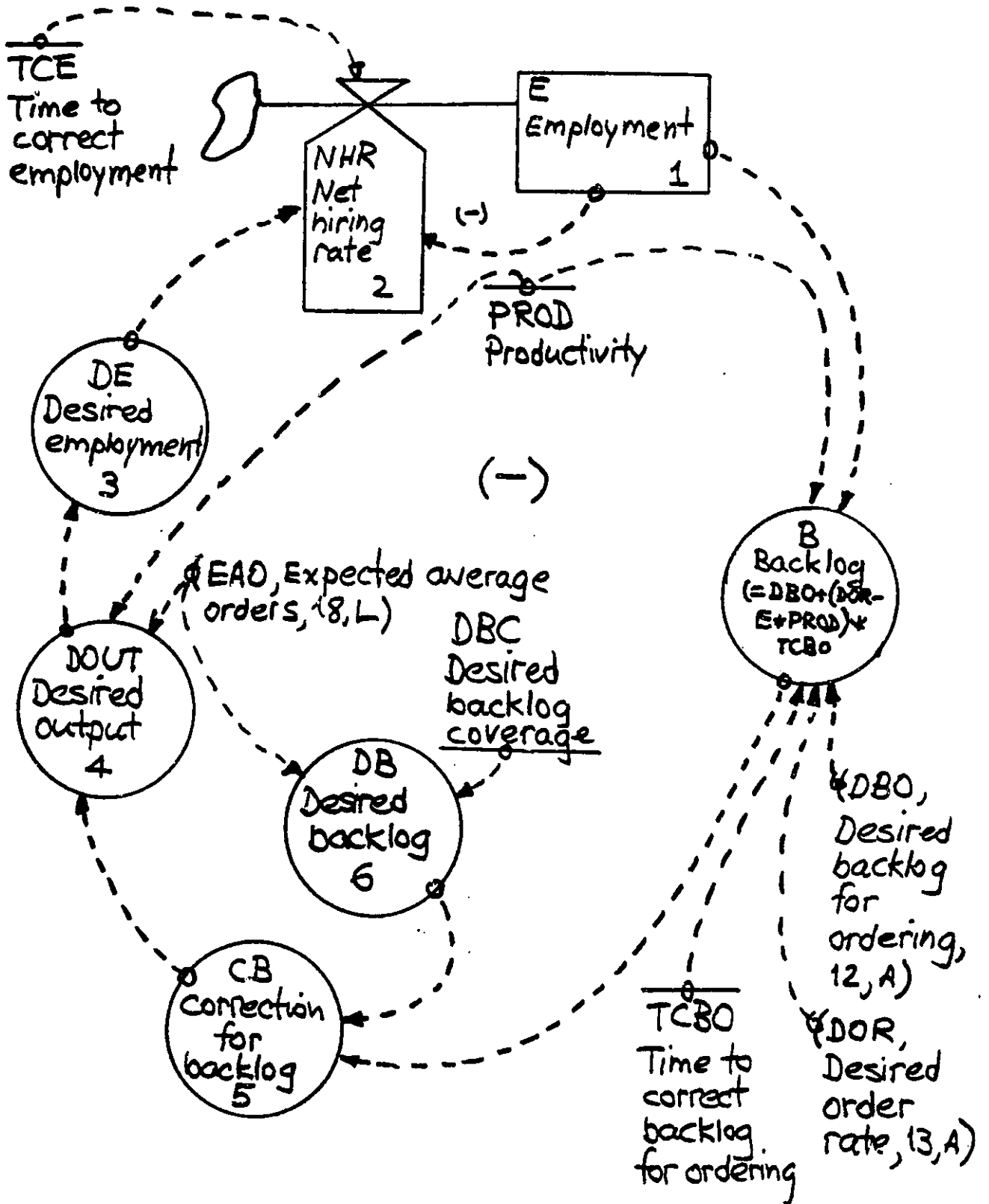


Figure 3-11.  
 BACKLOG B EFFECTIVELY CHANGED TO AN AUXILIARY VARIABLE



### 3.2. ADDING CROSS-LINKS BETWEEN SUBSYSTEMS

The preceding section used the principle on the origin of oscillations to enunciate a category of parametric and structural changes that reduce a system's tendency to oscillate. This section uses the same principle on the origin of oscillations as the basis for another type of structural change to reduce or eliminate oscillations. Oscillations, according to the principle, can originate from a loop formed of phase-lag subsystems. Because of the loop structure, one subsystem on one side of the loop can be rising up away from its steady-state value even when another subsystem on the other side of the loop can be declining away from its steady-state value. It is possible for two subsystems to (at least temporarily) move in opposite directions because of the delays involved in one subsystem communicating around the loop with the other system. One common strategy for reducing a system's tendency to oscillate is to increase the ability of one subsystem to communicate with another by adding a direct link between them:

#### Cross-link

A cause-and-effect relationship that connects two subsystems that otherwise would be distant from one another, in terms of the number of intervening levels and other cause-and-effect relationships.

Before formally stating a principle on cross-links, let us examine two

common situations, one physical and one economic, where cross-links help to stabilize oscillations.

A Descriptive Example: Vehicle Control. One future form of transportation currently under intense investigation is a motor roadway, much like current highways, with the vehicles under automatic control. In theory, traffic on such a roadway could move far more rapidly and with much smaller spacing between vehicles than is possible under manual operation. One question which must be answered before such a system becomes feasible is what should control the acceleration of each vehicle.

Figure 3-12 illustrates the simplest sort of control. Each vehicle has a position  $P$  and a velocity  $V$ , both of which are to be controlled by the vehicle's acceleration  $A$ . In the simple control scheme, the acceleration  $A$  of each vehicle depends only on the vehicle's position  $P$  and velocity relative to the positions  $P$  and velocities  $V$  of the vehicles immediately behind and in front of the vehicle being controlled. (This is approximately the situation when the vehicles are under manual control.) Unfortunately, this type of control yields instabilities when the vehicles are close together. These instabilities are probably familiar to most freeway drivers: a car in front brakes

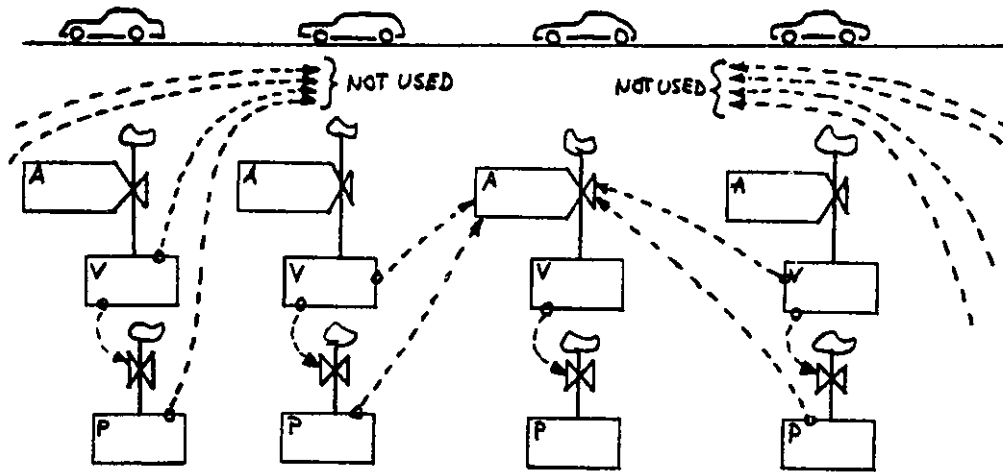


Figure 3-12.  
SIMPLE VEHICLE CONTROL SCHEME

and the car behind it brakes, a little more sharply than the first car, because the second car is now both moving too fast relative to the first car (which slowed down) and is also too close to the first car. Each succeeding car must brake more sharply, until the required deceleration exceeds the ability of some vehicle to brake, and a rear-end collision occurs. Similar but less extreme instabilities occur with "stop-and-go" traffic on very crowded freeways. In both cases, an initial disturbance propagates through the system, is amplified, and causes undesirable oscillations. The solution to this problem for manually-controlled vehicles (or vehicles automatically controlled with very local information) is to space the vehicles more widely, so that disturbances in the velocity and position of one vehicle will not cause such dramatic action in the next--the disturbance will not be amplified, and the system behaves more smoothly.

Another way of reducing the tendency of a string of vehicles to oscillate is to change the information used to determine the acceleration of each vehicle. It can be shown that the optimal control for vehicle acceleration  $A$  requires the acceleration  $A$  of each vehicle to respond to the position  $P$  and velocity  $V$  of every other vehicle in the chain of vehicles.\*6\* In effect, the optimal solution requires

---

\*6\*Indeed, experienced drivers probably often make a habit of looking at the two or three cars in front, by looking through the windows of the car immediately in front. This is why at least the author is uncomfortable when tall vans or trucks cut off the forward line of vision.

cross-links from every level in the system to every available rate. Such a control scheme would be considerably more costly than the simple scheme discussed earlier. Considerable attention has been given to the tradeoff between simplicity and performance. The results indicate that satisfactory results usually can be achieved by controlling the acceleration  $A$  of each vehicle using information from the neighboring five or six vehicles. (Obviously, information from more vehicles improves behavior further, but not by as much.) The performance of such compromise schemes comes quite close to the optimal performance with full cross-linking between all vehicles.

Another Descriptive Example: A Production-Distribution Chain. A very common problem of firms that sell to other firms is that the orders for their product fluctuate. In general, final demand for retail goods fluctuates less than the retailer's orders placed with the distributor, which in turn fluctuate less than the distributor's orders placed with the factory warehouse. Industrial Dynamics devotes considerable space to analyzing such a system; Figure 3-13 gives an overview of the system structure.\*7\* The overall structure is reminiscent of the structure of vehicle control: each sector is connected (initially, at least) only with its immediate neighbors. The vehicles attempt to maintain the appropriate position and velocity relative to their neighbors, and the firms attempt to maintain the appropriate inventories of goods and

-----  
\*7\*Forrester, Jay W., Industrial Dynamics (Cambridge, Mass.: MIT Press, 1961), pp. 137-186.

backlogs of unfilled orders relative to their neighbors. Both systems are capable of considerably amplifying disturbances, so that, for example, a small increase in demand to the retail sector can result in a large increase in demand to the factory sector, as each sector attempts to ship goods at an increased rate, and at the same time, to refill inventories depleted in earlier attempts to ship goods.\*8\* When goods finally arrive through the distribution chain, the goods arrive in such quantity that the sectors attempt to control excessive inventories and reduce their respective order rates. This reduction is in turn amplified as it moves through the distribution chain back toward the factory. The system thus continues to oscillate.

---

\*8\*This amplification is known as the inventory accelerator, and plays an important role in current business-cycle theory. See Mass, N. J., Economic Cycles (Cambridge, Mass.: Wright-Allen Press, 1975).

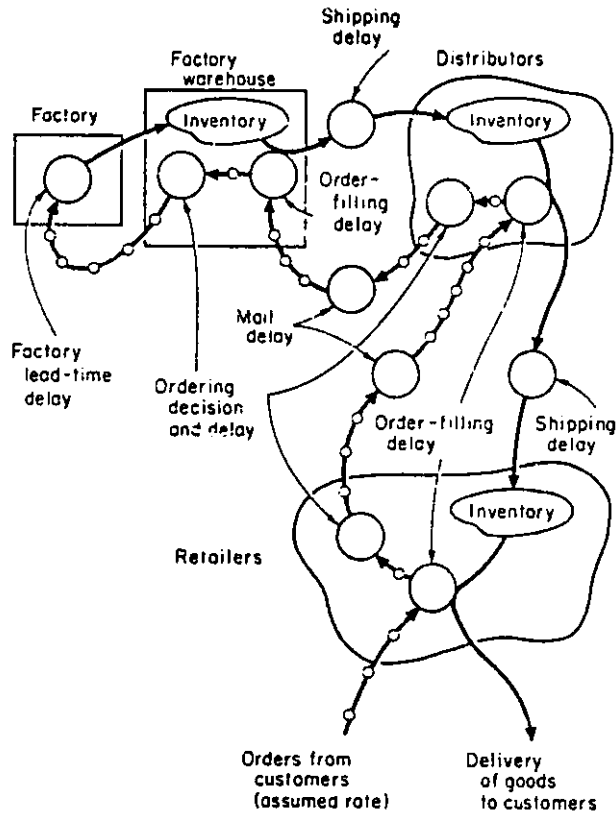


Figure 3-13.  
PRODUCTION-DISTRIBUTION SYSTEM  
(From Industrial Dynamics)

Appendix J of Industrial Dynamics reports modest success (reduction of oscillation) for a policy that uses information about retail sales at the factory level, thus establishing a cross-link between two otherwise-distant sectors.

A Principle. Figure 3-14 shows an abstract representation of adding a cross-link between otherwise-unconnected systems, such as in the examples above. The cross-link might be thought of as converting the three separate phase-lag subsystems into effectively only one phase-lag subsystem, through which a disturbance can propagate more rapidly than

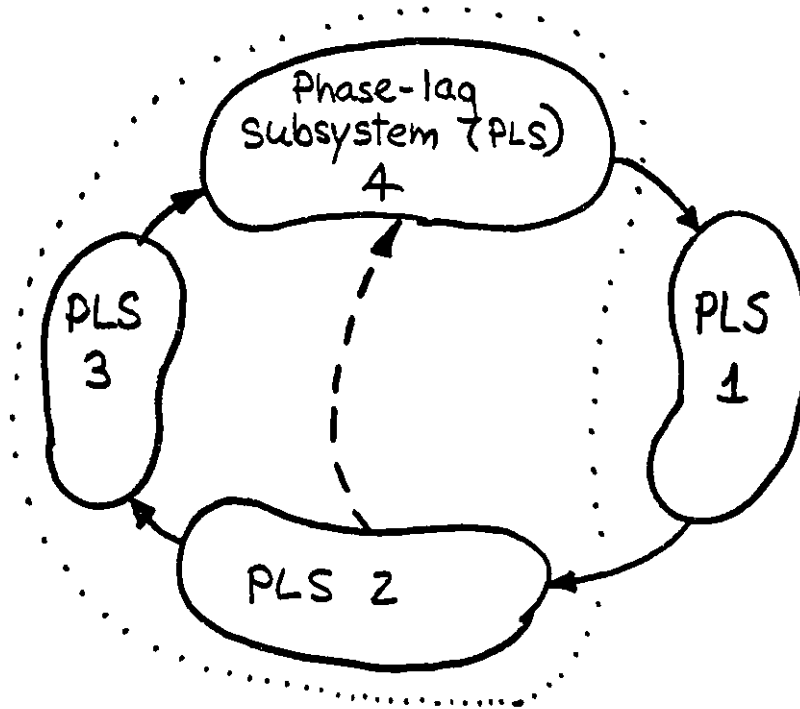


Figure 3-14.  
ABSTRACT REPRESENTATION OF ADDING A CROSS-LINK

it could through a series of three phase-lag subsystems. The principle on the origin of oscillations implies that oscillations can only arise from structures in which a disturbance can cause one phase-lag subsystem to depart from its steady-state condition, even while the other subsystems are at their steady-state conditions. Adding a cross-link reduces the ability of the phase-lag subsystems to move independently of one another, and thus the ability to oscillate out-of-phase with one another.

Adding a minor loop can be considered as one way of adding a cross-link. Figure 3-15 shows the equivalence between adding a minor



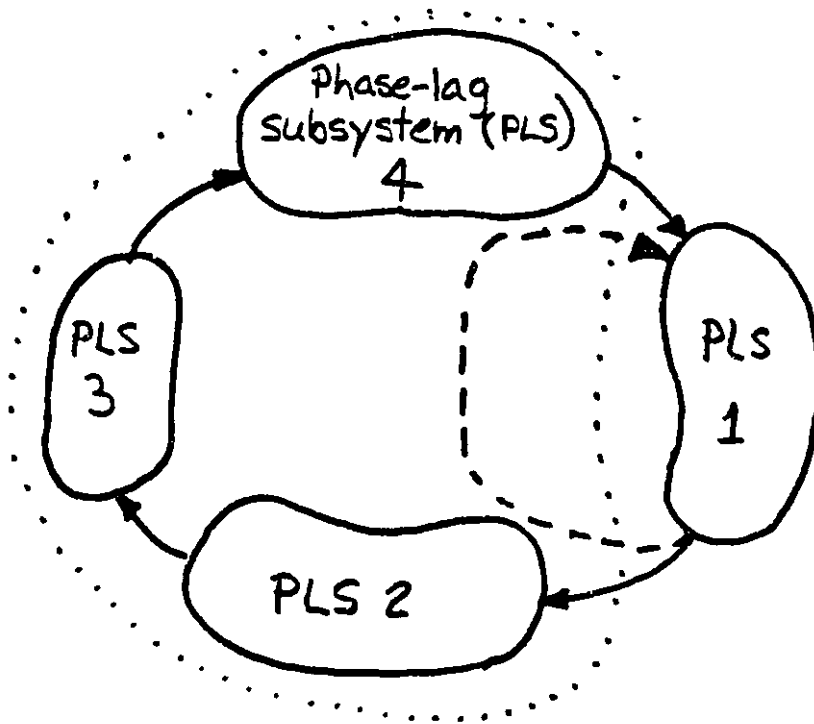


Figure 3-15.  
ABSTRACT REPRESENTATION OF ADDING A MINOR LOOP

loop around phase-lag subsystem one and adding a cross-link between the input to phase-lag subsystem two and the output of phase-lag subsystem four. Adding the minor loop reduces the ability of phase-lag subsystem one to oscillate out-of-phase with the other subsystems, first by reducing the magnitude of whatever oscillations it does show, and second, by coupling the movements of phase-lag subsystem one more closely to the output of phase-lag subsystem four.

Finally, reducing a system to an effectively-first-order system can be considered as an extreme form of adding a cross-link, where not only does one add a cross-link to reduce the ability of

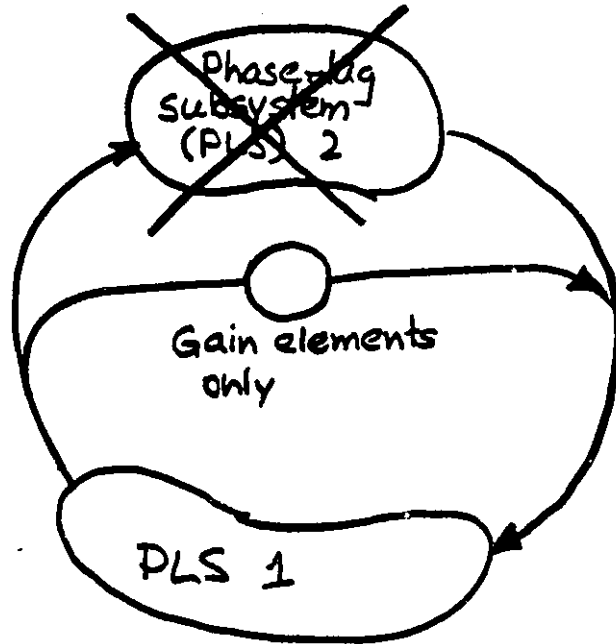


Figure 3-16.  
ABSTRACT REPRESENTATION OF REDUCTION TO A FIRST-ORDER SYSTEM

phase-lag systems to fluctuate out-of-phase with one another, but also one eliminates one of the phase-lag subsystems (which is replaced entirely by the cross-link). Figure 3-16 shows an abstract representation.

The previous examples and figures suggest a relatively simple principle:

Principle on adding cross-links

Adding cross-links between subsystems that reduce the ability of the subsystems to move out-of-phase with one another can reduce the tendency of the system to oscillate.

This principle is fairly abstract, but quite intuitive if interpreted, for example, in terms of the production-distribution system. Imagine going into someone's office, seeing them furiously ordering things to meet a high perceived demand, and saying "Excuse me, but the fellow who orders from the fellow who orders from you doesn't really have that much demand, so he's not going to be ordering much. Knowing that things are going to calm down in a bit, you'd better cool it with your own ordering and shipping. Otherwise, you'll be stuck in a bit with very few orders to fill." Structurally, your taking information from one part of the system and having it used in another corresponds to a cross-link. This example is illustrative of the way one can utilize the principle on adding cross-links in general: Examine the behavior and identify two sub-systems on an oscillatory loop that move substantially opposite one another. (In other words, not necessarily 180 degrees out of phase, but certainly from 90 to 270 degrees out of phase.) Then make a structural change that gives one subsystem advance warning about the disturbance that will be arriving due to a disturbance in the other subsystem.

The principle on adding cross-links also provides guidance during model refinement. If the model is unrealistically unstable, one can seek to identify real cross-links that are missing from the model structure. If two sectors of the model have very few connections

between them, one should always suspect missing cross-links: In a model of a firm, is it really true that the financial department has no direct influence on production planning? On occasion, informal ties between different departments (cross-links) are important in a system's behavior, and can be identified by interviewing, in this example, the people who do production planning.

Several caveats are in order about the practical ability of cross-links to reduce oscillation. First, not all cross-links in fact reduce oscillation, for if a given cross-link stabilizes a system, then reversing a sign in that cross-link may destabilize the system. The same warning applies to phase shifts as well: if a given cross-link (containing a level) produces a given phase shift, and stabilizes the system, then changing the phase shift by 180 degrees will probably destabilize the system.\*9\* In other words, one must consider the dynamics of the surrounding system to choose the phase and gain of a cross-link that will in fact stabilize the system; the proper phase and gain aren't implicit in the principle, which states only that such a cross-link exists.

The second caveat on cross-links concerns the magnitude of behavioral change one can obtain from adding cross-links. The idea of adding a cross-link is that the information coming through the

---

\*9\*Another possibility, less common but possible nonetheless, is that if the cross-link exerts a powerful-enough influence on the oscillation, changing the sign or phase shift may switch the system to a new oscillatory mode. Appendix M of Industrial Dynamics gives an example.

cross-link will in some sense oppose the information coming out of a phase-lag subsystem: when one says go up, the other says down, so the disturbance is not propagated as well. One may not be able to realistically attach enough credibility to the information coming through the cross-link to allow it to effectively oppose the information coming through normal channels. Indeed, the cross-link used in Appendix J of Industrial Dynamics is weighted heavily in relation to information from normal sources, and still adding the cross-link produces only a modest decrease in the system's tendency to oscillate.

The third and final caveat on cross-links concerns alternate methods of stabilizing a system. Forrester observes in Appendix J of Industrial Dynamics that policy changes within sectors (using only information from within each sector) provide performance equivalent or superior to that provided by adding a cross-link. This observation has since held true in other production-distribution systems as well. A simple explanation is suggested by the abstract representations in Figures 3-14 to 3-16: if the ability to oscillate depends on the ability of each phase-lag subsystem to propagate a disturbance with some phase lag, the oscillation--that is, the propagation of the disturbance--can in theory (and often in practice) be interrupted by alterations in a single phase-lag subsystem which attenuate the disturbance.

At present, there is no characterization of systems in which adding a cross-link is very effective, as opposed to only modestly effective in reducing oscillations; Section 6.3 discusses this and

related unanswered questions about the effect of cross-links on system behavior.

An Optimal Control Viewpoint. Figures 3-14 through 3-16 suggest that all of the methods of reducing a system's tendency to oscillate discussed thus far are special cases of adding cross-links. One can go much further: this subsection demonstrates mathematically, using only calculus, that the best way to stabilize a system is in general to add all possible cross-links (for a restricted but useful class of systems and definitions of "best"). The mathematical demonstration is due to Brockett.\*10\* Figure 3-17 shows schematically the system being analyzed: The behavior of a dynamic system is modified by raising or lowering the control inputs  $\underline{u}(t)$ . The mathematics that follows serves to demonstrate that optimally, each control input in general should have a cross-link originating from every state variable in the system. Readers not concerned with the intermediate mathematical steps can skim directly to the discussion of the optimal solution.

---

\*10\*Brockett, R.W., Finite-Dimensional Linear Systems (New York: Wiley and Sons, 1970).

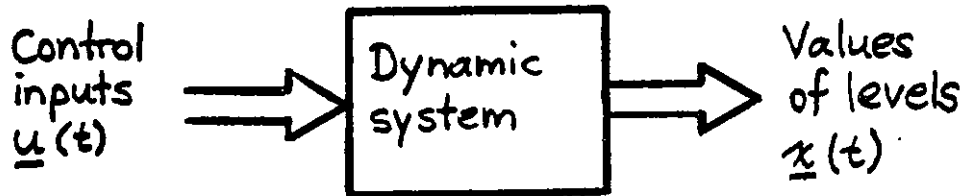


Figure 3-17.  
SCHEMATIC OF OPTIMAL CONTROL PROBLEM

The dynamics of the system are described by

$$\left( \frac{d}{dt} \right) \underline{x}(t) = \underline{A}(t) \underline{x}(t) + \underline{B}(t) \underline{u}(t)$$

where  $\underline{x}(t)$  is the vector of levels,  $\underline{u}(t)$  is the vector of control inputs, and  $\underline{A}(t)$  and  $\underline{B}(t)$  are (possibly time-varying) matrices that couple the state and the inputs to the rates. The levels have been normalized so that their equilibrium values are zero. Note that the equation above can provide an approximate (i.e., linearized) description of a nonlinear system as well.

The problem, then, is to find a set of control inputs as a function of the levels  $\underline{x}(t)$  and possibly time. (Since the structure may be time-varying, we should allow for the possibility of a control law or policy that varies with time.) The criterion for a good control policy is to minimize

$$V = \int_{t_1}^{t_2} \underline{x}'(t) \underline{L}(t) \underline{x}(t) + \underline{u}'(\underline{x}(t), t) \underline{u}(\underline{x}(t), t) dt + \underline{x}'(t_2) \underline{Q} \underline{x}(t_2)$$

where  $\underline{L}$  and  $\underline{Q}$  are positive definite matrices that define the cost in some sense of departing from equilibrium during and at the end of, respectively, the time the behavior is being controlled, and the cost of controlling the system is just

$$\int_{t_1}^{t_2} \underline{u}'(\underline{x}(t), t) \underline{u}(\underline{x}(t), t) dt$$

or the integral of the magnitude of the control vector. The minimization should hold for whatever initial conditions  $\underline{x}(t_1)$  hold. Note that the criterion above can be considered as the linear and quadratic terms in a Taylor series approximation of a more general criterion around an optimal trajectory (for which the linear term is zero).

Although the mathematical form of the problem as stated, is fairly simple, the form is that of an approximation to a much larger set of problems. Thus, we might expect the characteristics of the solution to appear in a much larger set of situations as well.

To solve this optimization problem, we first examine the properties of the following quadratic form:

$$\underline{x}'(t) \underline{K}(t) \underline{x}(t) \Big|_{t_1}^{t_2}$$



(The motivation for the examination is that the optimal cost criterion  $V$  may very well have that form, since its definition is completely quadratic, and the initial conditions  $\underline{x}(t_1)$  are the only remaining determinants of optimal cost, once the optimal policy has been determined.) For simplicity,  $\underline{x}(t)$ ,  $\underline{u}(\underline{x}(t), t)$ ,  $\underline{A}(t)$ ,  $\underline{B}(t)$ ,  $\underline{L}(t)$ , and  $\underline{K}(t)$  are denoted by  $\underline{x}$ ,  $\underline{u}$ ,  $\underline{A}$ ,  $\underline{B}$ ,  $\underline{L}$ , and  $\underline{K}$ , respectively.

$$\underline{x}' \underline{K} \underline{x} \Big|_{t_1}^{t_2} = \int_{t_1}^{t_2} (d/dt) (\underline{x}' \underline{K} \underline{x}) dt$$

From the definition of an integral, and using the chain rule and the system equations, the quadratic form becomes

$$\begin{aligned} &= \int_{t_1}^{t_2} [\underline{A} \underline{x} + \underline{B} \underline{u}]' \underline{K} \underline{x} + \underline{x}' \underline{K} [\underline{A} \underline{x} + \underline{B} \underline{u}] + \underline{x}' ((d/dt) \underline{K}) \underline{x} dt \\ &= \int_{t_1}^{t_2} \underline{x}' \underline{A}' \underline{K} \underline{x} + \underline{u}' \underline{B}' \underline{K} \underline{x} + \underline{x}' \underline{K} \underline{A} \underline{x} + \underline{x}' \underline{K} \underline{B} \underline{u} + \underline{x}' ((d/dt) \underline{K}) \underline{x} dt \\ &= \int_{t_1}^{t_2} [\underline{u}' \underline{x}'] \begin{bmatrix} \underline{0} & \underline{B}' \underline{K} \\ \underline{K} \underline{B} & (d/dt) \underline{K} + \underline{K} \underline{A} + \underline{A}' \underline{K} \end{bmatrix} \begin{bmatrix} \underline{u} \\ \underline{x} \end{bmatrix} dt \end{aligned}$$

So

$$\begin{aligned} 0 &= \int_{t_1}^{t_2} [\underline{u}' \underline{x}'] \begin{bmatrix} \underline{0} & \underline{B}' \underline{K} \\ \underline{K} \underline{B} & ((d/dt) \underline{K}) + \underline{K} \underline{A} + \underline{A}' \underline{K} \end{bmatrix} \begin{bmatrix} \underline{u} \\ \underline{x} \end{bmatrix} dt \\ &\quad - (\underline{x}'(t_2) \underline{K}(t_2) \underline{x}(t_2) - \underline{x}'(t_1) \underline{K}(t_1) \underline{x}(t_1)) \end{aligned}$$

Returning to the expression for cost,

$$V = \int_{t_1}^{t_2} [\underline{u}' \underline{x}] \begin{pmatrix} \underline{I} & \underline{0} \\ \underline{0} & \underline{L} \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{x} \end{pmatrix} dt + \underline{x}'(t_2) \underline{Q} \underline{x}(t_2)$$

Adding the identity for zero derived above,

$$\begin{aligned} V &= \int_{t_1}^{t_2} [\underline{u}' \underline{x}'] \begin{pmatrix} \underline{I} & \underline{0} \\ \underline{0} & \underline{L} \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{x} \end{pmatrix} dt + \underline{x}'(t_2) \underline{Q} \underline{x}(t_2) \\ &+ \int_{t_1}^{t_2} [\underline{u}' \underline{x}'] \begin{pmatrix} \underline{0} \\ \underline{KB} \end{pmatrix} ((d/dt)(\underline{K})) + \begin{pmatrix} \underline{B}' \underline{K} \\ \underline{KA} + \underline{A}' \underline{K} \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{x} \end{pmatrix} dt \\ &- (\underline{x}'(t_2) \underline{K}(t_2) \underline{x}(t_2) - \underline{x}'(t_1) \underline{K}(t_1) \underline{x}(t_1)) \\ &= \int_{t_1}^{t_2} [\underline{u}' \underline{x}'] \begin{pmatrix} \underline{I} \\ \underline{KB} \end{pmatrix} \begin{pmatrix} \underline{L} + ((d/dt)(\underline{K})) + \underline{KA} + \underline{A}' \underline{K} \\ \underline{B}' \underline{K} \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{x} \end{pmatrix} dt \\ &+ \underline{x}'(t_1) \underline{K}(t_1) \underline{x}(t_1) + (\underline{x}'(t_2) [\underline{Q} - \underline{K}(t_2)] \underline{x}(t_2)) \end{aligned}$$

Thus far, nothing has been assumed about  $\underline{K}(t)$ ; the equations above hold for all differentiable choices of  $\underline{K}(t)$ . The time has come to make a fortuitous choice for  $\underline{K}(t)$ . Let

$$\underline{K}(t_2) = \underline{Q}$$

to cancel out the final term in the expression for  $V$  above, and let

$$\dot{\underline{K}} = -\underline{AK} - \underline{KA} - \underline{L} + \underline{KBB}'\underline{K}$$

so that we "complete the square" with the expression for  $V$  simplifying to

$$\begin{aligned}
 V &= \int_{t_1}^{t_2} [\underline{u}' \underline{x}'] \begin{pmatrix} \underline{I} & \underline{B}' \underline{K} \\ \underline{K} \underline{B} & \underline{K} \underline{B} \underline{B}' \underline{K} \end{pmatrix} \begin{pmatrix} \underline{u} \\ \underline{x} \end{pmatrix} dt + \underline{x}'(t_1) \underline{K}(t_1) \underline{x}(t_1) \\
 &= \int_{t_1}^{t_2} [\underline{u} + \underline{B}' \underline{K} \underline{x}]' [\underline{u} + \underline{B}' \underline{K} \underline{x}] dt + \underline{x}'(t_1) \underline{K}(t_1) \underline{x}(t_1)
 \end{aligned}$$

The expression for the cost  $V$  above holds for any choice of  $\underline{u}$ . The choice of  $\underline{u}(\underline{x}(t), t)$  that minimizes  $V$  is obviously

$$\underline{u} = [-\underline{B}' \underline{K}] \underline{x}$$

or more formally

$$\underline{u}(\underline{x}(t), t) = [-\underline{B}'(t) \underline{K}(t)] \underline{x}(t)$$

We have a boundary condition and a differential equation for  $\underline{K}(t)$ , so we can numerically compute  $\underline{K}(t)$  and thus  $-\underline{B}'(t) \underline{K}(t)$ . The expression in brackets gives the parameters of a control policy, independent of the initial or subsequent states.

Looking again at the form of the optimal control policy, it is clear that each element of the control vector--each channel through which the system's behavior can be manipulated--to be optimally

controlled must in principle respond to the value of every level in the system.\*11\* In other words, within the constraints posed by a limited number of control inputs, the optimal solution calls for adding all possible cross-links.

-----

\*11\*It is undoubtedly possible to construct examples where many elements of  $-\underline{B}(t)\underline{K}(t)$  are zero, that is, where the optimal control does not always depend on every level. The probability of such a condition occurring by happenstance, however, seems remote. For this to occur, for some element  $ij$  of  $K$ ,

$$[(d/dt)K]_{ij} = [-\underline{AK} - \underline{KA} - \underline{L} + \underline{KBB}'K]_{ij} = 0 \quad \forall t \in [t_1, t_2]$$

which would seem fairly unlikely in a model well-focused on a particular set of dynamic behavior modes.

### 3.3. ADDING A MINOR NEGATIVE LOOP WITH A DELAY

Sections 3.1 and 3.2 in effect answered the question, "Given a desire to diminish or eliminate oscillations, what parametric or structural changes can accomplish this?" Section 3.3 provides one answer to an opposite sort of question: "Given a contemplated structural change, what is the effect on oscillatory behavior?" The structural change thus analyzed in Section 3.3 is a minor negative loop containing a delay.

An Example: Influence of Availability on Ordering. Section 2.2 described one constraint on the size of the backlog  $B$  in the employment-backlog system, where orderers attempted to maintain a given number of orders in the backlog. There is also another effect that regulates the size of the backlog  $B$ , which operates over longer time horizons. If the order backlog of a company is too high, orderers are not getting their orders filled right away. There is a delivery delay, which, if it persists at high values for a significant amount of time, will cause the company's customers, the orderers, to go elsewhere to obtain the products they want. As the company loses customers, the order rate is diminished, all other things being equal, so that not as many new orders are added to the backlog, which tends to reduce the size of the backlog. Inversely, if the backlog is very low, orders are filled quickly, which makes the company's product attractive to customers, because the customers need not maintain as much of their own stocks of the product, and yet the product will always be readily

available. So a low delivery delay increases the order rate, which tends to drive up the order backlog; the product-availability loop is negative.

Figure 3-18 shows how the influence of product availability can be modeled in the employment-backlog system. The order rate OR is modulated by the multiplier from delay on ordering MDO, which responds to the perceived delay for ordering PDO. This variable is a two-year smoothing of the actual delivery delay DD, which is the average time it takes an order to be filled. DD is computed by dividing the backlog B by the rate at which orders are filled, the output OUT.

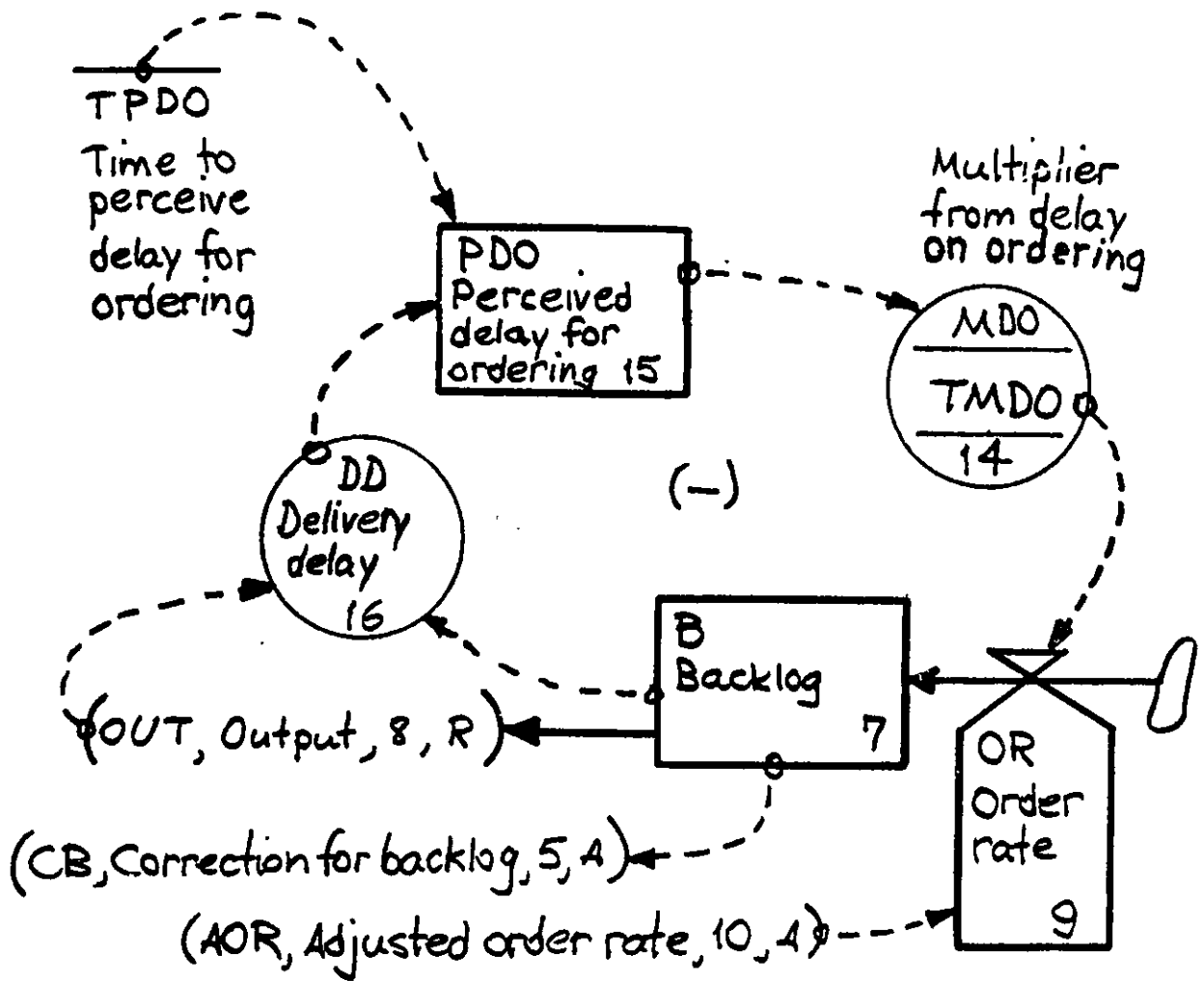


Figure 3-18.  
PRODUCT-AVAILABILITY LOOP

The equations for backlog B, output OUT, and order rate OR are already familiar from Section 2.1:

$$B.K = B.J + (DT)(OR.JK - OUT.JK)$$

$$B = BN$$

$$BN = 660$$

- B - BACKLOG (UNITS)
- OR - ORDER RATE (UNITS/YEAR)
- OUT - OUTPUT (UNITS/YEAR)
- BN - BACKLOG INITIAL (UNITS)

7, L  
7.1, N  
7.2, C

OUT.KL=E.K\*PROD\*(1-WTO)+TOUT.K\*WTO 8, R  
WTO=0 8.1, C

OUT - OUTPUT (UNITS/YEAR)  
E - EMPLOYMENT (MEN)  
PROD - PRODUCTIVITY (UNITS/YEAR/MAN)  
WTO - WEIGHTING ON TEST OUTPUT (DIMENSIONLESS)  
TOUT - TEST OUTPUT (UNITS/YEAR)

OR.KL=(AOR.K)(MDO.K) 9, R

OR - ORDER RATE (UNITS/YEAR)  
AOR - ADJUSTED ORDER RATE (UNITS/YEAR)  
MDO - MULTIPLIER FROM DELAY ON ORDERING  
(DIMENSIONLESS)

The multiplier from delay on ordering MDO simulates the long-term effect of product availability (represented by perceived delay for ordering PDO). Figure 3-19 shows a graph for MDO. For consistently low product availability (represented by a high PDO), the multiplier diminishes the order rate OR below what it otherwise would be. This diminution represents the company losing customers. At the other end of the graph, high product availability (low perceived delivery delay PDO) stimulates ordering.

MDO.K=TABLE(TMDO,PDO.K/DBC,0,4,.5) 14, A  
TMDO=1.3/1.2/1/.75/.5/.3/.2/.15/.1 14.1, T

MDO - MULTIPLIER FROM DELAY ON ORDERING  
(DIMENSIONLESS)

TMDO - TABLE FOR MULTIPLIER FROM DELAY ON ORDERING  
PDO - PERCEIVED DELAY FOR ORDERING (YEARS)  
DBC - DESIRED BACKLOG COVEARGE (YEARS)



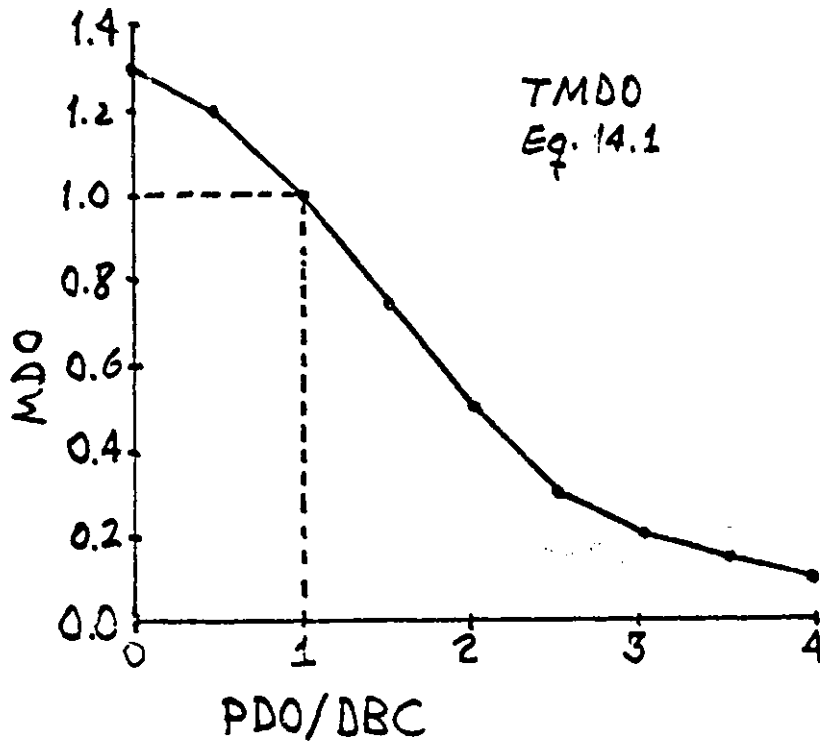


Figure 3-19.

TABLE FOR MULTIPLIER FROM DELAY ON ORDERING PMDO

The perceived delay for ordering PDO represents the perception of product availability over an averaging period of two years. In the basic employment-backlog model the time to perceive delay for ordering TPDO is set at 1E11, so that PDO effectively never varies from its initial value, and the multiplier from delay on ordering MDO remains equal to 1.0. In this section, MDO is activated by setting TPDO equal to two years.

$$PDO.K = PDO.J + (DT/TPDO)(DD.J - PDO.J)$$

$$PDO = DBC$$

$$TPDO = 1E11$$

PDO - PERCEIVED DELAY FOR ORDERING (YEARS)

TPDO - TIME TO PERCEIVE DELAY FOR ORDERING

DD - DELIVERY DELAY (YEARS)

DBC - DESIRED BACKLOG COVERAGE (YEARS)

15, L

15.1, N

15.2, C

The delivery delay DD is the average time taken for an order to be filled, which is just the order backlog B divided by the rate of filling orders, the output OUT.

DD.K=B.K/OUT.JK 16, A  
DD - DELIVERY DELAY (YEARS)  
B - BACKLOG (UNITS)  
OUT - OUTPUT (UNITS/YEAR)

Figure 3-20 shows the behavior of the employment-backlog system with the product-availability loop activated. For easy comparison, the behavior without MDO active is shown in the succeeding figure, Figure 3-21 (identical to Figure 2-14). Comparing the curves for backlog B reveals little overall change in the oscillatory behavior: a slightly shorter period and slightly less damping result from activating the product-availability loop, a minor negative loop with a delay.\*12\*

In Section 2.2, activating a minor loop around the level of backlog B increased the damping. Here, activating a minor loop with a delay around the backlog B decreases the damping. What is there about a delay that causes such a difference? The following subsection begins to explain.

-----

\*12\*"Adding a minor loop with a delay" here denotes adding a loop that contains one additional level, a delay, around an already-present level. "Minor loop with a delay" borders on being a misnomer (because minor loops contain no levels), but is retained for convenience.

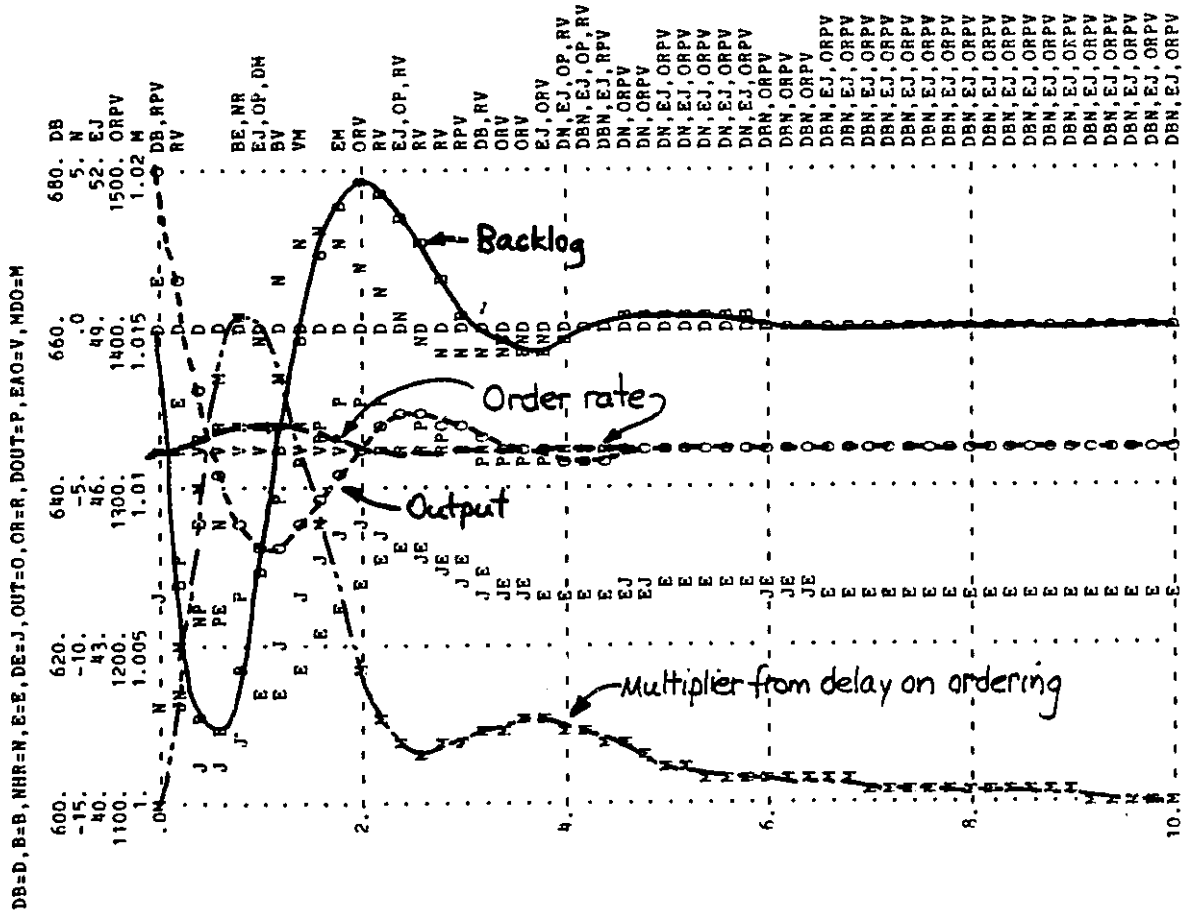


Figure 3-20.  
 EMPLOYMENT-BACKLOG SYSTEM WITH ACTIVE PRODUCT-AVAILABILITY LOOP

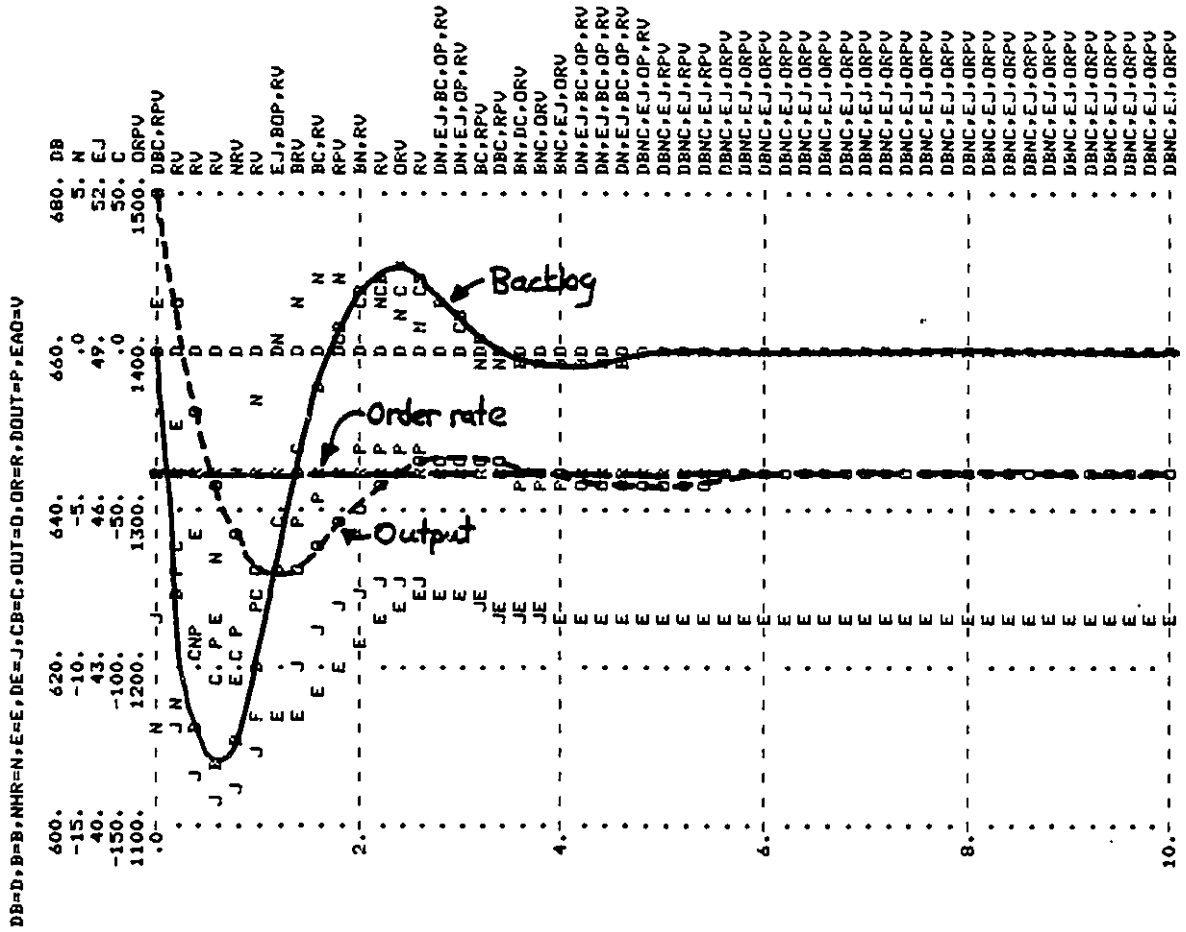


Figure 3-21.  
BASE SIMULATION (SAME AS FIGURE 2-14)

Explanation in Terms of Phase Shift. Before examining in detail the dynamics of the entire employment-backlog system with the multiplier from delay on ordering MDO activated, let us first examine the dynamics of only the subsystem shown in Figure 3-18.\*13\* We can isolate the subsystem by changing a parameter in the equation for the rate of output OUT, so that OUT responds to an exogenous test output TOUT, rather than the rate determined by the number of employees E within the system. By changing the weighting on test output WTO from 0.0 to 1.0, we eliminate the effect of employment E on output OUT, which will now respond only to the test output TOUT.

We can test the response of the backlog subsystem to a step in the rate of output OUT, by setting the test output step height TOSH to 132 units per year from zero. Figure 3-22 shows the result. The negative loop connecting the level of backlog B and the other level, the perceived delay for ordering PDO, causes a damped oscillation, with a period of about ten or twelve years. In effect, a step in the rate of output OUT resets the steady-state values for both levels. The

-----

\*13\*The motivation for examining the dynamics of the subsystem comes from two features of the simulation in Figure 20: first, the effect of a varying multiplier from delay on ordering MDO on the order rate OR is difficult to see because the behavior is still fairly damped, and the magnitude of the effect of MDO quickly becomes very small. Second, examining the curve for MDO reveals an apparently exponential transient with a time constant of about three years, upon which the oscillation is superimposed. The transient makes it difficult to analyze MDO or OR in terms of disturbance from steady-state values.

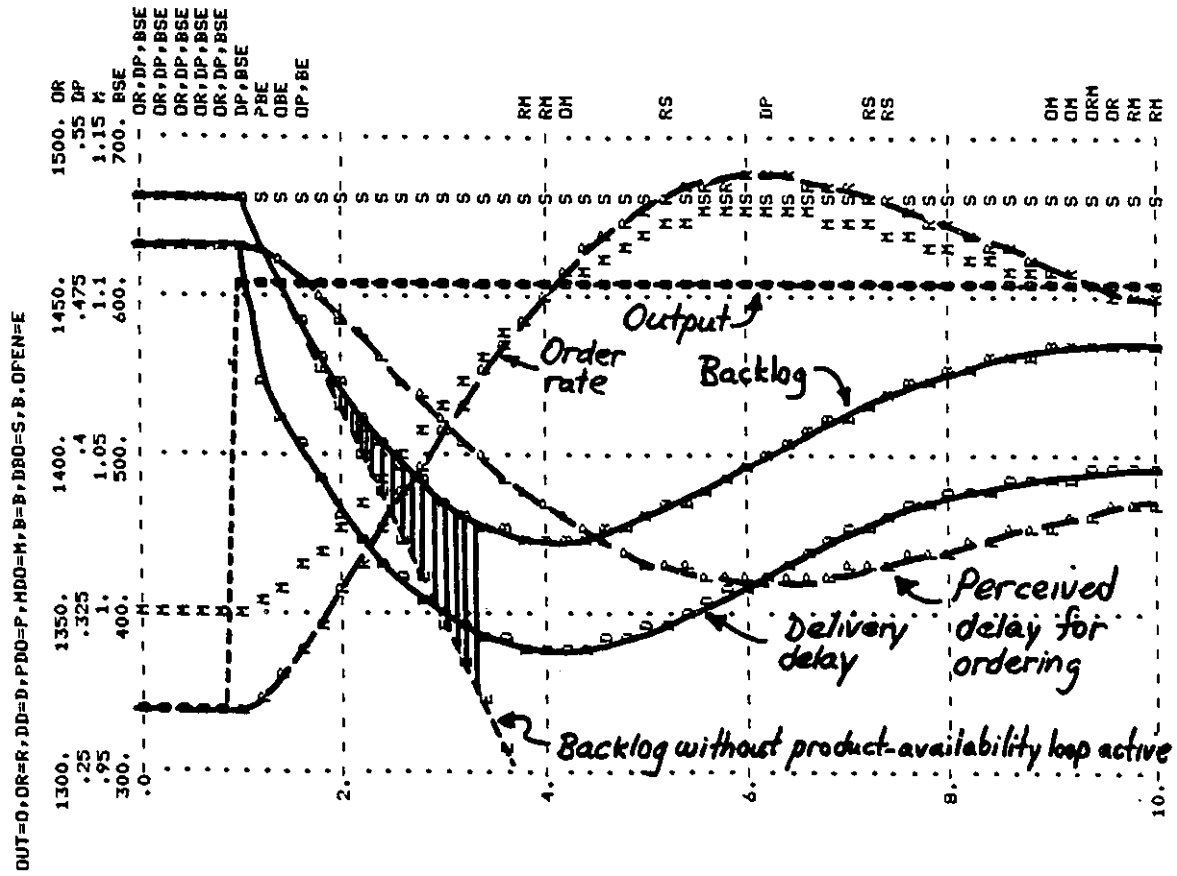


Figure 3-22.  
 STEP RESPONSE OF BACKLOG SUBSYSTEM OF FIGURE 3-18

perceived delay for ordering PDO must be lower than before to raise the multiplier from delay on ordering MDO and the order rate OR to equal the new higher rate of output OUT. The backlog B must therefore also equilibrate at a lower value to produce a lower delivery delay DD, which is averaged to produce PDO. Initially, then, both the backlog B and perceived delay for ordering PDO exceed their steady-state values. When the output OUT steps up at year 1, the backlog B begins to drop, causing the delivery delay DD to drop. Eventually, PDO (a two-year lag of DD) begins to drop. When the backlog B reaches its steady-state value, the perceived delay for ordering PDO has yet to reach its steady-state value, so the multiplier from delay on ordering MDO and the order rate OR are still below their respective steady-state values, so that the difference between the rate of output OUT and the order rate OR continues to reduce the backlog B. Thus, when the perceived delay for ordering PDO reaches its steady-state value, and the order rate OR and the rate of output OUT are equal, the backlog B and the actual delivery delay DD are below their respective steady-state values. So PDO continues to decline, and the oscillation continues, with the levels being alternately disturbed from their steady-state values.

What effect can a loop that oscillates with a period of ten or twelve years have on another loop that oscillates with a period around three years? One beginning point in answering this question is to reexamine Figure 3-22, which actually plots two simulations of the backlog subsystem: one with the multiplier from delay on ordering MDO

activated (which was discussed above), and another simulation with MDO set at 1.0. In this second simulation, there is no feedback from the backlog B to the order rate OR, so the initial discrepancy between the rate of output OUT and the order rate OR remains constant, and produces a steady downward-sloping ramp in the backlog B. Thus, the difference between the two curves for backlog B on Figure 3-22 results from the action of MDO. The difference between the two curves remains fairly small for about three-quarters of a year after the initial step. At that point, the curves begin to diverge markedly, with one continuing to decline (the curve for the simulation without an active MDO), and one beginning to level out (the curve for the simulation with MDO active). Thus, even though the resonant period of the subsystem is on the order of a decade or more, the negative loop begins to significantly alter the behavior of backlog B after only a few months.

The effects of the negative loop in the backlog subsystem on a three-year oscillation can be seen in Figure 3-23.\*14\* The backlog B still lags about 90 degrees behind its exogenous driving input, the rate of output OUT. Because of the two-year delay in perceiving the actual delivery delay, the perceived delay for ordering PDO lags another 48 degrees behind the backlog B. Thus, the order rate OR fluctuates with a

-----

\*14\*As noted previously, the backlog subsystem shows a long transient, which we can now associate with its ten- or twelve-year oscillatory period. To eliminate this transient, the simulation in Figure 3-23 is plotted starting at year 18 through year 23. By this time, the three-year oscillation has come quite close to its steady-state behavior.



phase lag of about 138 degrees, not very far away from a 180-degree phase shift.\*15\* Thus, due to the delay in the negative loop, the order rate OR fluctuates with a phase shift such that most of the time, it is augmenting the effects of the sinusoidal disturbance in the rate of output OUT. The striped areas enclosed by the order rate OR and its equilibrium value indicate the periods of time when the order rate OR is changing the backlog B in the same direction as the rate of output OUT. The dotted areas indicate times during which the effects of the order rate OR and the output OUT oppose one another on the backlog B. Clearly, at least in this case, the preponderant effect of adding the negative loop is to add to the effect of the disturbance, and increase the magnitude of oscillation.

The magnitude of oscillation of backlog B is increased due to the phase shifting of compensation through the multiplier from delay on ordering MDO. To what extent does this phase-shift effect depend on the value of the time to perceive delay for ordering TPDO (two years)?

-----

\*15\*The effect of the two-year delay in perceived delay for ordering PDO is somewhat diminished by the formulation for delivery delay DD, which varies not only in response to excursions in the backlog B, but also responds to excursions in the rate of output OUT. In Figure 3-23, the actual delivery delay DD leads fluctuations in backlog B by about 36 degrees. This must always be the case because of the way DD, B, the order rate OR, and the rate of output OUT are related. Imagine a trough (minimum) in backlog B, where the rate of output OUT has been declining, and has just equaled the order rate OR. (The backlog B attains its maxima or minima when OUT equals OR.) At that point, the delivery delay DD (which equals B/OUT) will already have passed its minimum point, because with backlog B unchanging and the output OUT declining, DD must already be beginning to rise.

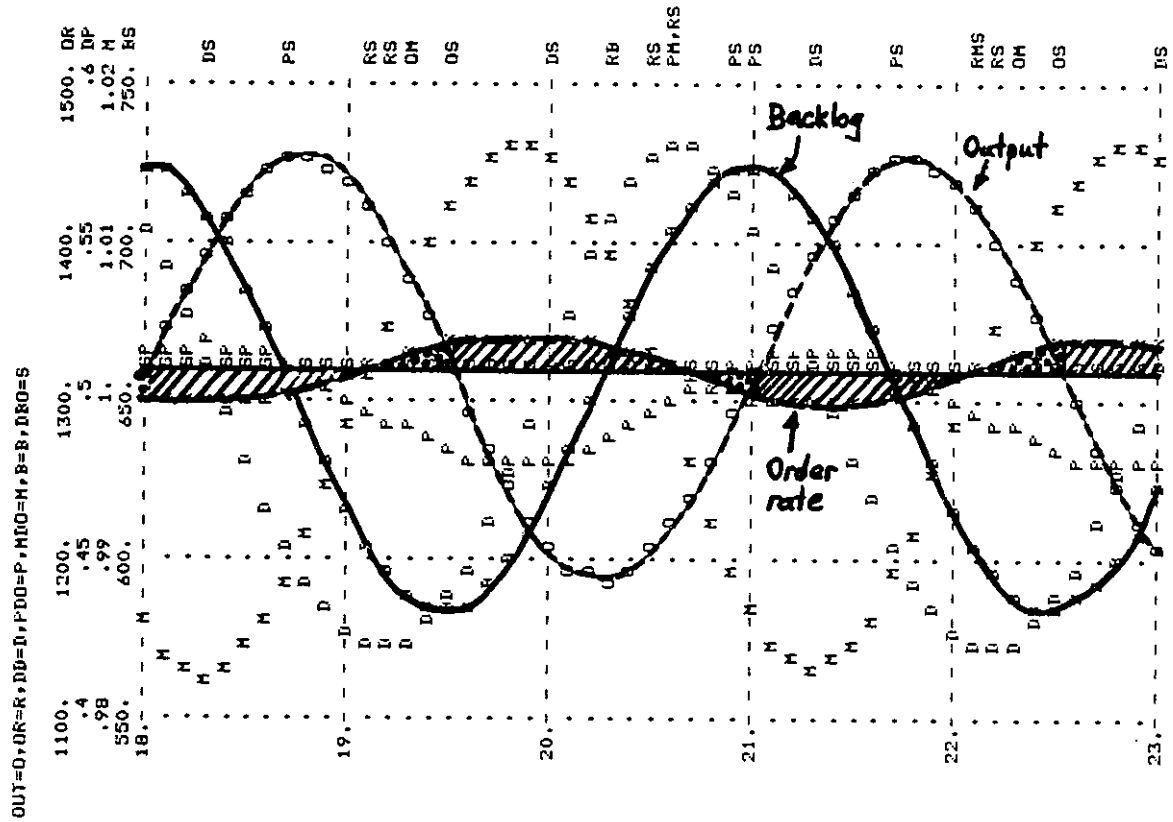


Figure 3-23.  
STEADY-STATE SINUSOIDAL RESPONSE OF BACKLOG SUBSYSTEM

relative to the period (three years)? Figure 3-24 shows a simulation with TPDO reduced by one-half to 1.0. Two phenomena are apparent. First, there is indeed a smaller phase shift, so that the amount of time that movements in the order rate OR oppose those of the rate of output OUT is larger relative to the amount of time that OR and OUT move the backlog B in the same direction. Second, the fluctuations in the order rate OR are larger, since the one-year delay attenuates the signal less than does the two-year delay. The smaller phase shift tends to diminish the magnitude of oscillations in the backlog B; the larger amplitude of fluctuation tends to increase fluctuations in B. The net effect of reducing the time to perceive delay for ordering TPDO is to leave the fluctuations in backlog B at about the same magnitude as when TPDO is 2.0. Apparently, then, the delay time must be fairly small relative to the period of oscillation (perhaps one-fifth or less) in order for the negative loop to increase rather than decrease the damping.

Consider now the effect of the product-availability loop in the context of the entire employment-backlog system. Figure 3-20 shows that between year 0.5 and year 2, the rate of output OUT and the order rate OR (which fluctuates as a result of the operation of the product-availability loop) do indeed move in opposite directions, so that the effect of the product-availability loop during the first two years is to increase the magnitude of movement in the backlog B. Figure 3-25 shows an expanded view of the system's behavior between



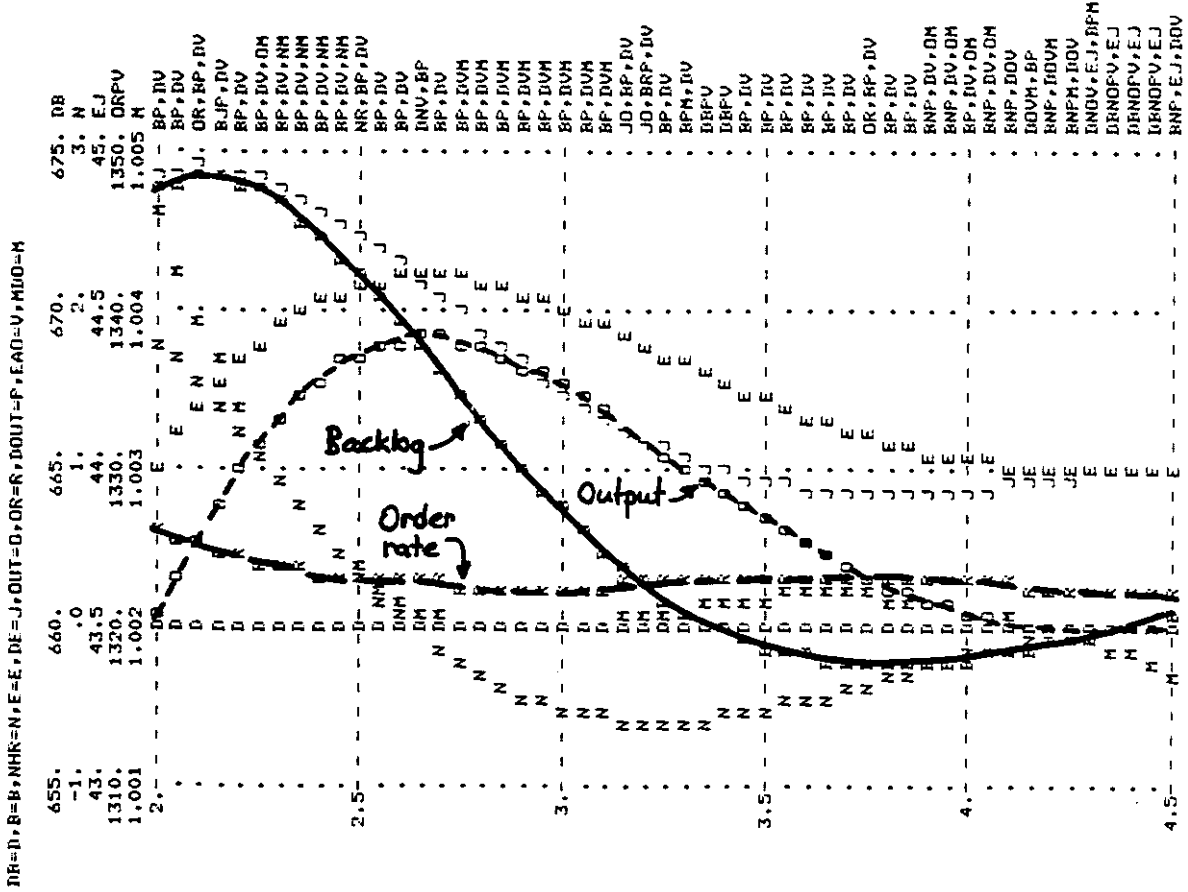


Figure 3-25.  
EXPANDED VIEW WITH ACTIVATED PRODUCT-AVAILABILITY LOOP

years 2 and 4.5. Roughly the same phase relationships obtain: the order rate OR decreases at about the same time that the rate of output OUT is increasing, and vice versa, which suggests that the effect of the minor negative loop with a delay is indeed to increase the magnitude of oscillation, and therefore to decrease the damping.

Figure 3-20 shows one additional effect of the minor loop with a delay that could not be directly predicted from the analysis of the backlog subsystem: the period of oscillation is shorter with the minor loop with a delay in operation. To see why, imagine the subsequent behavior of the system if the employment E were at its steady-state value, and the backlog B was below its steady-state value. (This happens in Figure 3-20 at about year 0.5.) The effect of the minor loop with a delay is to move the backlog B more rapidly toward its steady-state value (and beyond it as well), so that B will reach its steady-state value sooner, and employment E can cease its decline and begin to rise sooner. In other words, the minor loop with a delay increases the velocity of a system's movement through its cycles in state space.

A Principle. The description above explains the effect of adding a minor negative loop with a delay on oscillations in terms of phase shift. That explanation, however, does not suggest the ubiquity of the result. In terms of phase shift, the result seems to derive from a rather special relationship among the gains and phase shifts in the various parts of the system. Is there a way of describing the effect of

a minor negative loop with a delay such that it becomes clear why the effect happens so often? To begin that explanation, let us first identify the concepts in the previous explanation that are useful as independent, free-standing concepts.

One element of a useful explanation is contained in the commentary on Figure 3-22. That figure demonstrates that the added minor loop with a delay begins to affect the level around which it is added almost immediately; the magnitude of the effect, however, becomes larger as time passes. Compensation that arrives almost immediately will tend to oppose movement of the level. Compensation that arrives later (out-of-phase) will tend to augment the movement of the level. If the magnitude of the compensation becomes larger over time, the augmenting effect outweighs the opposing effect, and the added minor negative loop has an appreciable destabilizing effect on oscillations much shorter than its own resonant frequency. The natural period of the loop added to the backlog subsystem is around ten years, and it still has a destabilizing effect on the three-year oscillation of the employment-backlog system.

The other useful concept that can be abstracted from the previous explanation is that of accelerated movement through state space. The effect of adding the minor loop with a delay in the employment-backlog system was to accelerate the movement of the backlog B, both toward and beyond its steady-state value. It is this accelerated movement that causes both the shortened period and decreased damping of the system's oscillation. We can go beyond the concepts that

arose in the previous explanation to observe that the accelerated movement through state space of the backlog B is caused by a disturbance in the added level of the perceived delay for ordering PDO. This disturbance in turn arose from disturbances in the rate of output OUT and the backlog B. In other words, the accelerated movement through state space results from the addition of a pathway through which disturbances can propagate. This observation provides enough descriptive material to state a principle:

Principle on adding a minor negative loop with a delay

If a minor negative loop with a delay is added around a level already on an oscillatory loop, the added loop forms another pathway through which disturbances in the level can propagate back to the level. When the additional disturbance returns to the level, it moves the level more rapidly to and past its steady-state value, which results in a shorter period and less stable oscillation.\*16\*

Figure 3-26 illustrates the principle by showing the original system as one loop and the added loop as a pathway symmetrical with the original loop (the auxiliary variables are subsumed within the rate symbols for simplicity). The original system is one pathway through which a disturbance can propagate to cause oscillation in the backlog B. The added loop constitutes another system through which disturbances can propagate to cause oscillations in the backlog B.

---

\*16\*Less "stable" in this context means that the system exhibits a larger number of oscillations before coming essentially to equilibrium. In this sense, then, the oscillations described in the principle, which have shorter periods than the original oscillations, would be described as less stable, even if the time-constant of the exponential envelope around the oscillations remains the same.



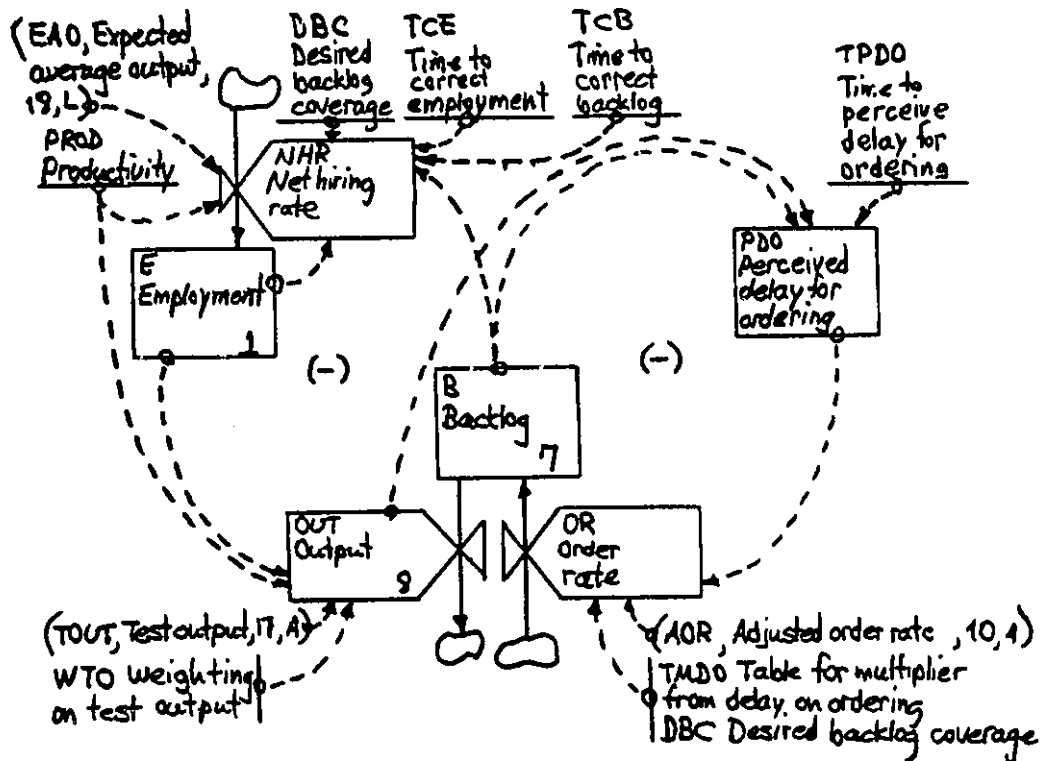


Figure 3-26.

ADDING A MINOR LOOP WITH A DELAY AS ADDING AN OSCILLATORY PATHWAY

In complex social and economic systems, it is a common occurrence to have a level be in part controlled by a minor negative loop with a delay: because it often takes time to perceive and act to change the system condition represented by the level, there are often delays in the negative loops whose goal is to control the level. Therefore, there are probably many opportunities in model formulation and policy design to use the principle on adding a minor negative loop with a delay. In policy design, such loops are immediate suspects for exacerbating the

oscillatory tendencies of the system. If that turns out to be the case, the system can be made more stable by either reducing the delay time (far enough that the loop effectively becomes a pure minor loop, which usually tends to stabilize oscillatory systems) or increasing the delay time (far enough that the delay filters out most of the inputs at the natural frequency of the system). In model formulation, if a preliminary model is unrealistically stable or has an unrealistically long response time, the principle above suggests that one scrutinize the real system for additional delays in minor loops presently in the model, and additional minor loops with delays not presently in the model. Many real decisions (price-setting, for example) are influenced by many streams of information, so that it is often relatively easy to identify additional streams of information, omitted from the preliminary model, that could have a significant influence on the system behavior by forming a minor negative loop with a delay.

There are caveats concerning the principle above: It seems possible that if the gain of the added minor negative loop with a delay is large enough, the added loop will dominate in controlling the level, rather than the original loop. This domination may cause oscillation and damping more characteristic of the added loop than the original system. At this point, all one can say is that there is very little qualitative knowledge existant about such phenomena; Section 6.3 suggests a means of investigation.

The second caveat concerns the delay time. If it is made short enough, the minor negative loop with a delay (which previously

may have destabilized the system's oscillations) effectively becomes a pure minor negative loop (which usually tends to stabilize the oscillations).

The example of the employment-backlog system may seem somewhat specialized, for the natural period of the added loop considerably exceeds the natural period of the original oscillating loop. We have seen that the slower loop is capable of exaggerating oscillations resulting from the more rapidly-oscillating loop. If the original system were the slowly-oscillating system, we would expect the oscillatory tendencies of the added loop to increase. In other words, there is a symmetry between the original loop and the added loop, and both loops acting in concert tend to produce more movement of the common level than either does alone, regardless

of which loop has the shorter natural period.\*17\* \*18\*

-----

\*17\*The magnitude of the effect of adding a minor loop with a delay will of course depend on the period and damping ratios of the two loops. One suspects that coupling two loops with identical natural periods would result in a considerable decrease in damping, in contrast to the mild decrease shown in the example in this section. Section 6.3 discusses questions for further research that arise out of the configuration shown in Figure 3-26, such as determining the natural period and damping of the total system as a function of the natural periods and damping ratios of the individual loops.

\*18\*The preceding material offers a tentative explanation for a curious phenomenon in economic dynamics. National economies, which are made of numerous disparate oscillating loops, might be expected to resonate at numerous frequencies. In fact, the detectable resonant frequencies are rather widely separated at 4, 20, and 50-year periods. This separation could be due to the ability of coupled negative loops (such as those in Figure 3-26) to become entrained at a resonant frequency higher (shorter period) than the resonant frequency of either loop oscillating independently. If a 4-year business cycle is able to entrain loops with resonant periods up to 15 years into the 4-year cycle, then a 20-year independent natural period is the next periodicity that will not simply be "captured" by the 4-year oscillation.

## CHAPTER FOUR

### POSITIVE LOOPS

Chapter 4 concludes the exposition begun in Chapter 3 on the effects of various types of structural changes on oscillatory behavior. Thus far, the exposition has described primarily the contribution of negative loops to oscillation. Chapter 4 completes the exposition by considering the effects of positive loops on oscillatory behavior, and demonstrating the symmetry between the effects of positive and negative loops.

CONTENTS

4.1.	Positive Loops Changing Response Time	207
	An Example: Saving	207
	Another Example: A Driven Pendulum	213
	A Further Example: Price Maintained by Tradition	216
	A Principle	221
	Further Examples: Cold Hands and the Flu	226
4.2.	Adding a Minor Positive Loop with a Delay	229
	An Example: The Influence of Backlog on Ordering	229
	A Principle	238
	Symmetrical Effect of Positive and Negative Loops	240
4.3.	Use of Principles When They Don't Hold	248
	An Example: Time to Average Output TAO	248
	Another Example: Changing TCB and TAO	253

#### 4.1 POSITIVE LOOPS CHANGING RESPONSE TIME

An Example: Saving. Let us begin consideration of positive loops with an extremely simple first-order system. Figure 4-1 shows a level of savings  $S$  such as might be kept in a savings bank for retirement purposes. An exogenous deposit rate  $DR$  can also increase savings  $S$ . The withdrawal rate  $WR$  is proportional to savings  $S$ , so that the more savings  $S$  are available, the more they will be used. The withdrawal rate  $WR$  forms a negative loop with savings  $S$ . If the savings are in a bank that pays interest, the interest payment  $IP$  will be proportional to savings  $S$ , and will increase savings  $S$ , so that interest payments  $IP$  form a positive loop.

The savings  $S$  are determined in a level equation, with the deposit rate  $DR$  and interest payments  $IP$  increasing the level, and with the withdrawal rate  $WR$  decreasing savings  $S$ . The level is initialized at \$10,000, representing an amount a person might possess when he or she retires.

$S.K=S.J+(DT)(DR.JK+IP.JK-WR.JK)$	1,L
$S=SN$	1.1,N
$SN=1E4$	1.2,C
S	- SAVINGS (DOLLARS/YEAR)
DR	- DEPOSIT RATE (DOLLARS/YEAR)
IP	- INTEREST PAYMENTS (DOLLARS/YEAR)
WR	- WITHDRAWAL RATE (DOLLARS/YEAR)
SN	- SAVINGS INITIAL (DOLLARS)

The withdrawal rate  $WR$  is formulated as a constant function of the existing savings  $S$ , so that no withdrawals can be made when

there are no savings. The withdrawal fraction WF is set at 0.08, so that withdrawals will exhaust the level of savings S with a time constant of 12.5 years, without any deposits or interest payments.

WR.KI=S.K*WF	2,R
WF=.08	2.1,C
WR	- WITHDRAWAL RATE (DOLLARS/YEAR)
S	- SAVINGS (DOLLARS)
WF	- WITHDRAWAL FRACTION (FRACTION/YEAR)

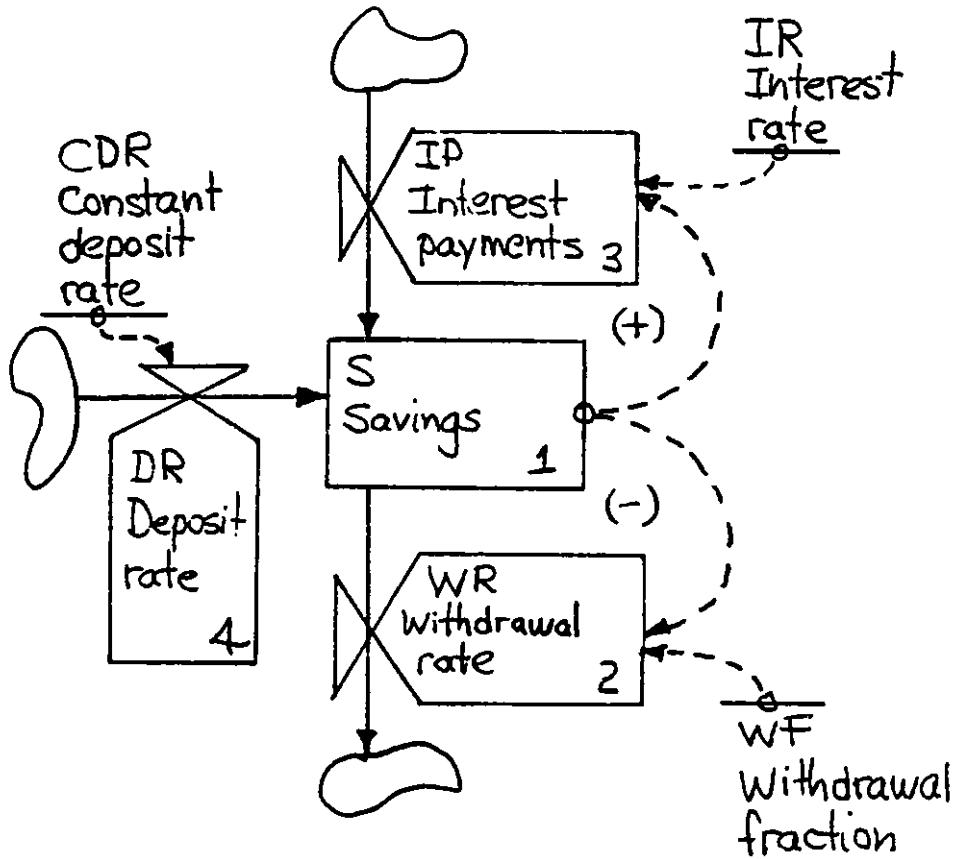


Figure 4-1.  
SAVING SYSTEM



Interest payments IP are proportional to the level of savings S. The interest rate IR is initially set at 0.0, so that the effect of the withdrawal loop alone on the behavior can be seen clearly. In a later simulation the interest rate IR will be set at 0.5.

IP.KL=S.K*IR		3, R
IR=0		3.1, C
IP	- INTEREST PAYMENTS (DOLLARS/YEAR)	
S	- SAVINGS (DOLLARS)	
IR	- INTEREST RATE (FRACTION/YEAR)	

The deposit rate DR is set at a single constant deposit rate CDR. This constant is initially set at zero to exhibit the consequences of the withdrawal loop alone on the behavior. In a later simulation, the constant deposit rate CDR will be set to \$250.00 per year.

DR.KL=CDR		4, R
CDR=0		4.1, C
DR	- DEPOSIT RATE (DOLLARS/YEAR)	
CDR	- CONSTANT DEPOSIT RATE (DOLLARS/YEAR)	

Figure 4-2 shows a comparative plot of the behavior of the savings system for four different sets of parameters. Curve A shows the behavior when withdrawals are made, but with no interest payments or deposits. Savings S declines with a time constant of  $(1/WF) =$

$(1/0.08)=12.5$  years. Curve B shows the behavior of the saving system with the positive loop involving interest payments IP activated by setting the interest rate IR equal to 0.05. The effective time constant of decline in savings S can be computed as follows:

$$\begin{aligned}\text{Effective time constant} &= S/((d/dt)S) \\ &= S/(IP-WR) \\ &= S/(S*IR-S*WF) \\ &= 1/(IR-WF) \\ &= 1/(0.05-0.08) \\ &= 1/(-0.03) \\ &= -33.3 \text{ years}\end{aligned}$$

The equations above imply that the interest rate IR and the withdrawal fraction WF can be set to give a very long time constant of behavior. For example, if the withdrawal fraction WF is reduced from 0.08 to 0.06, the effective time constant of the decline of the savings S becomes 100 years. This large value occurs because the effective time constant responds to the difference of the two system parameters, as well as their magnitude. Thus, a system involving both positive and negative loops can exhibit behavior with a time horizon much longer than might be inferred from any of the time-related parameters within the system structure.

Curve C shows the behavior of savings S when deposits are made at a rate of \$250 per year but with no interest payments or withdrawals. Curve C shows savings S increasingly linearly.

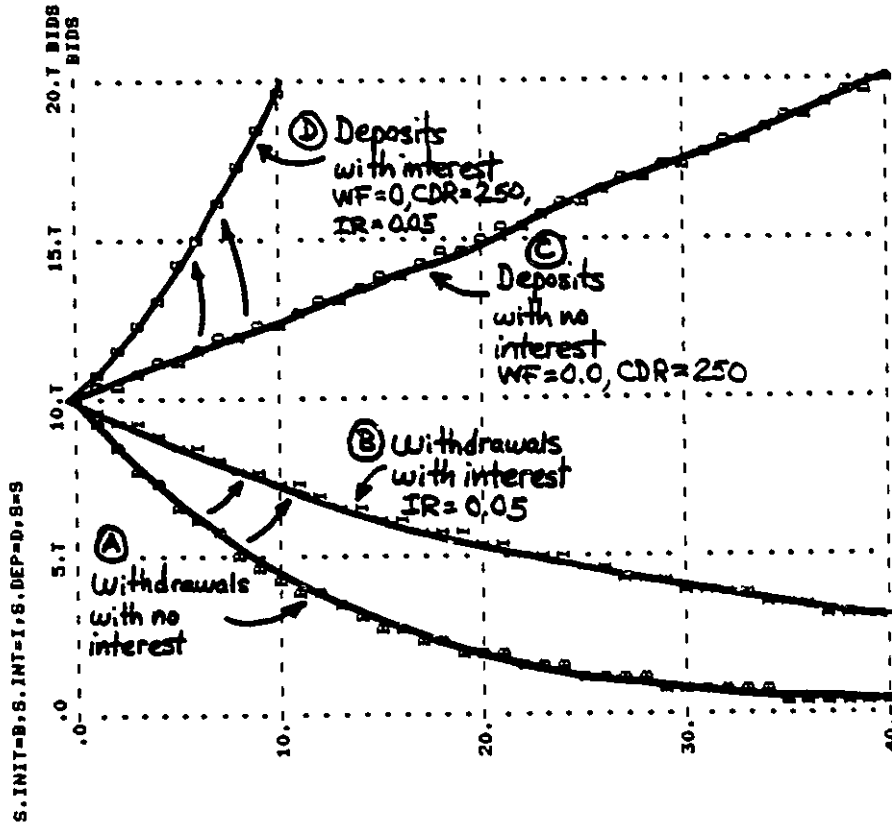


Figure 4-2.  
BEHAVIOR OF SAVINGS

Activating the positive loop involving interest payments IP causes savings S to rise very much faster: Curve A reaches \$20,000 in ten years, whereas Curve C reaches this value in 40 years.

To begin to develop concepts that go into a principle concerning the effect of positive loops on behavior, how might we characterize the impact of activating the positive loop involving interest payments IP? One way of describing these particular results is to say that the positive loop destabilized the behavior. However, the examples in Chapters 2 and 3 indicated that adding negative loops does not always stabilize a system's behavior. By analogy, we should be wary of similar cases where positive loops do not necessarily destabilize a system's behavior. Indeed, Section 4.2 gives an example of a positive loop that has a slightly stabilizing effect on the employment-backlog system's behavior. A more precise way of characterizing the behavior in Figure 4-2 is to say that adding the positive loop tended to augment the disturbance present in the system's levels, so that the levels move either more rapidly away from their steady-state values (if the levels were already diverging), or more slowly toward their steady-state values (if the levels were already converging). For convergent oscillatory systems, whether or not this slower movement toward steady-state values can produce either more damping or less damping depending on the characteristics of each specific system.

Another Example: A Driven Pendulum. Another example of activating a positive feedback loop occurs when the spring-mass system described in Section 2.1 is pushed with a force proportional to its velocity. Whichever direction the mass is going, the driving force pushes it faster in that direction. As the velocity of the mass diminishes, so does the driving force. Figure 4-3 shows a model of this driven mass, with the positive loop connecting the velocity  $V$ , the force from driving  $FD$ , and the acceleration  $A$ .

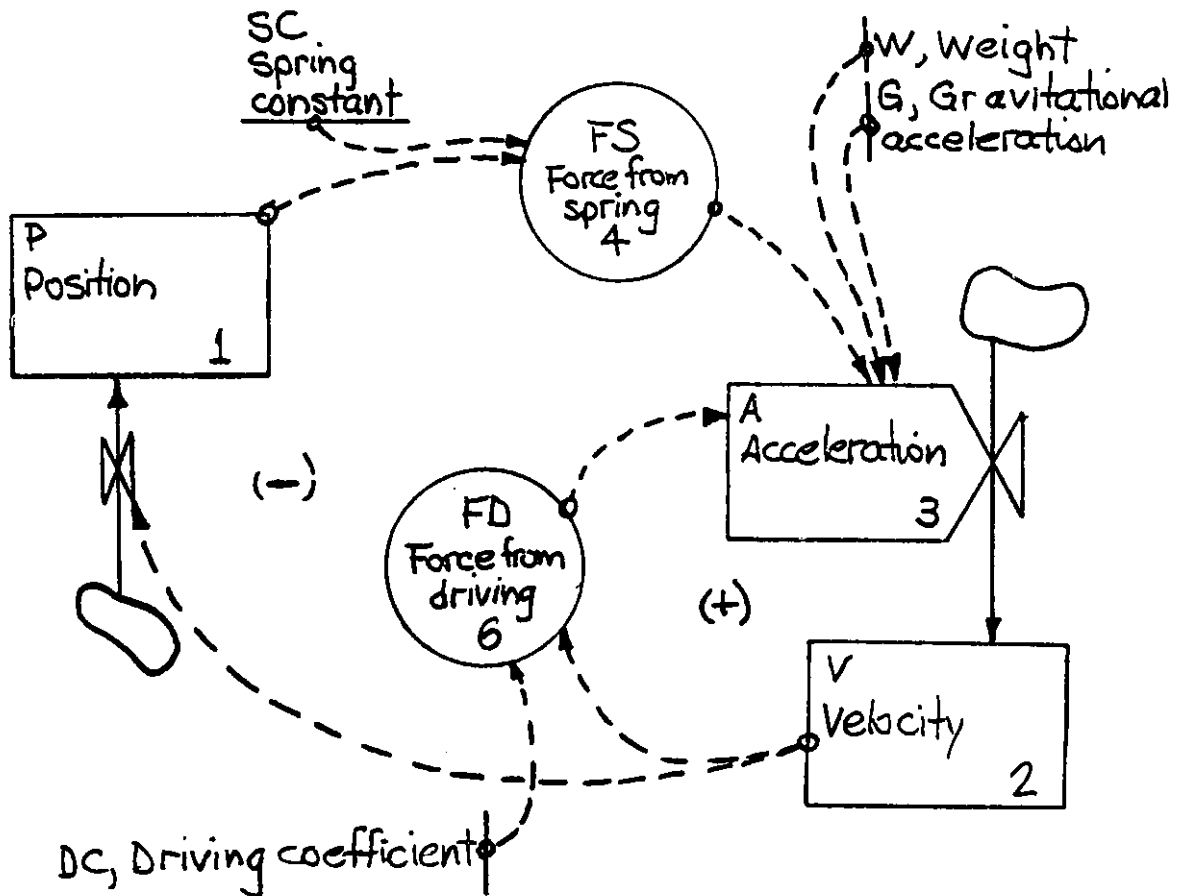


Figure 4-3.  
POSITIVE LOOP FORMED BY DRIVING THE SPRING-MASS SYSTEM

The force from driving FD is one of the forces that results in the acceleration A of the mass, which has already been defined in Chapter 2 as:

$$A.KL=(FS.K+FF.K+FD.K)/(W/G) \quad 3, R$$

$$W=160 \quad 3.1, C$$

$$G=32 \quad 3.2, C$$

A - ACCELERATION (FEET/SECOND/SECOND)  
 FS - FORCE FROM SPRING (POUNDS)  
 FF - FORCE FROM FRICTION (POUNDS)  
 FD - FORCE FROM DRIVING (POUNDS)  
 W - WEIGHT (POUNDS)  
 G - GRAVITATIONAL ACCELERATION (FEET/SECOND/SECOND)

The force from driving FD is assumed proportional to the velocity V. The constant of proportionality, the driving constant DC, is equal to 0.0 in the basic model, and is set to 3 pounds/foot/second to represent driving. Obviously, other policies of driving the system exogenously will have different effects on the behavior; anyone who has pushed a swing knows the swing can be pushed either to increase the swings, or to decrease the swings. The present policy, however, is both easily visualized, and corresponds to a simple structural addition, a minor positive loop.

$$FD.K=DC*V.K \quad 6,A$$

$$DC=0 \quad 6.1,C$$

FD - FORCE FROM DRIVING (POUNDS)  
 DC - DRIVING CONSTANT (POUNDS/FOOT/SECOND)  
 V - VELOCITY (FEET/SECOND)

Figure 4-4 shows the behavior that results from activating the positive loop. The oscillations diverge quite sharply. The causes for

the divergence can be seen by examining the values of the two levels at time 4.8, when the position P has risen to zero. Without the positive loop, a position P of zero would cause the acceleration A (exerted by the spring alone) to also equal zero. As position P continued to rise without the positive loop, the acceleration A would become negative and the velocity V would begin to decline. In contrast Figure 4-4 indicates that the velocity V continues to rise when position P equals zero: the rise is due to the force from driving FD, which continues to give the mass a positive acceleration A, thus causing the velocity V to rise further. Thus, the positive loop causes the velocity V to return to its steady-state value later than it would have without the influence of a positive loop. (This lengthened response time is the cause of the divergent oscillation; the longer the velocity V stays positive, the further position P will rise, so that the positive loop causes a larger disturbance over each cycle of the oscillation.)

The period of oscillation in Figure 4-4 can be inferred approximately from the time of the peak in position P around time 7.2. At that time, the position P has moved through one complete cycle, so the period is around 7.2 seconds, as opposed to 6.28 seconds for the original system in Figure 2-3.

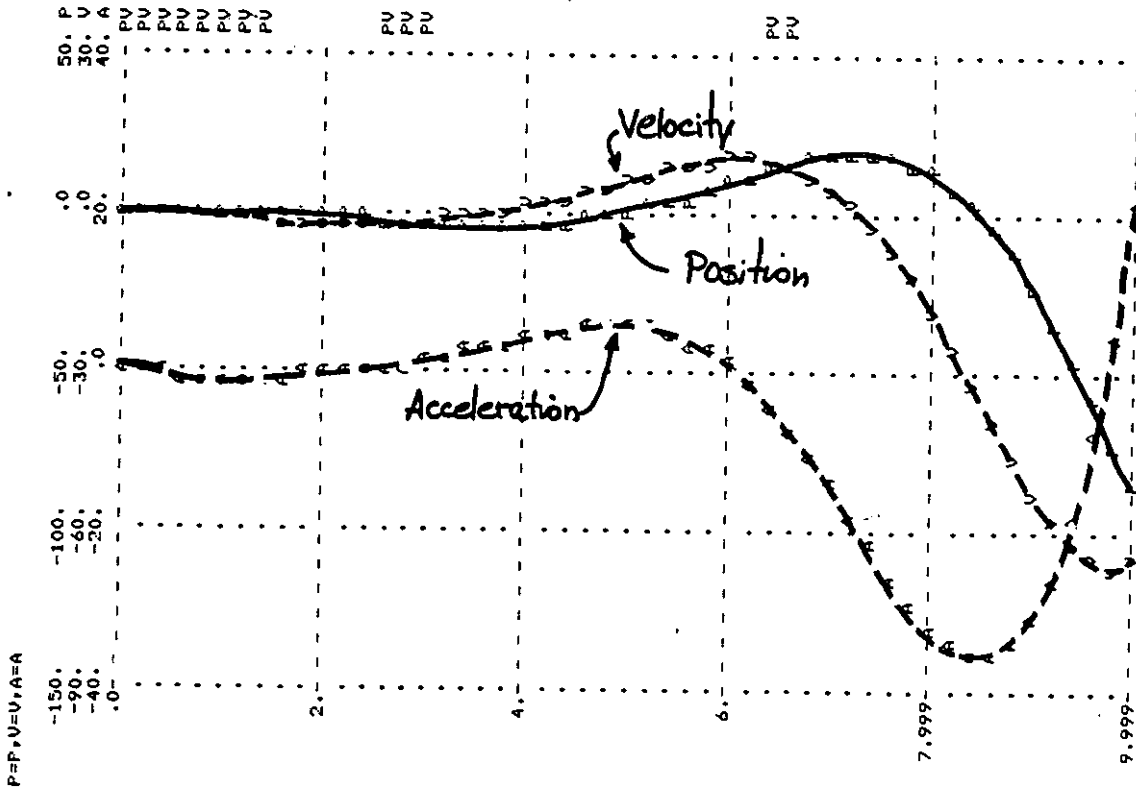


Figure 4-4.  
BEHAVIOR OF DRIVEN SPRING-MASS SYSTEM

A Further Example: Price Maintained by Tradition. Figure 4-5 shows a representative of a very large class of positive loops that occur in socio-economic systems. These loops occur in conjunction with a negative loop that attempts to control a system variable. In this case, a negative loop changes the price  $P$  that a company charges for its product to attempt to maintain a gross margin  $GM$  of revenues over cost. If the price  $P$  should fall, the gross margin  $GM$  also falls, which causes the change in price from margin  $CPM$  to make the change in price  $CP$  positive, which raises the price  $P$ . The goal of this negative loop



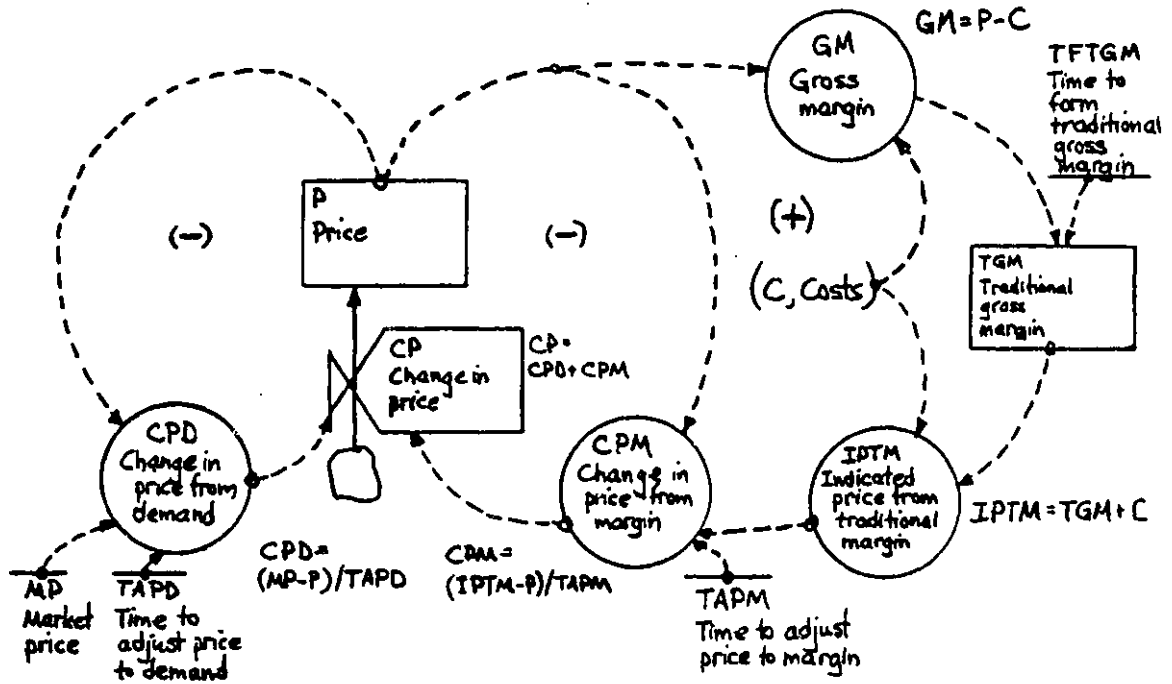


Figure 4-5.  
REGULATION OF PRICE BY GROSS MARGIN

is to maintain the gross margin GM equal to the traditional gross margin TGM.

What forms the firm's goal for profitability, the traditional gross margin TGM? In many institutions, the long-standing goals that regulate short-term behavior seem to be formed mostly on the basis of what has been the traditional performance. This formation of a traditional performance is represented by the long-term smoothing in the equation for the traditional gross margin TGM. The loop going

through TGM is positive; an increase in price P increases the gross margin GM, which will eventually increase the traditional gross margin TGM. A rise in this variable causes rises in the indicated price from traditional margin IPTM, the change in price from margin CPM, the change in price CP, and price P. The sign of the loop is positive; a change in one variable on the loop propagates around the loop to result in a change of the same sign. The gain, however, may not be nearly large enough to cause divergent behavior. In fact, the steady-state open-loop gain of such positive loops involving goals is usually 1.0 or less.\*1\*

Another negative loop attempts to regulate the price P with respect to competitive prices, represented here by the market place MP. If the price P becomes too high, the management of the firm perceives that the price may cause a loss in market share, so that the management will tend to lower prices, represented by the change in price from demand CPD. In this simple model, there is a separate time constant that regulates the adjustment of price in response to competition from other firms, a time to adjust price to demand TAPD.

---

\* 1.\*In brief, the gain is often 1.0 because the loop represents the formation of goals that allow the system to function at a scale of operation exactly equal to some measure of the current scale of operation. One example in the employment-backlog system is the positive loop closed by the perceived delay for backlog adjustment PDBA, which is discussed in Section 4.2. Another example in the employment-backlog system is the positive loop closed by the expected average orders EAO, which is discussed in Section 4.3. These positive loops are common enough to warrant a name of their own; I suggest calling positive loops with an open-loop, steady-state gain of 1.0 "a unity-gain loop."

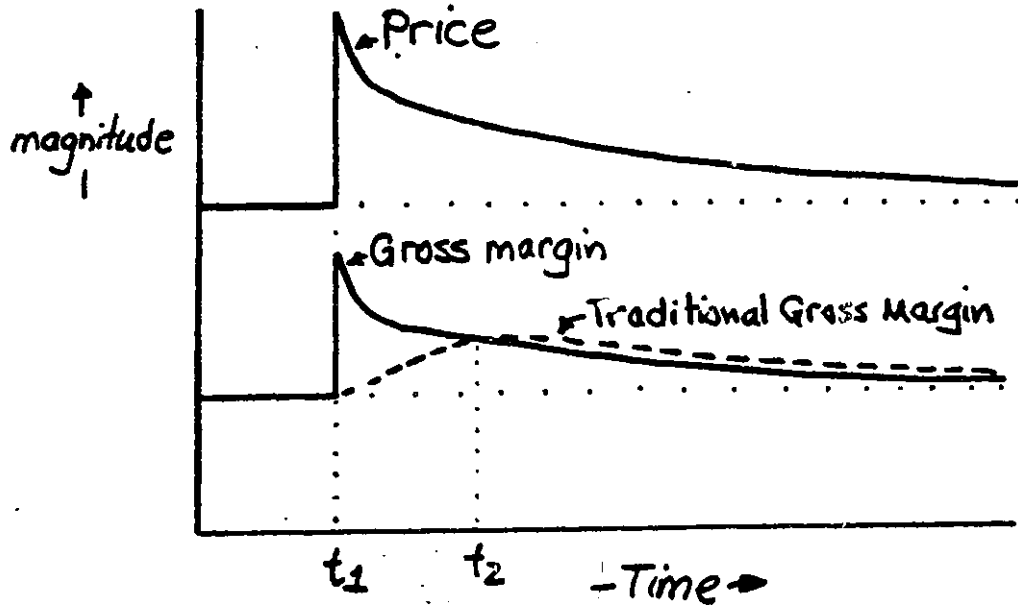


Figure 4-6.  
RECOVERY OF PRICE LEVELS

Figure 4-6 shows a sketch of the response of the price system shown in Figure 4-5 to a disturbance in the level of price  $P$ . A larger price  $P$  produces a larger gross margin  $GM$ , which eventually results in a larger traditional gross margin  $TGM$ . Between times  $t_1$  and  $t_2$ , the gross margin  $GM$  is larger than the traditional gross margin  $TGM$ , so that the influence of margins on the price  $P$  is to reduce  $P$ . After the time  $t_2$ , however, the traditional gross margin

TGM has risen, and the falling price P has reduced the gross margin GM so that the gross margin GM is actually below the traditional gross margin TGM. After time  $t_2$ , the influence of margins on price P is to raise P to attempt to maintain the now-traditional gross margin GM. Only the influence of the negative loop representing price competition from other firms continues to reduce price P. The positive loop involving traditional gross margin TGM tends to maintain and prolong the disturbance in price P. In fact if one calculates the time constant for the decline in price P for the equations shown on Figure 4-5, the time constant turns out to be much longer than any of the time constants in the system. If the time to adjust price to margin TAPM equals 2.0, the time to adjust price to demand TAPD equals 2.0, and the time to form traditional gross margin TFTGM equals 3.0 months, the time constant of decline is 72 months.\*2\* Thus in this situation, as with the simple saving system in Figure 4-1, the existence of a positive loop produces behavior with a time horizon much longer than might be expected from the values of the individual time-related parameters of the system.

---

\*2\*The system in Figure 4-5 is entirely linear, so that its behavior can be characterized by its eigenvalues, which can be calculated by any of a number of simple methods. The system has two eigenvalues, one of which corresponds to a time constant of 0.98 months, representing the time it takes for the two negative loops to adjust the price to some compromise between the market price MP and the indicated price from traditional margin IPTM. The 72-month time constant represents the time required for the traditional gross margin TGM to decline, given that the negative loop going through IPTM tends to maintain TGM.

A Principle. The preceding examples suggest a relatively simple principle:

Principle on adding a positive loop

Adding a positive loop around a level tends to augment whatever disturbance is present in a system's levels, so that the levels move more slowly toward their steady-state values. The system thus takes longer to complete a cycle in an oscillation, or to smoothly approach a steady-state value.\*3\*

The principle above can be justified heuristically by examining the consequences of raising the gain of a positive loop. Figure 4-7 illustrates. Curve A on the top part of the figure shows smoothly-convergent behavior, such as might occur in a first-order system or a highly-damped, higher-order system. The convergence is usually caused by a negative loop.\*4\* If a positive loop is activated with a gain so high that the original negative loop is unable to cause the

---

\*3\*In the case of smoothly divergent behavior, the statement in the principle that "the level moves more slowly toward its steady-state value" can be rephrased more accurately as "the level moves more rapidly away from its steady-state value." Thus, the principle implies that activating a positive loop increases the response time of a level returning to its steady-state value, or decreases the response time of a level diverging from its steady-state value. For example, the savings system in Figure 4-2 with deposits and no withdrawals produces the smoothly-divergent behavior shown in curve C. Activating the positive interest loop produces curve D, whose "response time of divergence" is clearly shorter than in curve C without the positive loop.

\*4\*Convergent behavior can also arise from purely positive loops; the fourth-order example of a purely positive loop in Section 2.1 has an eigenvalue of  $-1.0$  which corresponds to convergent behavior with a time constant of  $1.0$ . Such behavior modes, however, if they exist, are rarely the behavior mode of interest.

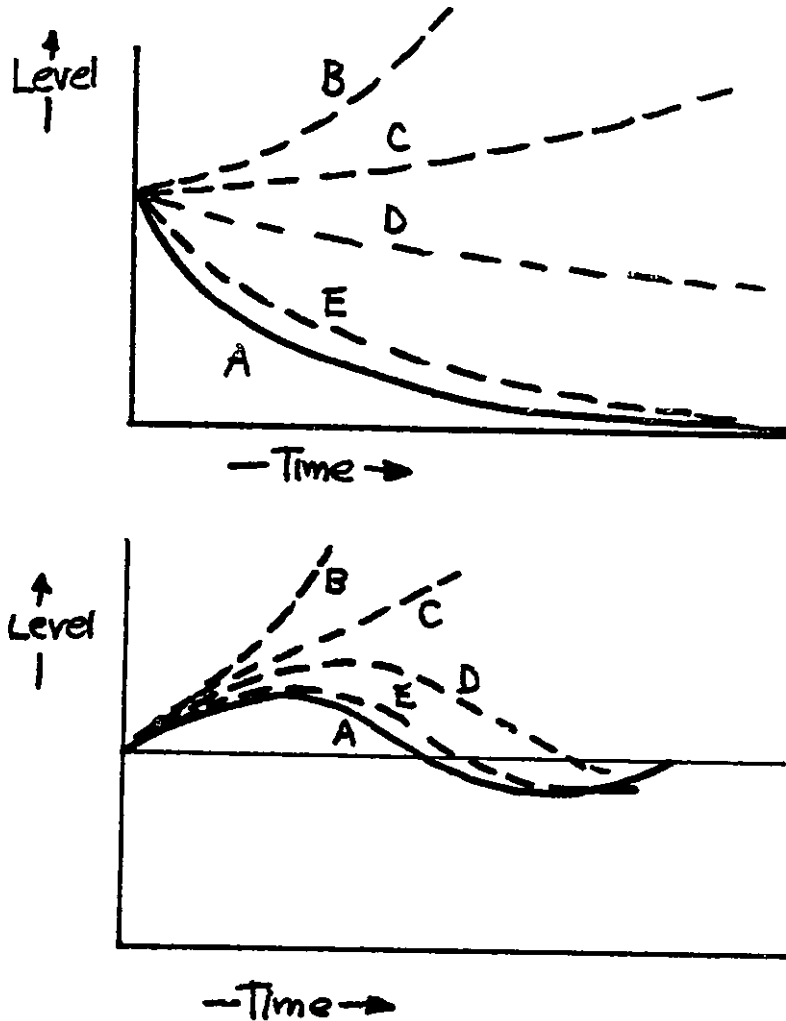


Figure 4-7.  
HEURISTIC FOR EFFECT OF POSITIVE LOOP ON RESPONSE TIME

system to converge to an equilibrium, then a curve such as B might result. Reducing the gain of the positive loop somewhat might result in the behavior shown by Curve C, where the state of the system still diverges, although less rapidly than in Curve B.

If the gain of the positive loop is reduced still further, the negative loop or loops that caused the original convergence would again be able to dominate the behavior and cause convergence. However, the convergence would be much slower than the original, as in Curve D. In comparison to the original convergent behavior, the time taken in Curve D to converge is very much longer than the time taken in Curve A. If the gain of the positive loop is reduced still further, a curve such as Curve E might result, where the trajectory is quite close to the original trajectory, but again, the response time is still longer than the original response time.

The lower part of Figure 4-7 shows a similar lengthening of response time for an oscillatory system.\*5\* Curves B and C represent the effect of activating a positive loop around a level in the originally-convergent system with a very high gain. As the gain is reduced (corresponding to changing from Curve B to C to D), the negative loops that cause the trajectory to return to the steady-state value eventually overcome the effect of the positive loop. Note that for a gain that just barely suffices to bring the trajectory back toward the steady-state value, the period of oscillation is very much longer than that of the original trajectory. (Indeed, just such a lengthening occurred in Figure 4-4, where a positive loop lengthened the period

-----

\*5\*Section 2.1 shows that convergent oscillations can also arise from positive loops, with or without minor negative loops. For the purpose of presenting a simple heuristic, let us ignore such cases, as it is only quite rarely that one would wish to know the effect of activating another positive loop around a level in such a system.

of oscillation of the spring-mass system.)\*6\* Figure 4-5 shows another trajectory, Curve E, that might result from further decreasing the gain of the added positive loop. Curve E is fairly similar to the original trajectory, Curve A, although, as usual, the period of oscillation is lengthened by the presence of an added positive loop.\*7\*

The principle on adding a positive loop can prove useful both during model refinement and policy design. During the process of model refinement it regularly occurs that a seemingly realistic model shows oscillatory behavior with an unrealistically short period. The principle implies that we should examine the real cause-and-effect relationships to identify positive loops, which includes many goal-formation processes, such as in the example of price declines in Figure 4-5.

-----

\*6\*Notice that one cannot infer from Curve D whether activating the positive loop has increased or decreased the damping. In fact, either can happen, depending on the specific system. In Figure 4-4, a positive loop around the velocity  $V$  in the spring-mass caused less damping than the original trajectory. In contrast, the following section gives an example where activating a positive loop around the backlog  $B$  of the employment-backlog system causes slightly more damping. In both cases, however, the period is lengthened, similar to the lengthening shown in the heuristic in Figure 4-5.

\*7\*The lengthening of the period of oscillation by a positive loop has an interesting implication for the study of business cycles. It is often assumed that business cycles occur independently of, and superimposed upon, exponential economic growth arising from the positive loop of capital reinvestment. The principle here implies that those positive growth loops may have the effect of lengthening the periods of business cycles, especially the twenty-year and fifty-year fluctuations, where the time constant of exponential growth is of the same order of magnitude as the periods and the time constants within the oscillatory portions of the system.



In searching for effective policies, it is sometimes desired to enhance the system's ability to track an exogenous input. For example, one might desire to enhance a firm's ability to follow upward trends in sales. (Whether or not this implies that the firm will also follow random fluctuations--which are present in every market--is another issue.) The principle suggests that a faster response time can result from eliminating or reducing the influence of positive loops.

There is at least one exception to the principle above. If the structure that is added to form a positive loop contains enough levels to be able to produce a large phase shift at a significant gain, the added structure could produce a large enough phase shift that disturbances propagating through the added structure arrive at a time when they cause the level that was in the original system to move more rapidly toward its steady-state value, instead of more slowly, as the principle states. To put it more concretely, adding a piece of structure that forms a positive loop and produces a 180-degree phase shift is functionally equivalent to adding a minor negative loop, whose sign change also produces, in effect, a 180-degree phase shift.\*8\* Adding minor negative loops often causes increased damping.

---

\*8\*None of the structures in this thesis added to produce a positive loop can produce a 180-degree phase shift, since they have no more than one level, and the largest phase shift obtainable from one level without a sign change is 90 degrees.

Further Examples: Cold Hands, and the Flu. Most of us are familiar with the uncomfortable period that follows coming indoors after an extended period of outdoor activity during winter. The cause of the discomfort is quite simple: despite the warm environment (and often brisk rubbing) hands and feet continue to feel cold for an appreciable amount of time after coming indoors. The duration of cold hands and cold feet arises at least in part from a positive loop produced by the body's reaction to prolonged low temperatures in the extremities. At normal indoor temperatures, the normal homeostatic response of blood flow through the hands and feet reduces blood flow when the temperature becomes too low. This tends to produce a stable temperature. The right side of Figure 4-8 shows the negative loop just described.

Figure 4-8 also shows a response that occurs when extremities are exposed to prolonged cold. At some point, the body attempts to conserve heat and reduce heat loss by restricting the flow of blood through the extremities. This restriction, of course, allows the hands and feet to become quite cold. Even if the person then goes indoors, the blood flow is restricted in the hands and feet, so that little

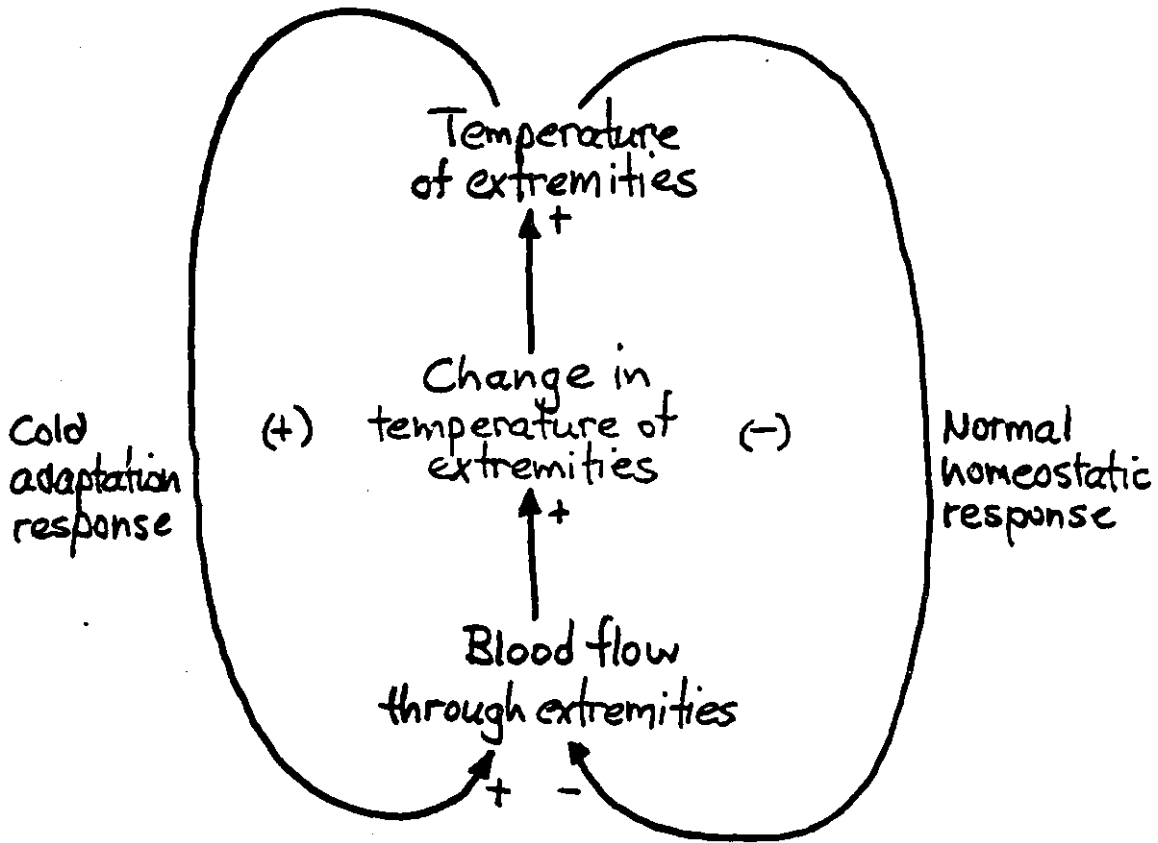


Figure 4-8.  
CONTROL OF TEMPERATURE OF HANDS AND FEET

heat is transmitted to the flesh of the hands and the feet, so they remain cold in relation to the indoor environment. This positive loop considerably prolongs the time it takes for the hands and feet to reach normal temperatures.

A similar and very much more uncomfortable prolongation of response time occurs in recovering from influenza. Normally, when one is hungry, one eats, which assuages the hunger. However, if one is sick

and hungry, the hunger also causes weakness and slows recovery from the disease. One of the unfortunate symptoms of the flu seems to be the inability to retain food. When one cannot retain food, one attempts to eat less often and one stays hungry. Both of these factors tend to maintain hunger and weakness. These effects form a positive loop, which is shown in Figure 4-9. The operation of this loop can cause the recovery from the flu to last several days. If one could only retain one good meal, the recovery could be very much quicker. In terms of the principle, this positive loop tends to augment the disturbance to the system (hunger, weakness, and disease), which slows the movement back toward the steady-state value of normal health.

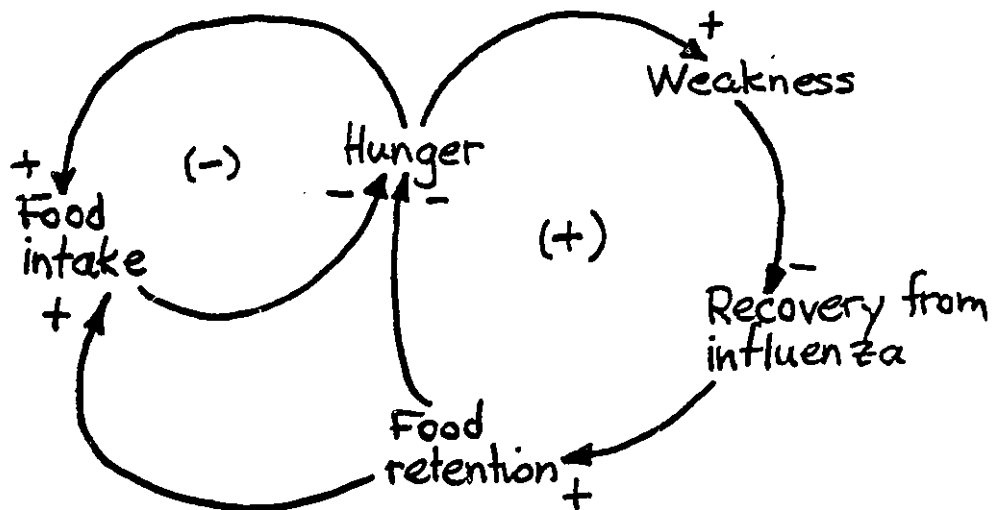


Figure 4-9  
RECOVERY FROM INFLUENZA

#### 4.2 ADDING A MINOR POSITIVE LOOP WITH A DELAY

An Example: The Influence of Backlog on Ordering. Section 2.2

described a minor negative loop around the level of backlog B in the employment-backlog system. That loop represented an attempt on the part of firms ordering units to maintain an appropriate level of orders in the backlog B. Figure 4-10 shows this minor negative loop.

The appropriate level of orders that the minor negative loop in Figure 4-10 attempts to maintain is the desired backlog for ordering DBO. DBO is formulated to maintain a desired coverage of the desired order rate DOR. The coverage is equal to the aggregate orderers' perception of how long an order in the backlog B takes to be filled. This perception is quantified as the perceived delay for backlog adjustment PDBA.

$$\text{DBO.K} = \text{PDBA.K} * \text{DOR.K} \quad 12, A$$

DBO	- DESIRED BACKLOG FOR ORDERING (UNITS)
PDBA	- PERCEIVED DELAY FOR BACKLOG ADJUSTMENT (YEARS)
DOR	- DESIRED ORDER RATE (UNITS/YEAR)

The formulation above ensures that, if the orderers' perceptions are accurate with respect to the delivery delay DD, maintaining the backlog B equal to the desired backlog for ordering DBO causes the actual rate at which orders are filled (the rate of output OUT) to equal the desired order rate DOR. This can be shown starting with the definition of the delivery delay DD:

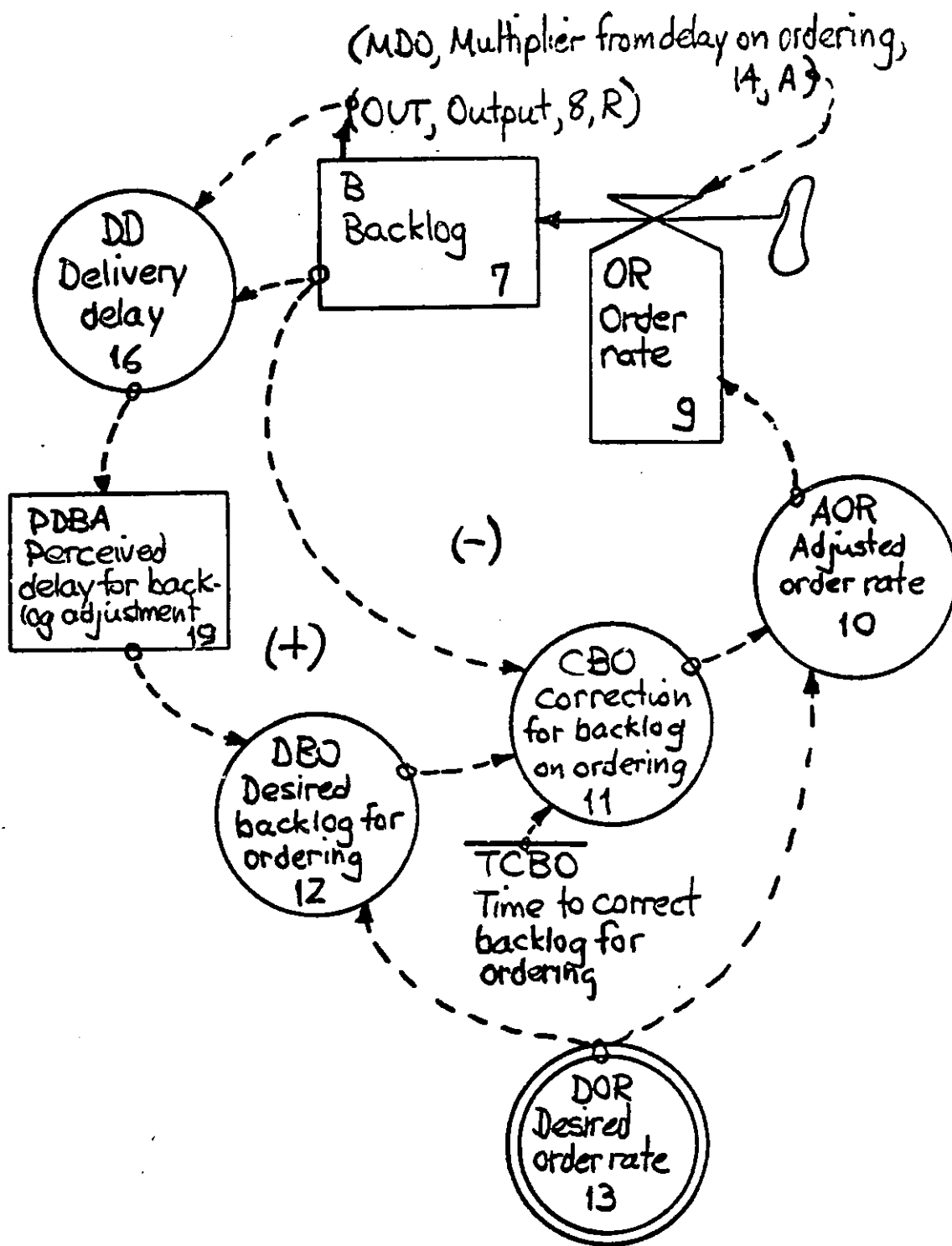


Figure 4-10.  
INFLUENCE OF BACKLOG ON ORDERING

$$DD = B/OUT$$

Assuming that the perceived delay for backlog adjustment PDBA equals DD,

$$PDBA = B/OUT$$

With the equation above, one can calculate the backlog B required to maintain deliveries of the units (which equals the rate of output OUT) at the desired order rate DOR by substituting DOR for OUT and solving for B:<sup>\*9\*</sup>

$$B = PDBA * DOR$$

The backlog B computed above specifies the goal at which the actual backlog B must be held in order to maintain the desired delivery rate, so the equation immediately above specifies the computation for the desired backlog for ordering DBO:

$$DBO = PDBA * DOR$$

The DYNAMO equation for DBO embodies this result.

The perceived delay for backlog adjustment PDBA represents a delayed and averaged perception of the current delivery delay DD. In previous simulations, the value of PDBA was held constant, despite variations in DD, by setting the time to perceive delay for backlog adjustment TPDBA at a very large number. In the simulation that follows, TPDBA has been reduced to 1.5 years. (A more realistic value might be somewhat shorter; the value of 1.5 years is chosen to

---

<sup>\*9\*</sup>The astute reader will notice that the calculations here arise from the viewpoint of an individual orderer, who does not perceive that his or her orders in any way effect the magnitude of the delivery delay DD. The individual orderers accept DD as given, and, under that assumption, attempt to control the number of units they each have on order.

show more clearly the effects of activating a minor positive loop with delay.)

PDBA.K=PDBA.J+(DT/TPDBA)(DD.J-PDBA.J)	19,L
PDBA=DBC	19.1,N
PTDBA=1E11	19.2,C

PDBA	- PERCEIVED DELAY FOR BACKLOG ADJUSTMENT (YEARS)
TPDBA	- TIME TO PERCEIVE DELAY FOR BACKLOG ADJUSTMENT (YEARS)
DD	- DELIVERY DELAY (YEARS)
DBC	- DESIRED BACKLOG COVERAGE (YEARS)

The loop closed by the perceived delay for backlog adjustment PDBA is positive; an increase in PDBA increases the desired backlog for ordering DBO, which increases the order rate OR, the backlog B, and the delivery delay DD. Thus an increase in PDBA propagates around the loop to cause a further increase in its value. We can calculate the steady-state open-loop gain of the positive loop in Figure 4-10, beginning with the delivery delay DD. If we break the loop at DD, and assume that some input value of DD exists (denoted by DDIN), we can calculate the delivery delay which would occur as a result of that DDIN propagating around the loop, the DDOUT. Taking the partial derivative of DDOUT with respect to DDIN yields the steady-state, open-loop gain of the positive loop. We know that in steady-state,

$$PDBA = DDIN$$

which implies that

$$\begin{aligned} DBO &= PDBA * DOR \\ &= DDIN * DOR \end{aligned}$$



How does the value of DBO influence the steady-state value of the next level in the loop, the backlog B? For B to be in equilibrium,

$$\begin{aligned} \text{OUT} &= \text{OR} \\ &= \text{DOR} + \text{CBO} \\ &= \text{DOR} + (\text{DBO}-\text{B})/\text{TCBO} \\ &= \text{DOR} + (\text{DDIN}*\text{DOR}-\text{B})/\text{TCBO} \end{aligned}$$

Solving for B,

$$\text{B} = \text{DDIN}*\text{DOR} - (\text{OUT}-\text{DOR})(\text{TCBO})$$

Finally, DDOUT comes from the regular equation for DD,

$$\begin{aligned} \text{DDOUT} &= \text{B}/\text{OUT} \\ &= (\text{DDIN})(\text{DOR}/\text{OUT}) + (\text{OUT}-\text{DOR})(\text{TCBO})/\text{OUT} \end{aligned}$$

If the employment E can be assumed to maintain the rate of output OUT at the desired order rate DOR (i.e.  $\text{OUT} = \text{DOR}$ ), then

$$\text{DDOUT} = \text{DDIN}$$

and so the steady-state, open-loop gain is

$$\begin{aligned} d(\text{DDOUT})/d(\text{DDIN}) &= d(\text{DDIN})/d(\text{DDIN}) \\ &= 1.0 \end{aligned}$$

Figure 4-11 shows the behavior of the employment-backlog system with the positive loop shown in Figure 4-10 activated. The time to correct backlog for ordering TCBO is set at 0.3 to allow the correction for backlog on ordering CBO to function. Also, the time to perceive delay for backlog adjustment TPDBA is set at 1.5 years. Since the correction for backlog on ordering CBO is active, the simulation in Figure 4-11 should not be compared with the base

simulation. Instead, the effect of activating the positive loop is most easily discerned by comparing Figure 4-11 with a simulation in which CBO is active, but the perceived delay for backlog adjustment PDBA is kept constant. Figure 4-12 shows that simulation; it is identical to Figure 2-25.

Consistent with the principle in the last section, activating the positive loop going through the perceived delay for backlog adjustment PDBA has significantly increased the response time of the system. Between years 1.0 and 4.5, the desired backlog for ordering DBO is below the steady-state value for backlog B, so that the effect of the positive loop during this time is to reduce the order rate OR and prolong the approach of the backlog B to its steady-state value. The backlog B smoothly approaches its steady-state value with a time constant of about one year, coming very close to its steady-state value at year 4.5. In contrast, the backlog B in the system without the active positive loop in Figure 4-12 already reaches its steady-state value at year 1.3, overshoots, and is very close to equilibrium at year 2.2.

Activating the positive loop has also eliminated the slight oscillation present in Figure 4-12; in Figure 4-11, the backlog B rises smoothly to its steady-state value, without overshoot. The lack of overshoot is due to the delay in perceiving the delivery delay DD. This can be seen by comparing the curves for delivery delay DD and the perceived delay for backlog adjustment PDBA. At time 2.0, the





actual delivery delay DD has peaked and is starting to decline. The perceived delay for backlog adjustment PDBA, however, is still well below the actual DD and still rising. Since PDBA is below its steady-state value, the desired backlog for ordering DBO (which is proportional to PDBA) is likewise below its steady-state value. Thus, even when the backlog B is approaching its steady-state value, the desired backlog for ordering DBO is still lower than both the backlog B and its own steady-state value. DBO thus depresses the order rate OR, and slows the rise of the backlog B. Were there little or no delay in perceiving the delivery delay DD, the restraining effect of the desired backlog for ordering DBO would disappear as backlog B approached its steady-state value.\*10\*

---

\*10\*If the time to perceive delay for backlog adjustment TPDBA is set at 0.5 years instead of the 1.5 years used in Figure 4-11, the resulting behavior shows much less damping than either Figure 4-11 or Figure 4-12. The desired backlog for ordering DBO responds much more quickly to the delivery delay DD, so that as the backlog B rises toward its steady-state value, the retarding effect of the desired backlog for ordering DBO diminishes as well. The backlog B is not restrained in this case from overshooting its steady-state value.

A Principle. The behavior change caused by adding a minor positive loop with a delay in Figure 4-11 is in many ways symmetric with the effect of adding a minor negative loop with a delay, whose effect is shown in Figure 3-20. The negative loop decreased the response time, and the positive loop increased the response time. The minor negative loop with a delay destabilized the system, and the minor positive loop with a delay stabilizes the system.\*11\* In both cases, adding the loop formed another pathway through which disturbances in a level could propagate back to the level. When the disturbance propagated around the added negative loop, it moved the level more rapidly to and past its steady-state value. When the additional disturbance propagated around the positive loop, it retarded the movement of the level toward its steady-state value, which resulted in a longer period and more stable oscillation. Just as a detailed examination of phase and gain in the backlog subsystem was performed with the minor negative loop with a delay, so could the same type of examination be performed with opposite signs for the minor positive loop. However, such an effort seems redundant, given the strong symmetry between the two systems. Instead, we can capitalize on this symmetry by stating a principle symmetrical to that stated in Section 3.3:

-----

\*11\*As in the principle on adding a minor negative loop with a delay, "destabilize" here denotes having more oscillations occur before essentially reaching equilibrium. Also "destabilize" is not necessarily the same as "increase the time constant of the exponential envelope," since the time constant of exponential approach in Figure 4-11 is longer than the time constant of the envelope of the "less stable" oscillations in Figure 4-12.

Principle on adding a minor positive loop with a delay

If a minor positive loop with a delay is added around a level already on an oscillatory loop, the added loop forms another pathway through which disturbances in the level can propagate back to the level. When the additional disturbance returns to the level, it retards the movement of the level toward its steady-state value, which results in a longer period and more stable oscillations.

As usual, it is not clear that a general rule can be stated for how short the delay in the added loop can be and still have a stabilizing, rather than destabilizing effect. In this particular case, a delay of less than 1.0 years causes the positive loop to destabilize the oscillations. Section 6.3 discusses a method for possibly arriving at general conclusions about the effect of the delay time on oscillations.

The principle here derives from a single example of a minor positive loop with a delay, just as the principle in Section 3.3 derives from a single example of a minor negative loop with a delay. However, because of the symmetry between the two examples, they can be considered as both supporting and exemplifying both principles.

As is true for most of the principles, the principle on adding a minor positive loop with a delay finds uses both in model refinement and policy design. During model refinement, a modeler may be uneasy about adding an obviously realistic positive loop to a model, even if the addition would probably make the period of oscillation longer and more realistic, given the folklore that associates positive loops with instability and divergent oscillations. The principle indicates that the folklore is not generally true for minor positive

loops with a delay. If the modeler wishes to stabilize a model's oscillations and at the same time lengthen the response time, the modeler can examine the real cause-and-effect relationships for positive loops with delays in them. In policy design, the principle indicates that if a minor positive loop with a delay is present in the system, the system's oscillations may very well be stabilized by increasing the gain of the positive loop, if at the same time the delay time is chosen to give the appropriate phase lag.

Symmetrical Effect of Positive and Negative Loops. Given the dependence of the arguments above on appeal to symmetry, this matter of symmetry between positive and negative loops warrants further examination. We can begin to put positive and negative loops into a uniform framework by regarding the polarity of a loop as the result of its parameters rather than regarding positive and negative loops as intrinsically different from one another. Figure 4-13 illustrates for a simple first-order loop. The rate  $R$  is equal to the level  $L$  times  $k$ , so that if  $k$  is less than zero, the loop is a negative loop, and if  $k$  is larger than zero the loop is a positive loop. Activating another negative loop around the level  $L$  corresponds to reducing the value of  $k$ . Activating another positive loop around the level  $L$  corresponds to increasing the value of  $k$ .

Consider now the behavior changes that result from activating positive or negative loops. These activations can be represented by changes in the system's parameters. (Indeed, the actual mechanism by



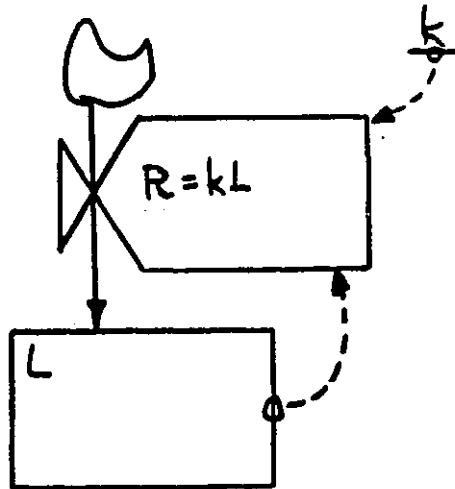


Figure 4-13.  
SIMPLE FIRST-ORDER SYSTEM

which loops are activated in the employment-backlog model is changing the parameters of the system.) We can consider the behavior of any system as a function of its parameters, even though for complex systems we may not be able to write down an explicit function that gives a behavioral characteristic (such as damping, period, or time constant of approach) as a function of its parameters. For example the behavior of the first-order system in Figure 4-13 can be described as a function of the parameter  $k$  as follows: if  $k$  is larger than zero, the undriven system will diverge exponentially from its initial condition with a time constant equal to  $1/k$ . If  $k$  is less than zero, the undriven system will converge exponentially toward zero with a time constant equal to  $1/k$ . If  $k$  is equal to zero, the level  $L$  will remain at its initial value. Similar but more complex relationships can be obtained

between the parameters and the behavior of the spring-mass system, or the employment-backlog system.

We can consider (in the abstract) any measurable behavioral characteristic as a function of one of the system's parameters as shown in Figure 4-14. This figure shows the relationship of a specific behavioral characteristic of a system (such as damping time constant, or period) as a function of a parameter value of the system such as  $k$  in Figure 4-13, which controls whether a loop is inactive, positive, or negative. The curve is smooth and locally monotonic. There are no sharp discontinuities, or reversals of slope at any maxima or minima. In practice, the behavioral characteristics of the majority of socio-economic systems show this well-behaved dependence on parameter values. For example in the employment-backlog system, Figure 3-2 indicates that reducing the time to correct backlog TCB decreases the damping and the period of oscillation. If damping and period are well-behaved functions of TCB, one would expect them to increase when TCB does; Figure 4-16 shows that this is indeed the case.

Given that one can expect the behavioral characteristic in Figure 4-14 to be a smooth, locally-monotonic function of the parameter, let us return to positive and negative loops. Parameter values to the left of the vertical bar in Figure 4-14 correspond to the activation of a negative loop. Parameter values to the right of the vertical bar correspond to the activation of a positive loop. If the behavioral

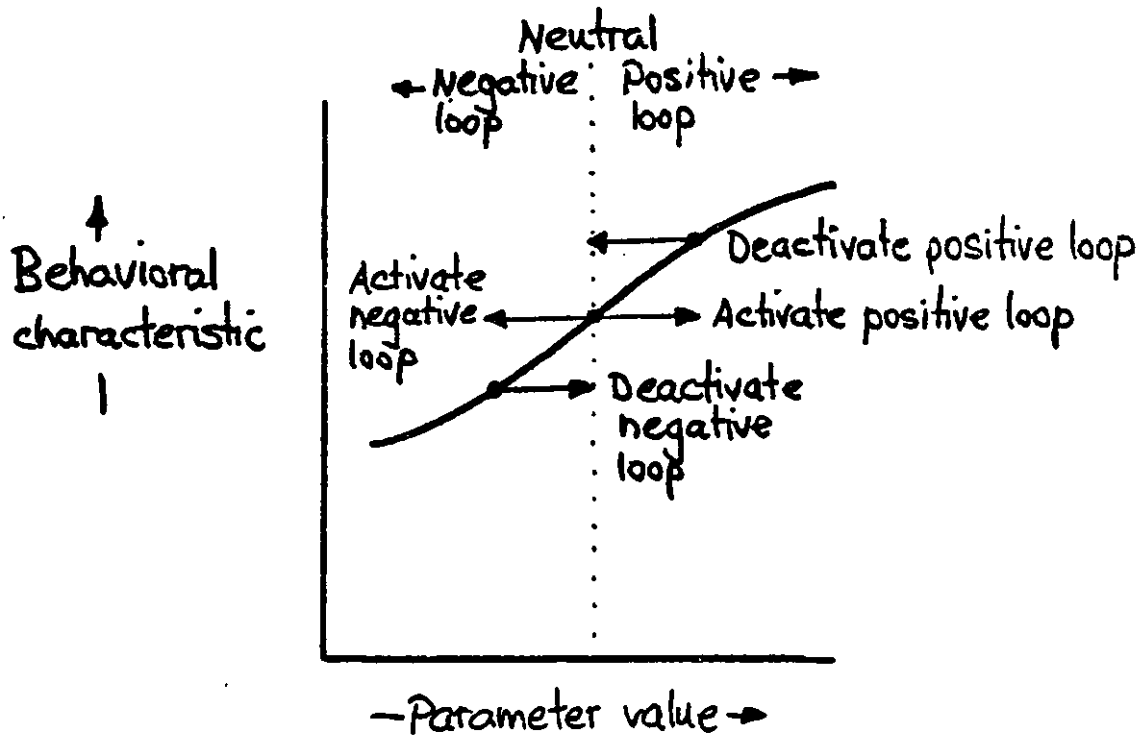


Figure 4-14.  
A BEHAVIORAL CHARACTERISTIC AS  
A FUNCTION OF A PARAMETER

characteristic is a smooth function of the parameter value (as it is on Figure 4-14), deactivating a negative loop has the same effect on the behavior as activating a positive loop; both of these actions correspond to increasing the parameter value. Similarly, deactivating a positive loop has the same effect on the behavioral characteristic as activating a negative loop; both of these actions correspond to decreasing the

parameter value. Therefore, for systems whose behavioral characteristics are (at least locally) monotonic functions of the parameter values, the effects of activating positive and negative loops will be symmetrical.

The symmetrical effect of activating positive and negative loops is borne out in examples given thus far.\*12\* Figure 3-20 shows that in the employment-backlog system, a minor negative loop with a delay around the backlog B decreases the damping and the period of oscillation; Figure 4-11 indicates that a minor positive loop with a delay around the backlog B increases the damping and the period of oscillation. It is common for minor negative loops to increase damping; it is common for minor positive loops to decrease damping.

To complete the symmetrical treatment of positive and negative loops in this thesis, the principle on positive loops changing response times can be restated for negative loops as follows:

Adding a negative loop around a level tends to diminish whatever disturbance is present in the level, so that the level moves more rapidly toward its steady-state value. The system thus takes less time to complete a cycle in an oscillation, or to smoothly approach a steady-state value.

---

\*12\*The single most flagrant exception is, as usual, the undamped second-order system exemplified by the spring-mass system. Although adding minor negative and positive loops changes the damping in opposite directions, adding a minor negative loop and adding a minor positive loop both increase the period (See Figures 2-3, 2-19 and 4-4).

With one exception, the principle above is borne out by all of the examples of adding a negative loop given in this thesis.\*13\* There is a class of exceptions, however, just as with the principle on positive loops: If the structure that is added to form a negative loop contains enough levels to be able to produce a large phase shift at a significant gain, the added structure could produce a large enough phase shift that disturbances propagating through the added structure arrive at a time when they cause the level that was in the original system to move more slowly toward its steady-state value, and not more rapidly as the principle states.

It should be noted that behavioral characteristics are not always smooth functions of parameter values, especially for appreciable changes in parameter values. Usually, a discontinuity arises because a positive loop finally gains dominance over a negative loop. For example, the pollution sector in the World Dynamics

-----

\*13\*The one exception is the undamped spring-mass system, in which adding a minor negative loop around the velocity  $V$  increased the period, as shown in Figure 2-19. This system is problematic in a number of areas. Either positive or negative minor loops increase its period of oscillation. It is the most common exception to the general rule that increasing a coupling time constant increases the damping of an oscillation in a loop. It is also the most common system for which the normal rules for choosing  $DT$  break down.

model shows a sharp discontinuity in its response to additional pollution.\*14\* During the smooth growth phase of the World Dynamics model, increases in the rate of pollution generation slightly increase the steady-state value of pollution. However, a point is reached where the presence of pollution inhibits the processing and degradation of pollutants, so that the outflow rate of the level of pollution is dramatically decreased. At that point, further increases in the rate of pollution generation result in runaway levels of pollution; there is no more quasi-steady-state. Another example is Shaffer's model of the criminal justice system.\*15\* Small perturbations in the crime rate do not significantly effect the average crime rate until prisons approach their carrying capacity of prisoners. At that point, prisoners begin to be released, and the deterrent effect of imprisonment begins to be eroded, and increases in the crime rate lead to further erosion of deterrence, which leads to still more crime. Thus, when prisons are adequate, the crime rate is stable, and when prisons are at or near capacity, the crime rate increases exponentially. Numerous other examples could of course be found, but the lesson is clear: there are on occasion

---

\*14\*Forrester, J.W., World Dynamics (Cambridge, Mass.: Wright-Allen Press, 1971). The dynamics of a similar pollution sector are analyzed in Goodman, M.R., Study Notes in System Dynamics (Cambridge, Mass.: Wright-Allen Press, 1974).

\*15\*Shaffer, W.A., "Court Management and the Massachusetts Criminal Justice System", Unpublished Ph.D. Dissertation (Cambridge, Mass.: MIT, 1976).

sharp discontinuities in the system's response to parameter changes or perturbations. While these exceptions to continuity form a class of systems too large to ignore, it nonetheless remains true that for the most part behavioral characteristics are continuous functions of parameter values, and there is a symmetry between the activation of positive and negative loops.

#### 4.3. USE OF PRINCIPLES WHEN THEY DON'T HOLD

The principles enunciated in this thesis seem to be true in a large enough fraction of cases that they can be enunciated as working rules. However, there are always exceptions. In such cases, the principles are in fact still useful, as principles lead one to form expectations which when broken facilitate one in detecting something out of the ordinary. To make this point clear, this section describes two instances that occurred in the process of writing Sections 4.1 and 4.2.

An Example: Time to Average Output TAO. In the basic employment-backlog model in Figure 2-12, the desired output DO<sub>UT</sub> and the desired backlog DB are both based upon a measure of expected level of activity of the firm, the expected average output EAO. This variable has been held constant thus far. EAO is formulated as a first-order exponential smoothing of the actual rate of output OUT, as shown in Figure 4-15. The expected average output EAO closes a positive loop. An increase in EAO increases the desired output DO<sub>UT</sub>, the desired employment DE, the employment E and thus, the rate of output OUT. Increases in OUT cause a further increase in EAO. The open-loop, steady-state gain of this loop is 1.0; it is a unity-gain loop.



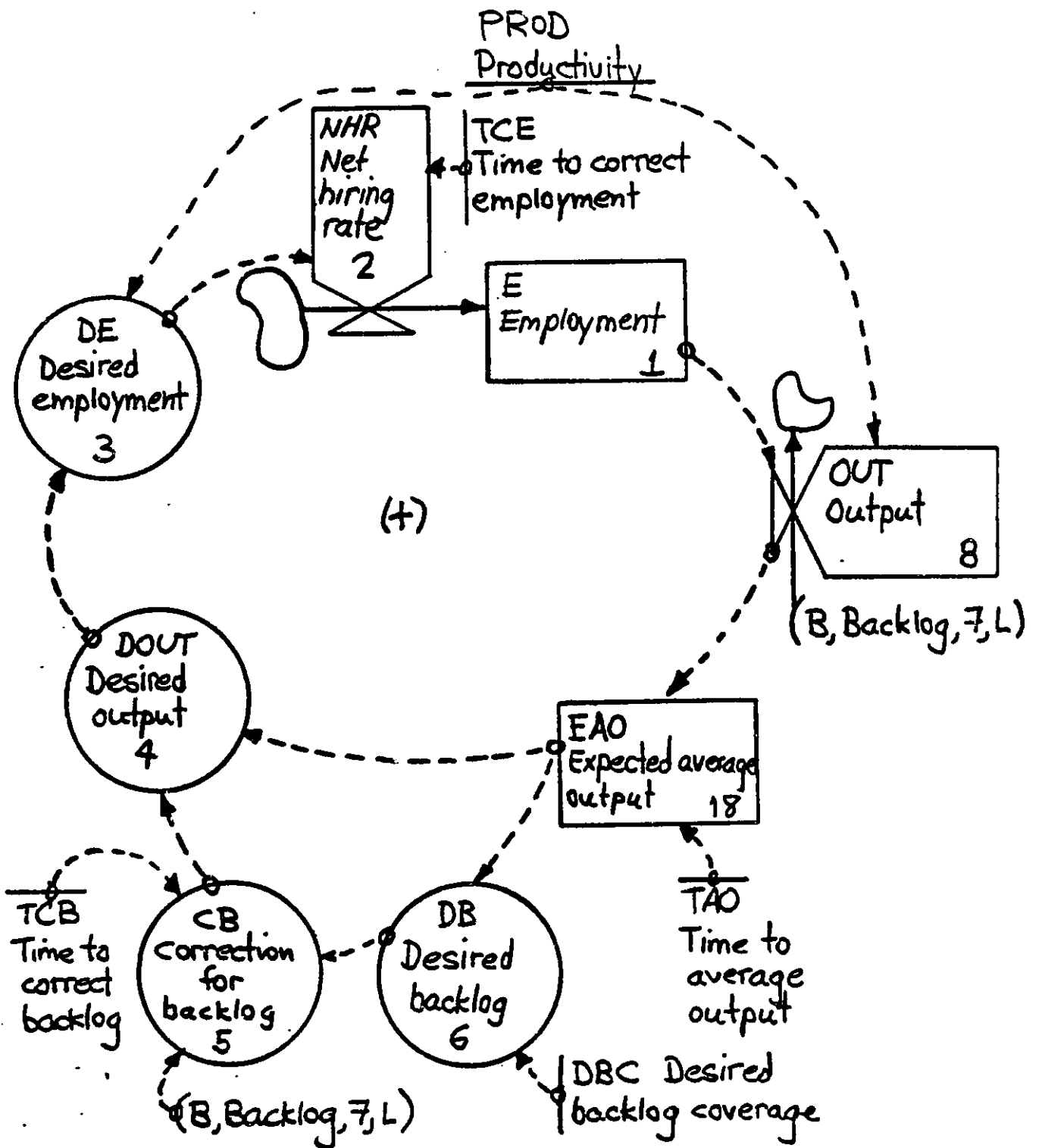


Figure 4-15.  
LOOP INVOLVING EXPECTED AVERAGE ORDERS EAO

Before giving an equation for the expected average output EAO, it will be helpful to review the two equations in which EAO is used. The first is desired output DOUT, in which the expected average output EAO is used as the baseline around which the correction for backlog CB varies the desired output DOUT.

$$\text{DOUT.K} = \text{EAO.K} + \text{CB.K} \quad 4, A$$

DOUT	- DESIRED OUTPUT (UNITS/YEAR)
EAO	- EXPECTED AVERAGE OUTPUT (UNITS/YEAR)
CB	- CORRECTION FOR BACKLOG (UNITS/YEAR)

The other use of the expected average output EAO is to determine the desired backlog DB of the firm. DB is computed in terms of the coverage (the desired backlog coverage DBC) of some expected rate of activity, the expected average output EAO.

$$\text{DB.K} = \text{EAO.K} * \text{DBC} \quad 6, A$$
$$\text{DBC} = .5$$

DB	- DESIRED BACKLOG (UNITS)
EAO	- EXPECTED AVERAGE OUTPUT (UNITS/YEAR)
DBC	- DESIRED BACKLOG COVERAGE (YEARS)

The expected average output EAO is defined as a simple first-order smoothing equation. The time to average output TAO has been set in the basic model at a very large value, so that EAO remains constant. In simulations where the expected average output EAO is to vary, TAO is set at 0.5 years.

EAO.K=EAO.J+(DT/TAO)(OUT.JK-EAO.J )	19,L
EAO=DOR	19.1,N
TAO=1E11	19.2,C
EAO	- EXPECTED AVERAGE OUTPUT (UNITS/YEAR)
TAO	- TIME TO AVERAGE OUTPUT (YEARS)
OUT	- OUTPUT (UNITS/YEAR)
DOR	- DESIRED ORDER RATE (UNITS/YEAR)

Since allowing the expected average output EAO to vary would activate a positive loop, I expected that activating EAO by setting the time to average output TAO at 0.5 years would increase the period of oscillation. In fact, no value of TAO produced any perceptible effect on the oscillation. After obtaining those results, another look at the flow diagram indicated the causes: increases in expected average output EAO also increased the desired backlog DB, which cause decreases in the correction for backlog CB. In other words, when the level of activity is expected to be large, the present backlog becomes less adequate, so that the management would want to slow down production to maintain the backlog. This effect on the desired backlog DB is of the opposite sign from the effect on the desired output DOUT. Thus, EAO also closes a negative loop that goes through the desired backlog DB. Which is more influential, the positive loop through EAO or the negative loop through EAO?

We can compute the net effect of variations in the expected average output EAO on the desired output DOUT, substituting the model equations,

$$\begin{aligned} \text{DOUT} &= \text{EAO} + \text{CB} \\ &= \text{EAO} + (\text{B} - \text{DB}) / \text{TCB} \\ &= \text{EAO} + \text{B} / \text{TCB} - \text{EAO} * \text{DBC} / \text{TCB} \\ &= \text{B} / \text{TCB} + \text{EAO} (1 - \text{DBC} / \text{TCB}) \\ &= \text{B} / 0.5 + \text{EAO} (1 - 0.5 / 0.5) \\ &= 2\text{B} \end{aligned}$$

In other words, because the time to correct backlog TCB was chosen exactly equal to the desired backlog coverage DBC, the expected average output EAO has exactly no effect on the desired output DOUT.

Without the principle on the effect of positive loops on response time, I probably would have done one simulation with the time to average output TAO set at 0.5 years, seen very little effect, concluded that the system was insensitive to TAO, and gone on to some other topic. As it happened, the principle led me to expect the period to lengthen, and when it did not, I knew something out of the ordinary had happened. In this case, it was the existence of an additional negative loop going through expected average output EAO, and the happenstance choice of parameters that caused EAO to have no effect on the system behavior. Given the time pressures that normally accompany research, this peculiarity would almost certainly have gone unnoticed without the principle.

Another Example: Changing TCB and TAO. After having deduced that for the original parameters of the employment-backlog system, the expected average output EAO would have no effect on system behavior, I increased the time to correct backlog TCB in order to further experiment with the effects of the positive loop involving EAO. Figure 4-16 shows the behavior that results from increasing the time to correct backlog TCB from 0.5 to 1.5. This value renders the loop closed by EAO positive, even when the effect of desired backlog DB is considered.\*16\* The effect of increasing TCB in Figure 4-16 (in which EAO is not activated) is as expected: the damping and the response time are increased. (This effect is symmetrical with the results shown in Figure 3-2, in which decreasing TCB decreased the damping and response time.) Although one can not be sure from the figure whether or not the system is in fact oscillating, the response time can be judged by the time the backlog B requires to return to its steady-state value of 660 units, about 4.8 years.

The principle enunciated in Section 4.1 indicates that activating a positive loop (such as that going through the expected average output EAO) increases the response time of the system. Figure 4-17 shows the effect of setting the time to average output TAO

---

\*16\*The equations above indicate that if TCB exceeds the desired backlog coverage DBC, the partial derivative of DOUT with respect to EAO is positive, and the loop closed by EAO is positive. If, on the other hand, DBC exceeds TCB, the partial derivative is negative, as is the loop.



(the smoothing time constant for EAO) at 0.5. As in the previous simulation, the time to correct backlog TCB has been increased from 0.5 to 1.5. Activation of the positive loop has caused the resumption of perceptible oscillations. Such a decrease in damping, however, is not inconsistent with the indications of the principle. However, the time taken for the backlog B to reach its steady-state value of 660 units for the first time has decreased substantially, from 4.8 years in the previous simulation to about 3.5 years. This shortening of the response time is exactly the opposite of what the principle in Section 4.1 would predict.

On further considering the behavior shown in Figure 4-17, the possibility arose that the level of expected average output EAO might be involved in a new behavior mode, as opposed to merely modifying the employment-backlog oscillations. This hypothesis is borne out in the simulation shown in Figure 4-18. In this simulation, the time to correct backlog TCB is again set at 1.5, the time to average output TAO is again set at 0.5, and, in addition, the time to correct employment TCE is set at 0.1. In other words, the time to correct employment TCE has been set so small that for all practical purposes, the employment E equals the desired employment DE, and the original employment-backlog system is an effectively-first-order system. In Figure 3-1, a similar reduction of TCE totally eliminated oscillation in the employment-backlog system.

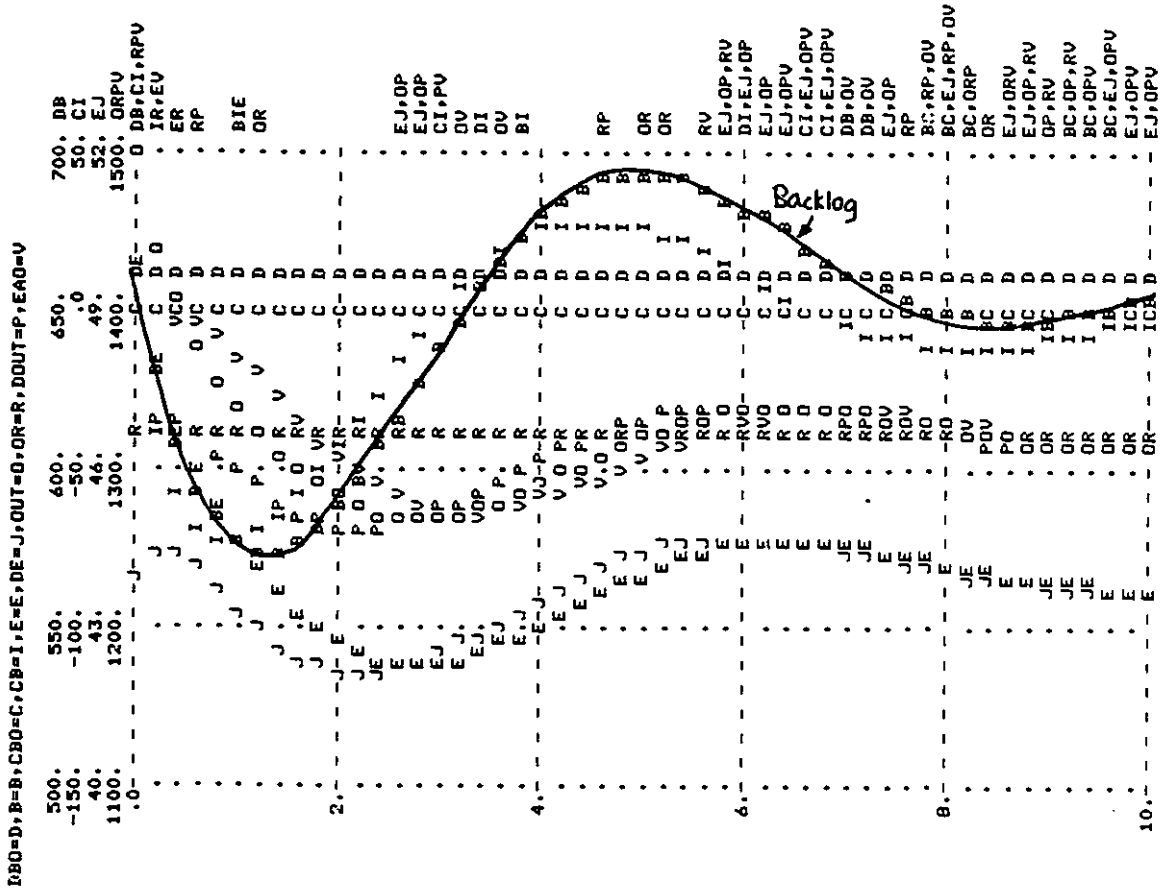


Figure 4-17.  
 SETTING TAO AT 0.5 AND INCREASING TCB FROM 0.5 TO 1.5



However, in Figure 4-18, oscillation still results, from propagating the disturbance around the loop containing the backlog B and the expected average output EAO. The dominant oscillatory mode of behavior has changed.\*17\*

In Figure 4-18, the initially-high employment E causes the rate of output OUT to exceed the order rate OR, which causes decline in backlog B, desired output DOUT, desired employment DE, employment E, the rate of output OUT, and the expected average output EAO. At time 1.2, EAO has declined back to its steady-state value. However, at time 1.2 EAO is declining because the current rate of output OUT is below its steady-state value (which occurs because the backlog B, the desired output DOUT, and hence the rapidly-responding employment E are below their respective equilibrium values). In short, at time 1.2, EAO is at its steady-state value and B is disturbed below its steady-state value.

When the rate of output OUT is below its steady-state equality with the order rate OR, the backlog B rises. At time 1.5, B has

-----

\*17\*One of the areas described in Chapter 6 as open for further investigation is the characterization of situations in which parametric or structural changes evoke a new behavior mode instead of merely modifying the original behavior mode. One form of evoking a new behavior mode occurs in Figures 4-17 and 4-18, where a new dominant oscillatory loop seems to have shifted because the original employment-backlog loop is highly damped and is no longer an efficient propagator of disturbances; the backlog-expected average output loop is more efficient at propagating disturbances, so the oscillations that emerge result from propagation through the latter loop.

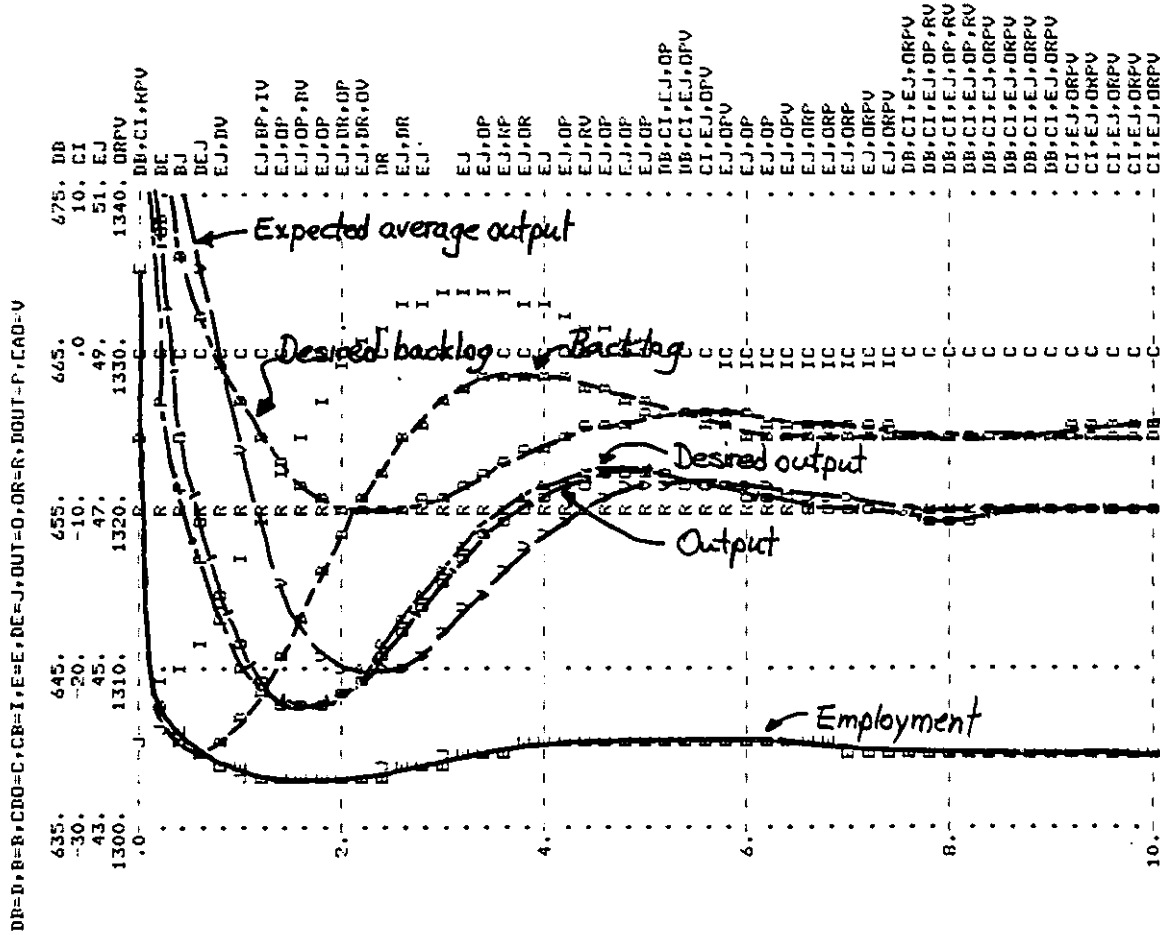


Figure 4-18. PREVIOUS CHANGES PLUS DECREASING TCE FROM 0.5 TO 0.1

risen to its steady-state value. The expected average output EAO, however, has dropped away from its steady-state value, following the (lower-than-steady-state) rate of output OUT. Thus, in going from time 1.2 to time 1.6, a disturbance in B has propagated into a disturbance in EAO. Oscillations continue after time 1.6 with the disturbance propagating back and forth from B to EAO.

Figure 4-18, then, implies that the dominant oscillatory loop in Figure 4-17 is no longer the original employment-backlog loop, but a new loop.

As in the previous example, a principle set up expectations about system behavior. When the actual behavior violated those expectations, the need for additional analysis was made very clear. Again as in the previous example, it seems unlikely that this property of the system could have been discovered without the principle.

CHAPTER FIVE:

SEPARATE SYSTEMS WITH A COMMON RANDOM INPUT

Sections 3.3 and 4.2 began to address some aspects of the general phenomenon of entrainment, where fluctuations in two or more oscillatory systems are drawn into a regular relationship with one another. In those sections, entrainment occurred as a result of both systems reacting to, or attempting to control, the same level. Chapter 5 analyzes another means by which entrainment occurs: two similar but completely separate systems can be subjected to the same random influences, so that to some extent, the behavior of the similar systems is also similar. One instance of such a configuration occurs when a real system and a model of that real system are both subjected to a known exogenous disturbance. Under what circumstances should the model behavior be expected to track real behavior, and what can be inferred from differences between the two? The principle developed in Chapter 5 begins to answer such questions.

CONTENTS

5.1.	A Heuristic Development of the Principle	262
	The Configuration and Its Significance	262
	An Example in the Employment-Backlog System	265
	A Heuristic Analysis	270
	A Principle	272
	Other Examples: Menstruation and Squirrels	274
5.2.	A Mathematical Development of the Principle	276
	Mathematical Approach	276
	General Linear Case	277
	Equations for a Simple System	282
	Numerical Results for the Simple System	288
	A Restatement of the Principle	307
	Observations on Mathematical Techniques	309

## 5.1 A HEURISTIC DEVELOPMENT OF THE PRINCIPLE

The Configuration and its Significance. Figure 5-1 shows the configuration being considered in this chapter. The similar, but not necessarily identical, dynamic systems are driven by the same random input. The systems have no causal connection with one another. In broad terms, the question this chapter seeks to answer is "Given that both systems oscillate, what is the relationship between the two oscillatory outputs?"

The question above occurs in several different contexts. This material was originally developed in the context of analyzing the question of why the multitudinous sectors of the economy seem to move roughly in unison in an aggregate three- to seven-year business cycle.\*1\* One means by which the various productive sectors of the economy might be driven to behave in unison is by being subjected, each and every one of the sectors, to the same exogenous influences. Changes in tax laws, changes in expectations about economic growth, wars, presidential elections, weather, and rapid oil price changes all have an impact on most productive sectors. The question quite naturally arises, "Can such random exogenous influences account for the unanimity with which productive sectors exhibit business cycles?"

-----

\*1\*Section 1.3 describes this aspect of the System Dynamics National Model development in more detail.

Another context within which questions arise about the relationships of the outputs of two separate systems with a common input occurs is in testing models against time-series data. For example, system one could be the actual system that determines hog prices in the U.S., which is influenced by the price of corn. (Corn is the principal food for hogs, and buying or growing corn is the principal cost of raising hogs.) System two could be a model of the U.S. hog price system, again subject to the known fluctuations in corn prices. In one study of U.S. hog prices, the outputs of the

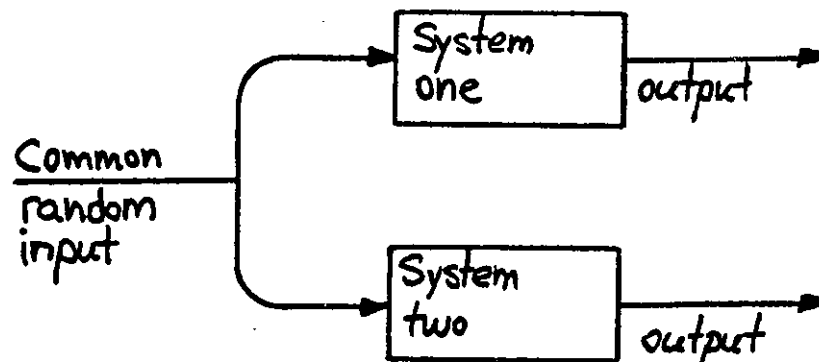


Figure 5-1.  
SEPARATE SIMILAR SYSTEMS WITH A COMMON RANDOM INPUT

two systems (real hog prices and simulated hog prices) seemed to accord very well with one another; Figure 5-2 illustrates.\*2\*

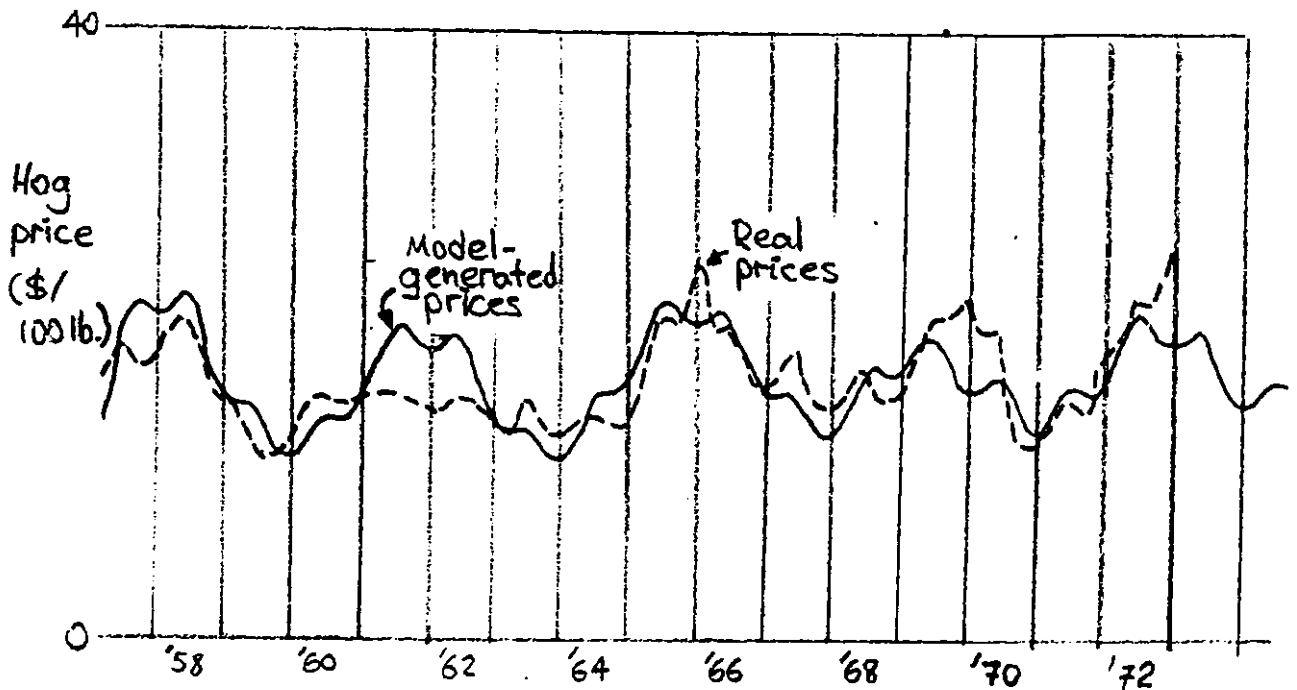


Figure 5-2.  
SIMULATED VERSUS REAL HOG PRICES  
(After Naill, et.al.)

---

\*2\*R.F.Naill, N.J. Mass, M.K. Simpson, and J. Randers, "Dynamic Modeling as a Tool for Managerial Planning: A Case Study of the U.S. Hog Industry," Proceedings of the Summer Computer Simulation Conference, Montreal, 1973.



Under what conditions should we expect a valid model of a system such as the hog price system to closely track the behavior of the real system? And, if the tracking is not exact, how would we interpret, say, a consistent phase difference between the simulated and the real outputs? To answer such questions, let us begin with a concrete example.

An Example in the Employment-Backlog System. Figure 5-3 shows the step response of two similar but not identical employment-backlog systems.\*3\* One system has the same parameters that have been used all along, which resulted in the initial condition response shown in Figure 2-14. The other system is identical to the first, except that the time to correct backlog TCB is raised from 0.5 to 1.0. Increasing TCB from 0.5 to 1.0 apparently does not much effect the damping time constant, although the period of oscillation is substantially increased. The original system, the faster system, requires about 1.8 years to traverse from the first trough to the first peak. The system with the larger TCB, the slower system, requires 3.0 years, implying that increas-

---

\*3\*Technically, the two trajectories were not generated simultaneously. Two successive simulations of the same employment-backlog system with two sets of parameters generated the two trajectories. Both trajectories, however, result from exactly the same sequence of pseudo-random inputs.

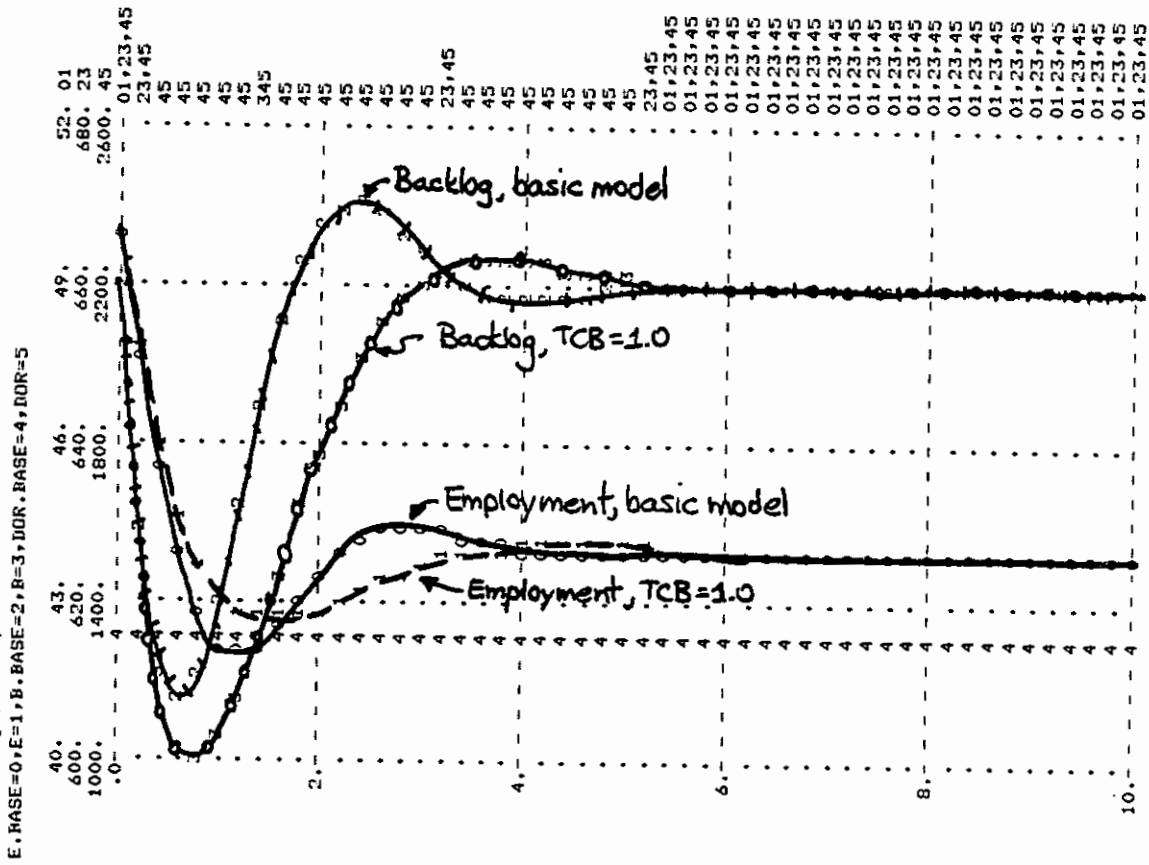


Figure 5-3.  
STEP RESPONSE OF TWO EMPLOYMENT-BACKLOG SYSTEMS

ing TCB has increased the natural period from about 3.6 years to about 6.0 years.\*4\*

Figure 5-4 shows the response of the two employment-backlog systems subjected to exactly the same random variations in the desired order rate DOR, which (because no extra loops are activated) equals the actual order rate OR. The random input to the desired order rate DOR is generated by a macro that produces first-order autocorrelated noise.\*5\* The standard deviation is set to zero in the basic model. In Figure 5-3, the standard deviation of noise SDN is set at 10 percent of the mean; the exponential smoothing that generates the first-order autocorrelation has a time constant (the time to correlate noise TCN) of 0.7 years.

DOR.K=1320*PKNS(1,SDN,TCN)	13, A
SDN=0	13.1, C
TCN=.7	13.2, C
DOR	- DESIRED ORDER RATE (UNITS/YEAR)
PKNS	- COLORED (PINK) NOISE FUNCTION
SDN	- STANDARD DEVIATION OF NOISE (DIMENSIONLESS)
TCN	- TIME TO CORRELATE NOISE (YEARS)

-----

\*4\*The measurement of natural period here is only approximate; in general, the maxima and minima are not exactly one-half of the natural period apart, due to transients and the "tilt" of the peaks due to damping. A correct method of assessing the natural period is finding the frequency at which a sinusoidal input produces the maximum steady-state response. However, the approximation used here suffices to indicate that the natural periods indeed differ significantly from one another, irrespective of their exact values.

\*5\*The macro definition for PKNS appears in the program listing in Appendix F. A discussion of the mathematical properties of this macro are given in K.R. Britting "DYNAMO Noise Generator," System Dynamics Group Working Paper D-2261, Alfred P. Sloan School of Management (Cambridge, Mass.: MIT, 1975).

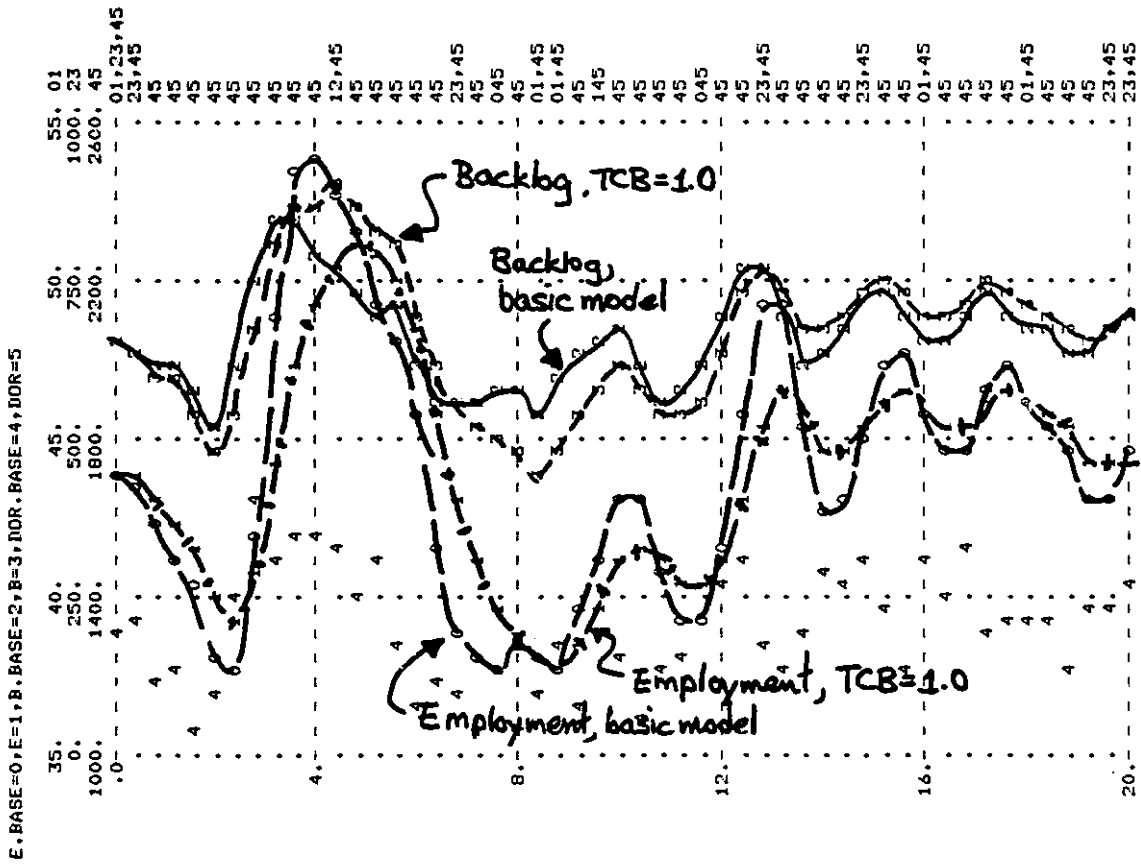


Figure 5-4.  
NOISE RESPONSE OF TWO EMPLOYMENT-BACKLOG SYSTEMS

Both systems start out in equilibrium, but fairly rapidly begin to show somewhat regular fluctuation. The trajectories of the two systems for both employment E and backlog B seem to move basically together. The major difference between the behavior of the two systems is that the basic model (the more rapidly-responding model) often both rises and falls first, prior to increases and declines in the variables of the slower system. For example, the employment E of the basic model peaks at year 4.0 and begins to decline. The slower system reaches its maximum at year 4.4 or 0.4 years later. The employment E of both systems thereafter declines, with the employment of the faster system lower at any given point during the decline than the employment E of the slower system. Later around year 8, the faster system again begins to reverse its direction (albeit somewhat erratically), while the slower-responding system reaches its extreme later. Similar differences occur later, as well as between the backlogs B of the two systems.

A Heuristic Analysis. The basic behavior of Figure 5-4 seems to be that the two systems fluctuate in parallel, with their differences in natural period manifesting themselves as phase differences. Term this behavior probabalistic entrainment, to distinguish it from the types of entrainment that can result from causal connections between the two systems. One way of rendering probabalistic entrainment more plausible is to consider the noise input as a series of impulse disturbances. Figure 5-3 shows the responses of the two systems to one such impulse disturbance (in this case, a disturbance in the employment E; disturbances to the backlog B have the same properties).

The largest deviation from steady-state values occurs between time 0.0 and time 1.8, during which both the slower system and the faster system are below their steady-state values. They are basically in phase, although the faster system does reach its trough before the slower system does. Because the natural periods of the two systems are different, the oscillations slowly become out of phase with one another. However, because damping occurs as the two systems drift out of phase, the magnitude of the out-of-phase oscillation between years 3.2 and 4.5 is quite small. Thus, if the two systems were subjected to a series of such impulses, one would expect the resulting behavior to be dominated by the larger in-phase portion of the response, rather than the relatively much smaller out-of-phase portion of the response.\*6\*

---

\*6\*The heuristic description above is the verbal equivalent of performing a convolution integral, which computes exactly the contribution of each portion of the impulse response to the behavior.

Furthermore, because the faster system moves toward and achieves its maxima and minima more rapidly than the slower system, one would expect some of the discrepancies in the trajectories of the two systems to manifest themselves as phase differences, exactly as happened in Figure 5-4.

A moderate amount of damping seems essential in producing the behavior described above. In the extreme case of no damping, where both systems maintain the same amplitude of oscillation forever, each at its own natural period, the trajectories of the two systems will move in and out of phase with one another, in a "beat" pattern. Even in cases of very light damping, the step responses of the two systems would show nearly as much magnitude in the out-of-phase portion of the response as the in-phase portion of the response, so that the sum of such responses still would not show a strong regular relationship. In contrast, consider a very heavily-damped pair of systems, where both responses will follow the profile of the common input -- sluggish systems will go wherever the input moves them, and recover through exponential decay. The resulting behavior will reflect the time-shape of the input much more than the internal dynamics of the two systems. The trajectories will therefore be very similar to one another.

Can one say anything more quantitative about the circumstances under which the two systems will exhibit probabilistic entrainment? From the description above, we can infer that two systems will show such entrainment when their oscillations damp out before the oscillations

become out of phase with one another. The time it takes for two inphase sinusoids to become out-of-phase is inversely proportional to the difference between the two natural periods.\*7\* Thus, systems whose natural periods are close to one another will take a relatively long time to get out-of-phase. Even very lightly-damped systems can show probabilistic entrainment if their periods are close enough together that the time it takes to get out of phase is still longer than the damping time constants. The more damped the systems are, the further apart their natural periods can be and still show probabilistic entrainment.

A Principle. The essence of the heuristics above can be embodied in the following principle (the principle is called preliminary because a more precise formulation of the principle is given in Section 5.2):

-----

\*7\*Consider the number of half-cycles  $N$  it takes for a faster system (with period  $P_1$ ) to get out-of-phase with a slower system (with period  $P_2$ ), if they start in phase. That event takes time  $N(P_1/2)$  to occur. Over the same time period, the slower system goes through  $N-1$  half cycles, which takes time  $(N-1)(P_2/2)$ . Since the time taken to get out of phase is the same for both systems, we can get  $N(P_1/2) = (N-1)(P_2/2)$ , which implies that  $N = P_2/(P_1-P_2)$ . The time taken to get out-of-phase is  $N*P_1 = P_1*P_2/(P_1-P_2)$ . If  $P_1=P_2$ , the time is quite properly infinite: they never get out of phase.



Preliminary principle on two similar systems subject to a common random input

Two similar systems can show significantly-similar time-profiles of response to a common random input when the time it takes the two oscillatory impulse responses to get out of phase is significantly longer than the damping time constants of the two systems. In such cases, the initial portion of the response of the two systems to each random impulse will dominate their behavior, since the latter portions are diminished by damping. In the initial portion of the impulse response, the two systems oscillate fundamentally in-phase, but with the system with the faster response time (usually implying a shorter period) peaking and troughing ahead of the slower system. Thus, differences in period manifest themselves as phase differences rather than as a "beat" phenomenon.

The principle above implies a rule of thumb for time-series testing. Suppose that a model of some system is moderately-damped (damping time constants approximately equal in magnitude to the period of oscillation), and the real system is also moderately damped as it should be if the model is realistic. In such cases, if the behavior of the real system is driven by a known exogenous influence, then the model, when subjected to that same known exogenous influence, should show behavior similar to the real system's output. Regular phase-lag or phase-lead relationships between the real output and model output can result from a difference in natural period between the two. Of course, if some other type of discrepancy occurs (unrealistic magnitude of response, occasional large deviations, and so on), the modeler must look elsewhere to explain the discrepancies.

Other Examples: Menstruation and Squirrels. Probabilistic entrainment can explain two otherwise inexplicable biological phenomena. One phenomenon is that women who live together for long periods of time tend to menstruate together. It is well known that women menstruate with a period that averages 28 days. The period, however, is not inviolable: many events are capable of influencing, however slightly, the progress of the menstrual cycle, including diet, emotional upsets, and especially light and dark cycles. Thus, when a number of women that live together for a long period of time and subject themselves to the same exogenous influences of diet, emotional climate, and light and dark cycles, we would expect the hormonal oscillations to become entrained.

The second inexplicable phenomenon is that squirrels hibernate every winter, but will hibernate once a year anyway if cut off from all environmental stimuli. In other words, a squirrel kept in a cage with constant light and dark cycles and constant temperature (and even constant electronic and magnetic influences) will continue to hibernate once a year. Given the existence of an internal oscillator with a period of one year, how is it that all wild squirrels hibernate only during the winter? One tenable hypothesis based on the existence of probabilistic entrainment, is that environmental stimuli have some impact on the one-year oscillations, so that if a squirrel's internal timing mechanism should ever begin to deviate from producing hibernation each winter, environmental conditions (relative balance of light and dark, average temperature, or other influences) would tend to straighten out the cycle.

Most long-distance air travelers have noticed a phenomenon

similar to the entrainment of squirrels' hibernation. Humans, like squirrels, have a natural cycle of sleeping, which for us lasts anywhere from 23 to 26 hours per cycle. But environmental and social stimuli suffice to keep the vast majority of the human population together on a 24-hour schedule. When, however, someone travels and changes time zones, the traveler's internal cycle is at odds with the light-dark cycle and the eating cycle. The traveler experiences "jet lag" for around three days while his or her internal cycle readjusts to those in the new environment, and his or her daily schedule becomes much like everyone else's in that time zone.

Admittedly, the random inputs to squirrels and travellers are not pure white processes; they contain predominant periods of one year and one day, respectively. Still, the principle on similar systems with a common random input helps one explain these phenomenon. Section 6.3 discusses two similar systems subject to a common sinusoidal input as a topic for future investigation.

## 5.2 MATHEMATICAL DEVELOPMENT OF THE PRINCIPLE

The simulation in the previous sections suggests a definite pattern of common response to a common random input. Differences in period between the two similar systems reveal themselves as an apparent phase difference between their oscillation. Two heuristic justifications for this behavior were given in the previous section; these qualitative concepts indicate that some probabilistic entrainment should occur, but give only sketchy indications of the extent to which it occurs under various conditions. The qualitative concepts provide one type of certainty about the behavior of the combined system; they assert that probabilistic entrainment should occur to some extent in one large class of oscillatory systems. A different type of certainty about the behavior of the two systems can be obtained from the mathematical analysis below, which provides an exact quantitative description of the behavior of a very limited class of systems. In the case of probabilistic entrainment, a mathematical analysis is clearly complementary (rather than redundant) to the more conceptual simulation analysis given in the previous section. Readers not interested in a mathematical development may skip this section entirely, or use the table of contents to locate specific subsections of interest.

Mathematical Approach. Entrainment can be very broadly defined as the presence of a regular and strong relationship between conditions in one sector and the corresponding conditions in another sector. For example,

sectors entrained in an aggregate business cycle might consistently show peaks in production around the same time, with only small phase differences among them. One way of describing such relationships is to ask, given conditions in one sector, what are conditions in the other sector likely to be? For entrained sectors, if one sector is peaking, other sectors are likely to have peaked or be about to peak.

The mathematical analysis which follows does no more than give a quantitative answer to the question "given conditions in one sector, what are the conditions in the other?" Of course, because the inputs to both sectors are random, the relationship between the two sectors can only be characterized probabilistically: given conditions in one sector, what are the conditions in the other sector most likely to be, and with what certainty can that be stated? In mathematical terms, this question is equivalent to asking for the conditional probability density function for the state of one system, given the state of the other. The following analysis derives that function.

General Linear Case. We begin by analyzing the equations of motion for two general time-invariant linear systems (which also approximates a pair of time-invariant nonlinear systems oscillating close to an equilibrium). The equations will then be applied to a specific pair of second-order oscillators, to derive a general expression characterizing the probabilistic entrainment as a function of the system parameters. Finally, the expression will be evaluated numerically to delineate the para-

meter values for which the systems show probabilistic entrainment.

The equations for the two general linear systems can be combined into one composite equation, so that the dynamics of both systems can be described in a single set of equations. If the individual sector equations are

$$\begin{aligned} \dot{\underline{x}}_1 &= \underline{F}_1 \underline{x}_1 + \underline{G}_1 v(t) \\ \dot{\underline{x}}_2 &= \underline{F}_2 \underline{x}_2 + \underline{G}_2 v(t) \end{aligned} \quad \underline{G}_1 = \underline{G}_2$$

then the composite system is

$$\begin{pmatrix} \dot{\underline{x}}_1 \\ \dot{\underline{x}}_2 \end{pmatrix} = \begin{pmatrix} \underline{F}_1 & \underline{0} \\ \underline{0} & \underline{F}_2 \end{pmatrix} \begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \end{pmatrix} + \begin{pmatrix} \underline{G}_1 \\ \underline{G}_2 \end{pmatrix} v(t)$$

The input gain matrices  $\underline{G}_1$  and  $\underline{G}_2$  are equal, so that the systems can respond in the same manner to their common input. To simplify the notation, let the composite matrices be denoted

$$\begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \end{pmatrix} = \underline{x}, \quad \begin{pmatrix} \underline{F}_1 & \underline{0} \\ \underline{0} & \underline{F}_2 \end{pmatrix} = \underline{F}, \text{ and } \begin{pmatrix} \underline{G}_1 \\ \underline{G}_2 \end{pmatrix} = \underline{G}$$

Assume that  $v(t)$  is a scalar-gaussian white noise signal, with a covariance factor  $q$ . It can be shown that the response to  $v(t)$  is also gaussian, with zero mean, and with covariance defined as

$$\underline{\Gamma}(t) = E((\underline{x}(t) - E(\underline{x}(t))) (\underline{x}(t) - E(\underline{x}(t)))')$$

The covariance changes through time according to this differential equation.\*8\*

$$\frac{d}{dt}\underline{\Gamma}(t) = \underline{F}\underline{\Gamma}(t) + \underline{\Gamma}(t)\underline{F}' + \underline{G}q\underline{G}'$$

If  $\underline{F}$ ,  $\underline{G}$ , and  $q$  are known, the equation above specifies the steady-state covariance by setting

$$\frac{d}{dt}\underline{\Gamma}(t) = 0$$

and solving the resulting linear equations for the elements of gamma. The mean and the covariance completely specify the gaussian probability density function, which takes the form

$$p(\underline{x}) = [(2\pi)^k |\underline{\Gamma}|]^{-1/2} e^{-1/2 \underline{J}(\underline{x})}$$

where  $k$  is the dimension of the state vector  $\underline{x}$ , the vertical bars denote the determinant (of the steady-state covariance matrix), and

$$\underline{J}(\underline{x}) = (\underline{x} - \underline{m})' \underline{\Gamma}^{-1} (\underline{x} - \underline{m})$$

---

\*8\*F.C. Schweppe, Uncertain Dynamic Systems, Prentice-Hall, Englewood Cliffs, N.J., 1973, p.60.

Since the equations specifying the system dynamics have no constant terms, and if the initial conditions are zero, then the system's behavior is zero mean, so  $\underline{m}$  equals zero.  $J(\underline{x})$  is therefore completely characterized by gamma, which can be expressed in terms of the two subsystems that define  $\underline{x}_1$  and  $\underline{x}_2$

$$\underline{\Gamma} = \begin{pmatrix} \underline{\Gamma}_{11} & \underline{\Gamma}_{12} \\ \underline{\Gamma}_{21} & \underline{\Gamma}_{22} \end{pmatrix}$$

Then

$$J(\underline{x}) = [\underline{x}'_1 \quad \underline{x}'_2] \begin{pmatrix} \underline{\Gamma}_{11} & \underline{\Gamma}_{12} \\ \underline{\Gamma}_{21} & \underline{\Gamma}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \end{pmatrix}$$

To begin to derive an expression for the conditional probability density of  $\underline{x}_1$  and  $\underline{x}_2$ , first consider the defining equation for the conditional probability density.

$$p(\underline{x}_2 | \underline{x}_1) = \frac{p(\underline{x}_1, \underline{x}_2)}{p(\underline{x}_1)}$$

As indicated above, the joint probability density of  $\underline{x}_1$  and  $\underline{x}_2$  is gaussian, so whatever the specific value of  $p(\underline{x}_1)$ , the conditional probability density will always depend on  $\underline{x}_2$  as follows:

$$p(\underline{x}_2 | \underline{x}_1) = \text{constant} * e^{-1/2 J(\underline{x}_1, \underline{x}_2)}$$



Through tedious matrix manipulations, it can be shown that

$$p(\underline{x}_2 | \underline{x}_1) = \text{constant} * e^{-1/2 \tilde{J}(\underline{x}_2)}$$

where

$$\tilde{J}(\underline{x}_2) = (\underline{x}_2 - \tilde{\underline{x}}_2)' \tilde{\underline{\Gamma}}^{-1} (\underline{x}_2 - \tilde{\underline{x}}_2)$$

$$\tilde{\underline{\Gamma}} = \underline{\Gamma}_{22} - \underline{T} \underline{\Gamma}_{11} \underline{T}'$$

$$\underline{T} = \underline{\Gamma}'_{12} \underline{\Gamma}_{11}^{-1} = \underline{\Gamma}_{21} \underline{\Gamma}_{11}^{-1}$$

$$\tilde{\underline{x}}_2 = \underline{T} \underline{x}_1$$

These equations have several interesting features. The state of the first system,  $\underline{x}_1$ , completely summarizes the available knowledge about the state of the composite system. The most likely state of the second system,  $\underline{x}_2$ -tilde, is therefore a function of  $\underline{x}_1$ . However, the certainty with which the state of the second system can be known,  $\underline{\Gamma}$ -tilde, is independent of the specific state of the first system. The expression for  $\underline{\Gamma}$ -tilde contains only the difference between two positive semi-definite matrices. The first matrix,  $\underline{\Gamma}_{22}$ , represents the unconditional covariance of the state of the second system, so the second term represents the decrease in uncertainty due to information about the state of the first system. Because of the negative sign, that information can only decrease the uncertainty, which is proper.

The equations give the correct results for the case in which the first system is exactly identical to the second system. Since the corresponding state variables in the two systems would be exactly equal, covariances between corresponding pairs of state variables should also be exactly equal. Therefore, all four submatrices of the unconditional covariance matrix  $\gamma$  will be exactly equal. The product of any one submatrix and the inverse of any other submatrix will be the identity matrix. Thus, the state transformation matrix  $\underline{T}$  equals the identity matrix, which gives  $\underline{x}_2$ -tilde equal to  $\underline{x}_1$ . The equations correctly indicate that the states of identical systems will indeed be identical. Moreover, the error covariance by which the estimated (conditional) state of the second system may be known,  $\gamma$ -tilde, equals a zero matrix for identical systems.

The equations also give the correct results if both systems are undamped:  $\underline{T}$  goes to the zero matrix, and at least the diagonal elements of  $\gamma$ -tilde go to infinity. If both systems are oscillating (each with a different period), then there is no information about the state of the second system in the state of the first: they could be in phase, or they could be out of phase.

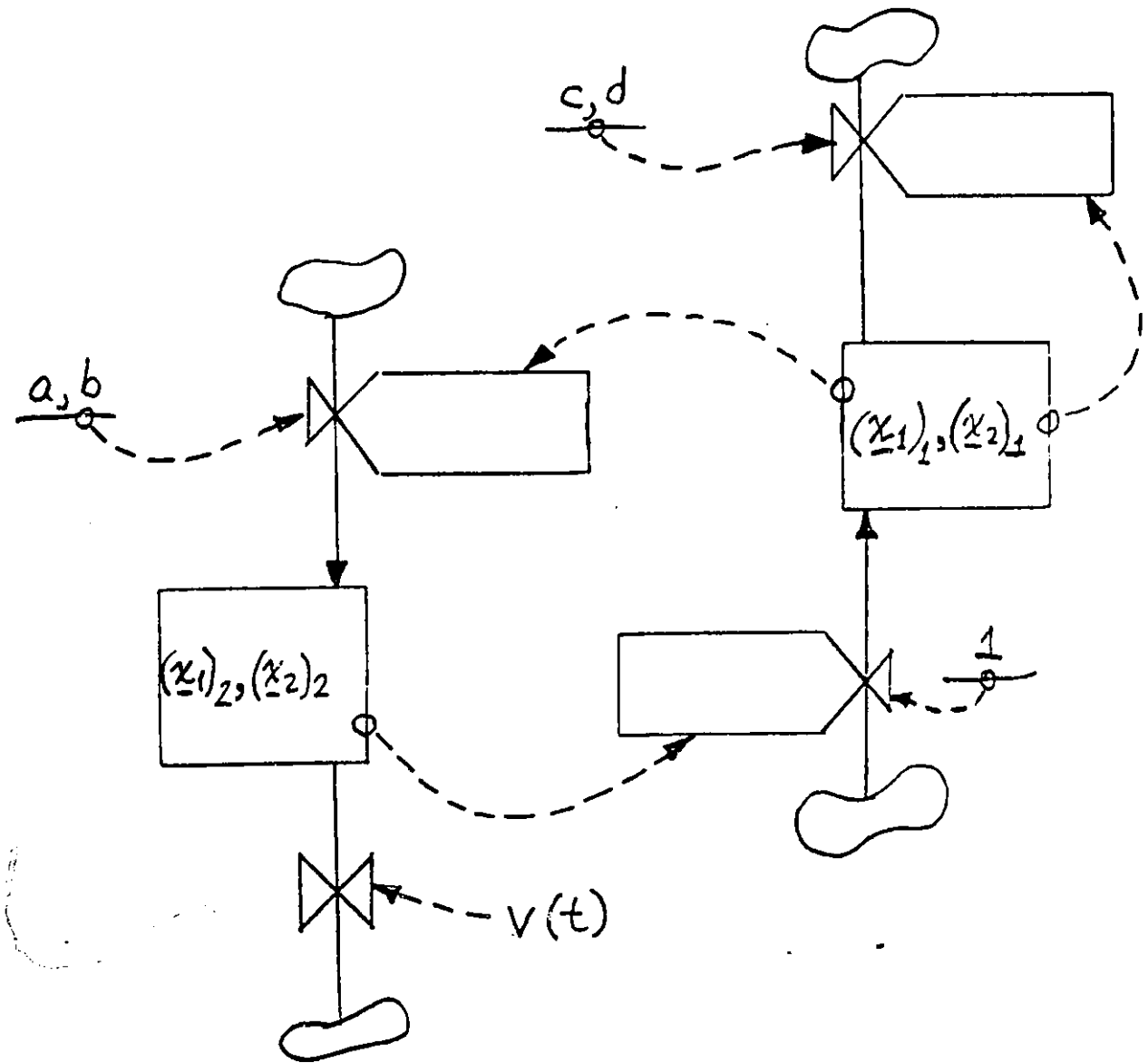
Equations for a Simple System. The equations given above cannot indicate the quantitative extent of probabilistic entrainment without specific values for the system structure and parameters. This subsection

uses two specific oscillatory systems to generate the conditional probability density function of the state of one system, given the state of the other. It is possible to perform such calculations for the employment-backlog system, but it seems more insightful to use even simpler systems, with a minimal number of parameters, and that equilibrate at zero.

Figure 5-5 shows the structures of two such systems. The parameters  $c$  and  $d$  control the damping, and the parameters  $a$  and  $b$  control the period (along to some extent with  $c$  and  $d$ ).

As indicated earlier, the steady-state unconditional covariance  $\gamma$  can be obtained by solving a set of linear equations resulting from setting the derivative of  $\gamma$  to zero:

$$\frac{d\gamma}{dt} = 0 = \underline{F}\gamma + \underline{G}q\underline{G}'$$



$$\dot{\underline{x}} = \begin{bmatrix} c & 1 & 0 & 0 \\ a & 0 & 0 & 0 \\ 0 & 0 & d & 1 \\ 0 & 0 & b & 0 \end{bmatrix} \underline{x} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} v(t)$$

Figure 5-5.  
FLOW DIAGRAM AND EQUATIONS FOR TWO OSCILLATORY SYSTEMS

For the pair of systems shown in Figure 5-5, that calculation results in the following submatrices of gamma:

$$\Gamma_{11} = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$$

$$\alpha = q/2ac$$

$$\beta = -q/2a$$

$$\gamma = q(c^2 - a)/2ac$$

$$\Gamma_{12} = \Gamma_{21}' = \begin{pmatrix} \delta & \xi \\ \rho & \nu \end{pmatrix}$$

$$\delta = q(c + d)/k$$

$$\xi = -q(b - a + cd + d^2)/k$$

$$\rho = q(b - a - cd - c^2)/k$$

$$\nu = q(c^2d - ac + cd^2 - bd)/k$$

$$k = (ad + bc)(c + d) - (a - b)^2$$

$$\Gamma_{22} = \begin{pmatrix} \eta & \sigma \\ \sigma & \lambda \end{pmatrix}$$

$$\eta = q/2bd$$

$$\sigma = -q/2b$$

$$\lambda = q(d^2 - b)/2bd$$

This completes the derivation of the conditional probability density function for the second system given the state of the first system. Consider now the original qualitative hypothesis, that for moderately-damped systems, differences in natural period manifest themselves in apparent phase differences between the two systems. Do the equations bear out the hypothesis? The state transformation matrix  $\underline{T}$  can be expressed in terms of the system parameters as follows:

$$\underline{T} = \underline{\Gamma}' \underline{\Gamma}^{-1}$$

$$\underline{\Gamma}'_{12} = \begin{pmatrix} \delta & \rho \\ \xi & \nu \end{pmatrix} = \frac{1}{k/q} \begin{pmatrix} c+d & b-a-cd-c^2 \\ -b+a-cd-d^2 & c^2d-ac+cd^2-bd \end{pmatrix}$$

$$\underline{\Gamma}^{-1}_{11} = -\frac{2c^2}{q} \begin{pmatrix} \frac{c^2-a}{c} & 1 \\ 1 & \frac{1}{c} \end{pmatrix}$$

$$\underline{T} = -\frac{2c^2}{k} \begin{pmatrix} (b-a)-a(1+\frac{d}{c}) & (\frac{b-a}{c}) \\ -d(b-a)+a(\frac{b-a}{c})-bc+\frac{ad^2}{c} & -b(1+\frac{d}{c}) \end{pmatrix}$$

If the transformation represented by  $\underline{T}$  truly induced only a phase shift, the angle between  $\underline{x}_1$  and  $\underline{x}_2$ -tilde would be constant, as illustrated in Figure 5-6. It is quite easy to compute the tangent of the angle  $\theta$  for two specific values of  $\underline{x}_1$ , the unit vectors  $\underline{e}_1$  and  $\underline{e}_2$ :

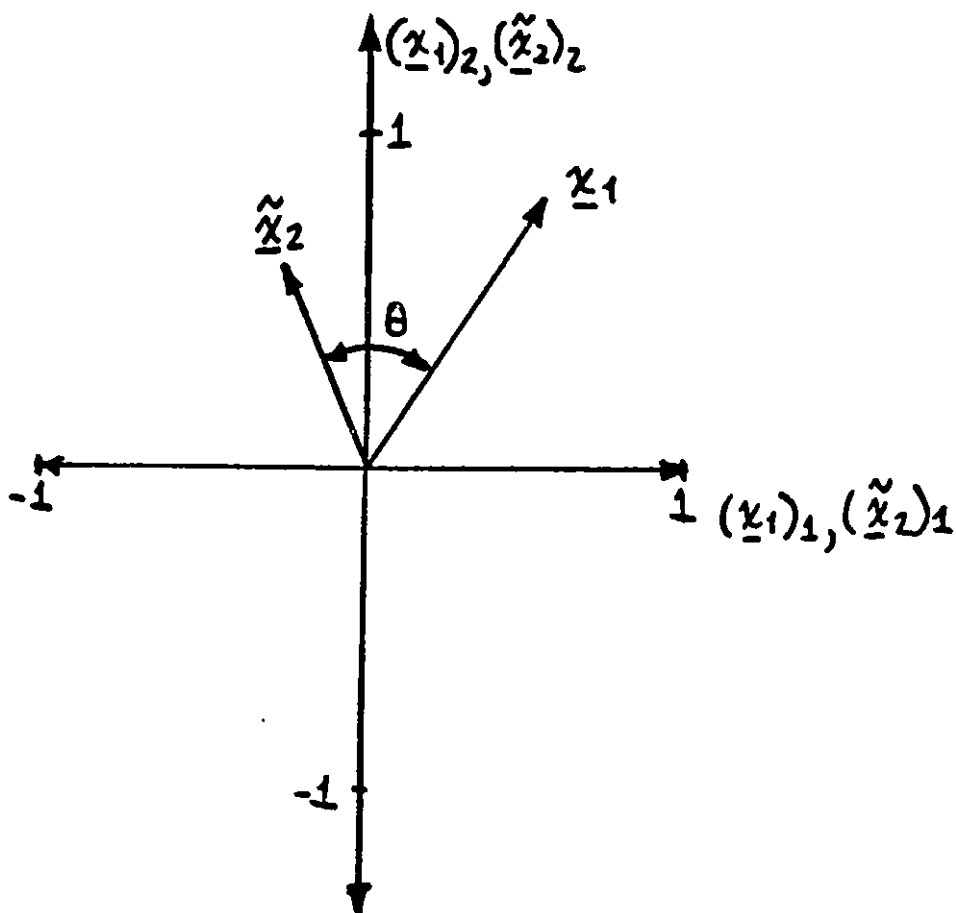


Figure 5-6.  
PHASE SHIFT FROM  $\underline{x}_1$  TO  $\underline{x}_2$

$$\underline{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \underline{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

The tangent values are just

$$\tan \theta_1 = \frac{(\underline{T}\underline{e}_1)_2}{(\underline{T}\underline{e}_1)_1} = \frac{T_{21}}{T_{11}}$$

$$\tan \theta_2 = \frac{(\underline{T}\underline{e}_2)_1}{(\underline{T}\underline{e}_2)_2} = -\frac{T_{12}}{T_{22}}$$

By inspection, the tangents of  $\theta_1$  and  $\theta_2$  are not equal. However, when the systems are identical

$$\theta_1 = \theta_2 = 0$$

and, because both angles are continuous functions of the system parameters, the angles will be nearly equal when the system parameters are nearly equal. Only numerical calculations can indicate the extent to which the state transformation matrix  $\underline{T}$  induces an approximate phase shift effect.

Numerical Results for the Simple System. Before considering the quantitative extent of the phase relationship, let us consider the quantitative extent of any consistent relationship. How might we characterize the consistency of a relationship--the confidence with which we can assert that the two system states will be related? There are a number of measures, called distance measures, that quantify the extent to which two probabilistic processes are similar; the divergence and the Battacharya distance are the most common.\*9\* Both of these measures are based on functions of the probability that a single given trajectory could have

-----

\*9\*Schweppe, op. cit., pg. 263.



been generated by either of two systems. Thus, both of these measures would deem very similar two identical undamped oscillatory systems, even though their outputs would not be in any way related to one another. We seek not a measure of how related systems are (by hypothesis, we know they are highly related), but of how related the two outputs are.

The well-known t-statistic provides a model for an index of confidence in the relationship of one system output to the other. One definition for the t-statistic for a given parameter estimate is just the magnitude of the estimated parameter divided by the standard deviation of the parameter estimate. The exact equivalent for the vector case is analytically intractable, so instead we can define an index with the corresponding quantities squared:

$$N = E_{\underline{x}_1} [(\underline{\tilde{x}}_2 | \underline{x}_1)' \underline{\tilde{\Gamma}}^{-1} (\underline{\tilde{x}}_2 | \underline{x}_1)]$$

The index N takes the "square" of the vector "divided by" the "square" of the "standard deviation," that is, multiplied by the inverse of

the conditional covariance matrix. The expectation is taken to cause the index to be a single number, instead of a function of the state of the first system.\*10\* We can derive N as a function of only the system parameters as follows.

Substituting the definition of  $\underline{x}_2$ -tilde,

$$N = E (\underline{x}'_1 \underline{T}' \underline{\tilde{\Gamma}}^{-1} \underline{T} \underline{x}_1)$$

-----

\*10\*Note that N is not the same as

$$\begin{aligned} & E_{\underline{x}_2, \underline{x}_1} [(\underline{\tilde{x}}_2 - \underline{x}_2)' \underline{\tilde{\Gamma}}^{-1} (\underline{\tilde{x}}_2 - \underline{x}_2)] \\ &= E_{\underline{x}_2, \underline{x}_1} [\text{tr}(\underline{\tilde{x}}_2 - \underline{x}_2) (\underline{\tilde{x}}_2 - \underline{x}_2)' \underline{\tilde{\Gamma}}^{-1}] \\ &= \text{tr} (E_{\underline{x}_2, \underline{x}_1} [(\underline{\tilde{x}}_2 - \underline{x}_2) (\underline{\tilde{x}}_2 - \underline{x}_2)'] \underline{\tilde{\Gamma}}^{-1}) \\ &= \text{tr} \underline{\tilde{\Gamma}} \underline{\tilde{\Gamma}}^{-1} = \text{tr} I = n \end{aligned}$$

where n is the order of each system

which from matrix algebra we know is equivalent to

$$N = E (\text{tr}(\underline{T}' \underline{\tilde{\Gamma}}^{-1} \underline{T} \underline{x}_1 \underline{x}'_1))$$

The trace operator  $\text{tr}$  takes the sum of the diagonal elements of a square matrix. Both the trace operator and the expectation operator are linear so they commute and associate.

$$\begin{aligned} N &= \text{tr}(E(\underline{T}' \underline{\tilde{\Gamma}}^{-1} \underline{T} E(\underline{x}_1 \underline{x}'_1))) \\ &= \text{tr}(\underline{T}' \underline{\tilde{\Gamma}}^{-1} \underline{T} E(\underline{x}_1 \underline{x}'_1)) \\ &= \text{tr}(\underline{T}' \underline{\tilde{\Gamma}}^{-1} \underline{T} \underline{\Gamma}_{11}) \end{aligned}$$

We can carry the computation further to arrive at an intuitively-appealing result. Substituting in the definitions of  $T$  and  $\tilde{\Gamma}$ ,

$$N = \text{tr}([\underline{\Gamma}_{11}^{-1} \underline{\Gamma}_{12} (\underline{\Gamma}_{22} - \underline{\Gamma}'_{12} \underline{\Gamma}_{11}^{-1} \underline{\Gamma}_{12})^{-1} \underline{\Gamma}_{12} \underline{\Gamma}_{11}^{-1}])$$

Using a matrix identity on the expression within brackets,

$$\begin{aligned} N &= \text{tr}([\underline{\Gamma}_{11}^{-1} + (\underline{\Gamma}_{11} - \underline{\Gamma}_{12} \underline{\Gamma}_{22}^{-1} \underline{\Gamma}'_{12})^{-1}]) \\ &= \text{tr}([\underline{\Gamma}_{11} - \underline{\Gamma}_{12} \underline{\Gamma}_{22}^{-1} \underline{\Gamma}'_{12}]^{-1} \underline{\Gamma}_{11} - I) \\ &= \text{tr}([\underline{\Gamma}_{11} - \underline{\Gamma}_{12} \underline{\Gamma}_{22}^{-1} \underline{\Gamma}'_{12}]^{-1} \underline{\Gamma}_{11}) - n \end{aligned}$$

The expression now inside the brackets should seem familiar. It has the same form as the expression for  $\tilde{\gamma}$ , the covariance of the estimate of  $\underline{x}_2$  given  $\underline{x}_1$ . The expression now within the brackets is the error-covariance that would result from estimating  $\underline{x}_2$  given  $\underline{x}_1$ . This symmetry is reassuring. It says, in effect, that there is a two-way relationship between the two systems: whenever knowing something about  $\underline{x}_1$  says something about  $\underline{x}_2$ , then knowing something about  $\underline{x}_2$  says something about  $\underline{x}_1$ .

The entrainment index  $N$ , then, takes a scalar measure, the trace, of the error covariance of an  $\underline{x}_1$  estimated from  $\underline{x}_2$ . That error covariance is inverted to give a high index of entrainment for low error covariance. The inverse of the error covariance is also normalized with respect to the noise present in the systems by multiplication by the unconditional covariance of  $\underline{x}_1$ , i.e., the covariance we can expect from  $\underline{x}_1$  without any knowledge of  $\underline{x}_2$ . In effect, the multiplication of the two matrices compares the error covariance an estimate of  $\underline{x}_1$ , with and without knowledge of  $\underline{x}_2$ . If knowledge of  $\underline{x}_2$  decreases the error covariance significantly, then the index of entrainment is large. Finally  $N$  is normalized with respect to the order of the systems by subtracting  $n$ .

Note that, analogously to the  $t$ -statistic that inspired its definition,  $N$  goes to infinity when the two systems are identical and  $\underline{x}_1$  equals  $\underline{x}_2$ . This implies that

$$\Gamma_{-11} = \Gamma_{-22} = \Gamma_{-12} = \Gamma_{-21}$$

So

$$\begin{aligned}
 N &= \text{tr}([\underline{\Gamma}_{11} - \underline{\Gamma}_{12} \underline{\Gamma}_{22}^{-1} \underline{\Gamma}'_{12}]^{-1} \underline{\Gamma}_{11}) - n \\
 &= \text{tr}([\underline{\Gamma}_{11} - \underline{\Gamma}_{11} \underline{\Gamma}_{11}^{-1} \underline{\Gamma}_{11}]^{-1} \underline{\Gamma}_{11}) - n \\
 &= \text{tr}([\underline{\Gamma}_{11} - \underline{\Gamma}_{11}]^{-1} \underline{\Gamma}_{11}) - n \\
 &= \text{tr}([\underline{0}]^{-1} \underline{\Gamma}_{11}) - n
 \end{aligned}$$

N thus approaches infinity as  $\underline{x}_1(t)$  approaches  $\underline{x}_2(t)$  for all systems with nonzero steady-state covariance.

If there is no relationship between  $\underline{x}_1$  and  $\underline{x}_2$ .

$$\underline{\Gamma}_{12} = \underline{\Gamma}_{21} = \underline{0}$$

So that

$$\begin{aligned}
 N &= \text{tr}([\underline{\Gamma}_{11} - \underline{\Gamma}_{12} \underline{\Gamma}_{22}^{-1} \underline{\Gamma}'_{12}]^{-1} \underline{\Gamma}_{11}) - n \\
 &= \text{tr}(\underline{\Gamma}_{11}^{-1} \underline{\Gamma}_{11}) - n \\
 &= \text{tr} \underline{I} - n \\
 &= n - n = 0
 \end{aligned}$$

So N, again analogous to the t-statistic, goes to zero for unrelated system outputs.

While the computation of  $N$  from the system parameters is complex, it can be done, since each of the matrices is already known as functions of the system parameters. Figure 5-7 shows the entrainment index  $N$  as a function of the ratio of the natural periods of the two systems, for several different values of the respective time constants.\*11\* Significant entrainment would seem to occur when  $N$  equals about 4. (The corresponding point for the  $t$ -statistic occurs where the parameter value divided by its standard deviation equals 2; a  $t$ -statistic of 2 is usually considered significant, although not exceptionally so.) As might be suspected from the simulations in the previous section, significant entrainment occurs for moderately-damped systems with fairly disparate natural periods. For example, with the damping time constants equal to the natural period of the second system, the entrainment index  $N$  exceeds 4.0 if the natural period of the first system is anywhere between 80% and 130% of the natural period of the second system. For damping time constants equal to half the natural period of the second system, the range of entrainment for natural periods of the first system is from 67% to 200% of the period of the second.

-----

\*11\*A DYNAMO program was written to implement the calculations described previously, and to graph the results. Appendix G lists the program and the rerun parameters and Appendix I defines the variables in the program.

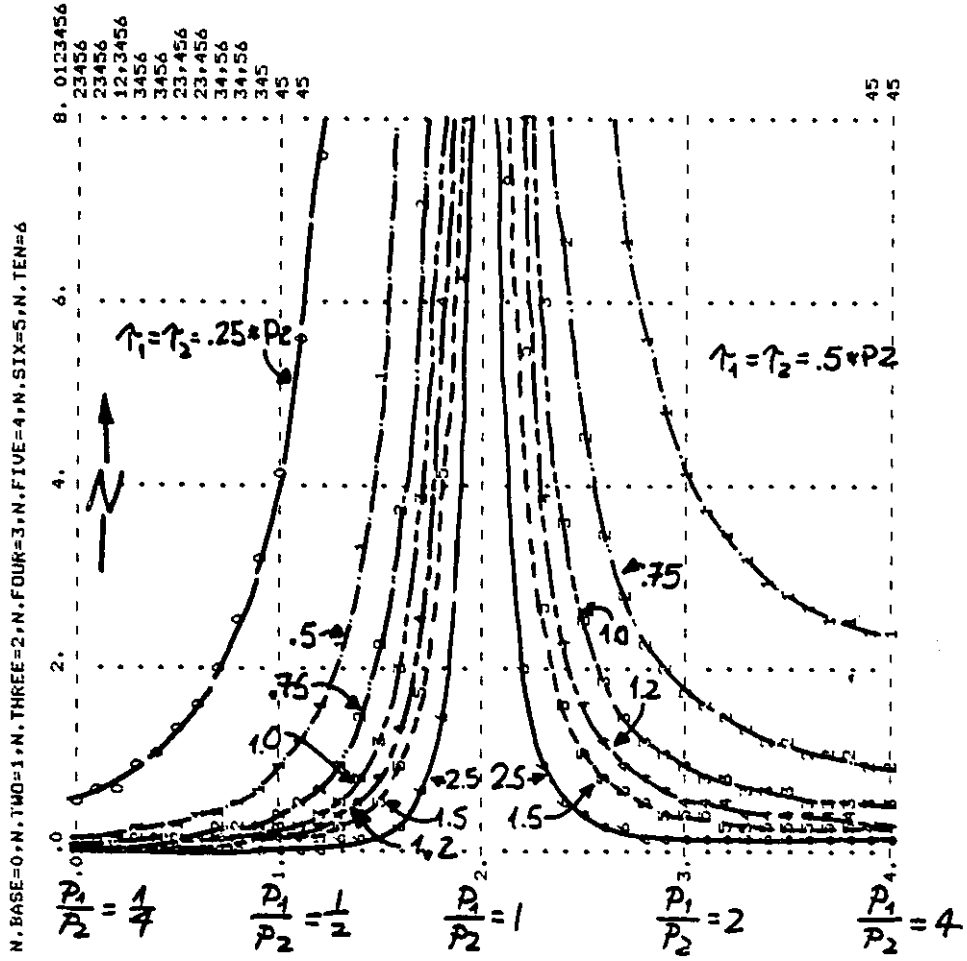


Figure 5-7.  
 ENTRAINMENT INDEX N FOR DIFFERENT PERIODS

The heuristic analysis in the previous section places considerable importance on differences between the periods, and suggests, by omission, that differences between the damping time constants of the two systems are not nearly so important. Figure 5-8 confirms this, by showing the same curves as in the previous figure, but with the damping time constants differing by 40% between the first and second system. However, the geometric means of the damping time constants for the corresponding curves are exactly equal. The resulting entrainment behavior, as characterized by the index  $N$ , is remarkably similar to the case of equal damping time constants.

As noted above, the phase shift between the state of the first system and the conditional state of the second system depends on the state of the first system. Different states produce different apparent phase shifts. We can examine the phase shift induced by the state transformation matrix  $\underline{T}$  for the two unit vectors  $\underline{e}_1$  and  $\underline{e}_2$ . The tangents of the resulting phase shifts are shown in Figures 5-9 and 5-10. The curves show an unmistakable trend: the system with the longer natural period will typically lag behind the system with the shorter period. (On Figure 5-9, the tangent of the phase angle jumps from negative values to positive values when the angle passes through 90 degrees; this discontinuity is a property of the tangent function). Although the phase shift angles are not identical for the two unit vectors, within the ranges of periods where significant entrainment can occur, the angles are close enough to result in an apparent phase



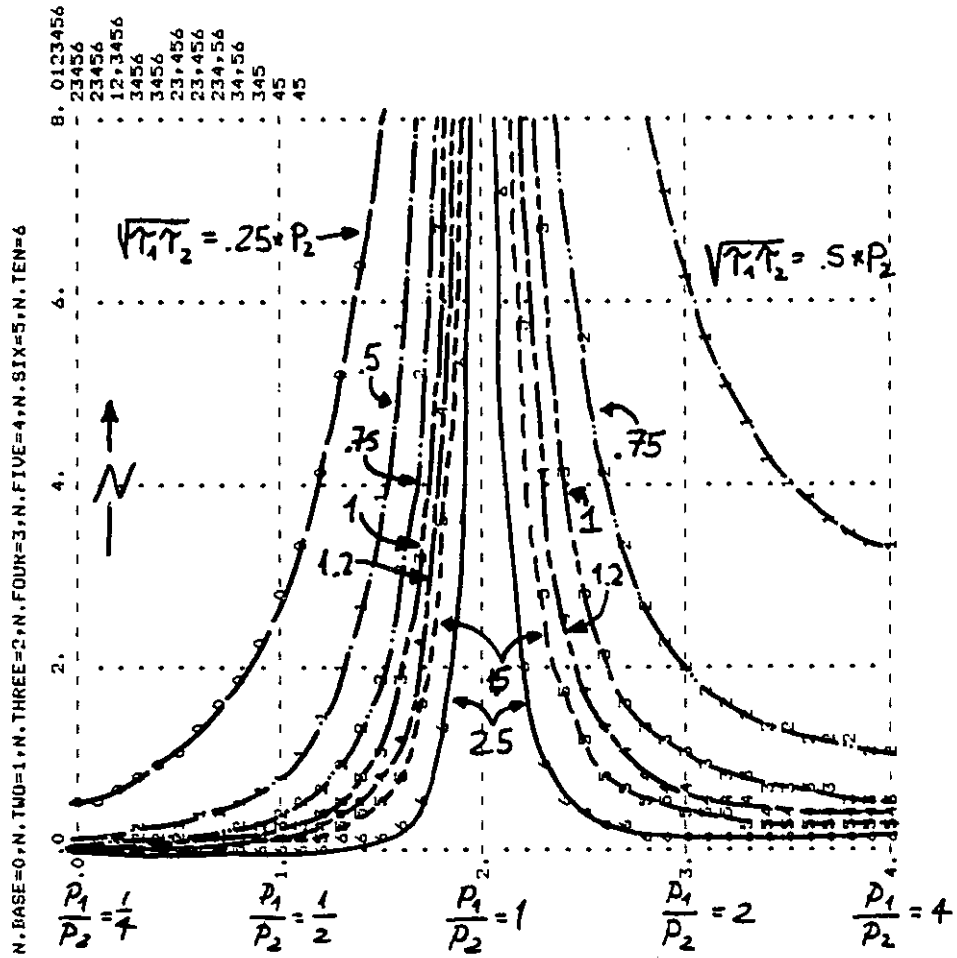


Figure 5-8.  
ENTRAINMENT INDEX N FOR UNEQUAL DAMPING TIME CONSTANTS

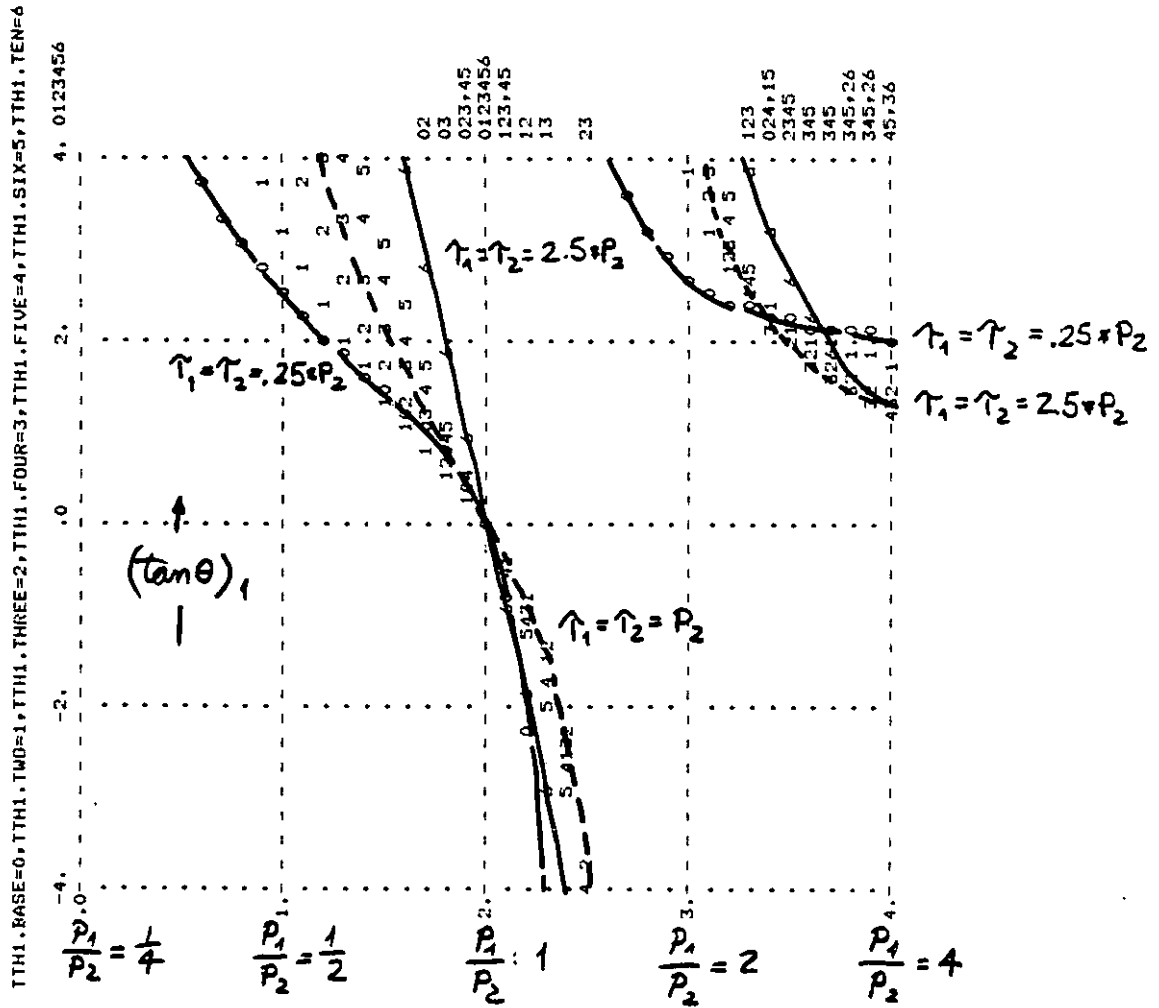


Figure 5-9.  
PHASE SHIFT FROM I FOR UNIT VECTOR 1

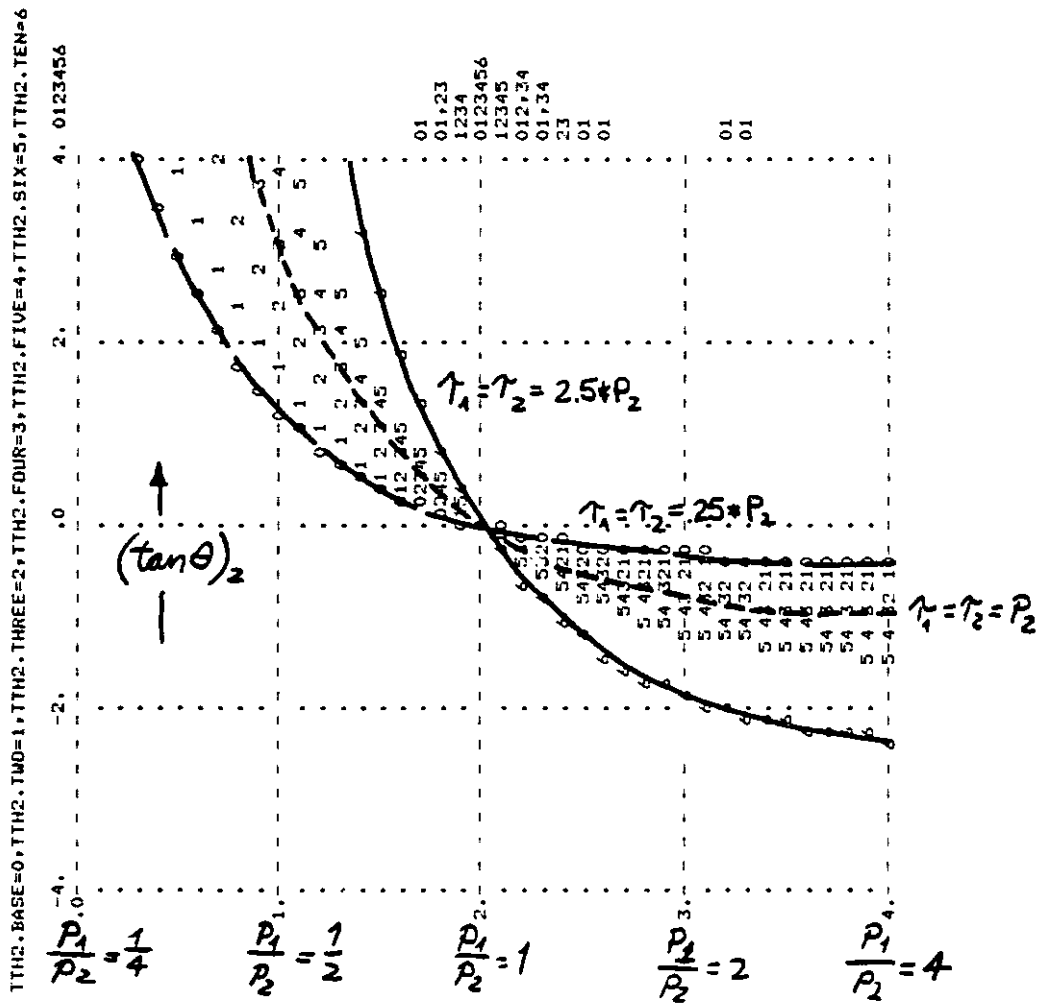


Figure 5-10.  
 PHASE SHIFT FROM T FOR UNIT VECTOR 2

relation between the two systems. For example, consider a pair of systems whose periods are at the outer limit of being able to show entrainment. The damping time constants are equal to the period of the second system, and the first system has a period 80% of the period of the second system.  $\theta_1$  is about 45 degrees, and  $\theta_2$  is about 22 degrees. In a more extreme case where the period of the first system is about 30% greater than the period of the second system,  $\theta_1$  is about 90 degrees and  $\theta_2$  is again about 22 degrees. Even these disparate angles have the proper signs, and are not so far apart as to destroy the appearance of a phase shift relation between the first and second systems.

The results discussed above describe a steady-state probabilistic entrainment. How long does it take for systems to fall into this entrainment? Can probabilistic entrainment occur quickly enough to happen in real life, or is it a statistical artifact? In the simulations in the previous section, the systems seemed to be entrained from the start, although that could result from both systems starting in exact equilibrium. Or was this just a coincidence? A mathematical analysis can provide more certainty about the time required for systems to fall into probabilistic entrainment.

The equations developed previously show that the entrainment index  $N$  can be expressed as a function of only the submatrices of the unconditional covariance matrix  $\gamma$  for the combined systems. Values for the elements of  $\gamma$  were derived from the differential equation

that describes how gamma changes over time. By implementing that differential equation directly, we can compute the entrainment index  $N$  as a function of time.

Figure 5-11 shows  $N(t)$  for several combinations of parameters.\*12\* By design, all of the combinations of parameters results in a steady-state  $N$  of about 4. The specific parameter values are given on the figure itself. The covariance matrices are initialized so that the states of both systems have their steady-state amount of randomness, but with no statistical relationship between the two systems. In general, the time taken for two systems to become entrained seems very close to the magnitude of their damping time constants. For example, the highly-damped systems whose entrainment index is shown on Curve 1 have very little momentum in their oscillations and can rather quickly show a similar response to current values of their common random input. In contrast, the lightly-damped systems whose entrainment index is shown on Curve 4 possess considerable momentum in their oscillations. A much longer time period is required for the common input fluctuations to nudge the systems into an entrained relationship. This result makes good intuitive sense, since the damping time constants characterize the time it takes for the system levels to "forget" things -- everything about the system's state, including being out of phase, disappears with these time constants.

-----

\*12\*The DYNAMO program that produced the plots is given in Appendix H. Definitions of the variables in the program appear in Appendix I.



All of the cases shown in Figure 5-11 result in steady-state values of the entrainment index  $N$  of about 4. However, most of the curves seem to overshoot and then return to the steady-state value. (Curve 4 does in fact merely overshoot and return to its steady-state value of about 4, but the eventual return requires many times the length of the simulations shown.) I suspect that the initial peaks in the entrainment index  $N$  represent the first time in which the two systems have the opportunity to oscillate in phase. However, as time passes, the differences between the two systems eventually cause them to fall somewhat out of entrainment. Certainly the peaks are not well understood, but further analysis of the phenomenon would seem to fall outside the scope of this thesis.

With the mathematical analysis completed, let us return for a moment to the heuristic development of Section 5.1. The discussion there suggested that the tendency for the outputs of the similar systems derives from the time taken for the oscillatory impulse response to damp out being shorter than the time the oscillations at the two different periods require to get out of phase. This suggests that it might be possible to form a simple heuristic measure of entrainment  $NH$ . Use the geometric mean of the damping time constants of the two systems as a scalar characterization of the damping time constants. Use the formula in footnote 7 to determine the time to get out of phase:

$$t_{\text{out of phase}} = \frac{P_1 P_2}{|P_1 - P_2|}$$

Taking the ratio of these two quantities, and multiplying by some constant to make the magnitude of NH comparable to the magnitude of N,

$$NH = \frac{P_1 P_2 k}{|P_1 - P_2| \sqrt{DTC_1 * DTC_2}}$$

To determine a value for k, use one of the sets of parameters on Figure 5-11 which results in N equal to 4:

$$NH = 4 = \frac{5.19 * 4 * k}{|5.19 - 4| \sqrt{4 * 4}}$$

Solving the equation yields a value for k of 0.91. (Other sets of parameters yield values of k fairly close to 0.91, which bodes well for the accuracy of the heuristic measure.) Thus, significant probabilistic entrainment occurs when

$$\frac{0.91 * (t_{out \text{ of phase}})}{(\text{geometric mean of damping time constants})} > 4$$

The equations above imply that if  $DTC_1 = DTC_2 = P_1$ , then  $P_2$  can be anywhere between 80 percent and 125 percent of  $P_1$ , and significant entrainment



can occur. This is in rough accord with the results on Figure 5-7. If we are willing to let  $NH=4 \times 0.91=3.64$  be considered significant entrainment, we can say that significant entrainment occurs when the time to get out-of-phase is equal to or greater than four times the geometric mean of the damping time constants.

To test more thoroughly the accuracy of the heuristic index of probabilistic entrainment  $NH$ , Figure 5-12 shows values of  $NH$  for the same periods and damping time constants that were used to display  $N$  in Figure 5-7. Although the shapes of the  $N$  and  $NH$  curves differ, careful inspection reveals that  $NH$  and  $N$  agree very well on which systems will and will not show probabilistic entrainment. If anything, the heuristic index  $NH$  is slightly more conservative than the mathematical index  $N$ , that is,  $N$  might indicate that borderline cases do show probabilistic entrainment, whereas  $NH$  would not so indicate.

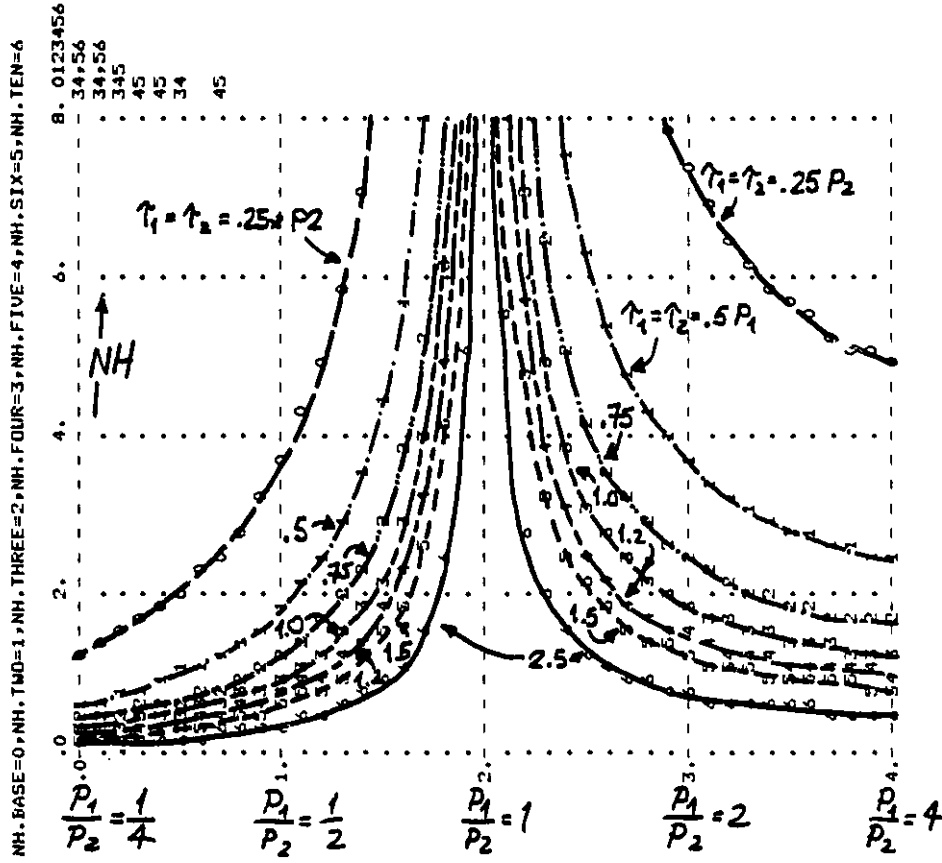


Figure 5-12.  
 HEURISTIC ENTRAINMENT INDEX NH FOR DIFFERENT PERIODS

A Restatement of the Principle. With the heuristic index of probabilistic entrainment fairly well validated by comparison to a much more rigorously-derived measure, we can restate the principle on probabilistic entrainment with more precision:

Principle on two similar systems subject to a common random input

Two similar oscillatory systems will show significantly-similar time profiles of response to a common random input when the time it takes for the two oscillations to go from in-phase to out-of-phase equals or exceeds four times the geometric mean of the two damping time constants. In such a case, the initial portion of the response of the two systems to each random impuse will dominate their behavior, since the latter portions are diminished by damping. In the initial portion of the impulse response, the two systems oscillate fundamentally in-phase, but with the system with the faster response time (usually implying a shorter period) peaking and troughing ahead of the slower system. Thus, differences in period manifest themselves as phase differences, rather than as a "beat" phenomenon.

The principle above seems to cover systems with one oscillatory behavior mode fairly well. But note that we can consider all of the preceding analysis as applying to one behavior mode of a linearized multimodal system as well. Thus we would expect productive sectors in a national economy, which seem to show moderatley-damped oscillations around both four- and twenty-year periods, to show probabilistic entrainment in both behavior modes.\*13\*

---

\*13\*See Mass, J.N., Economic Cycles (Cambridge, Mass.: Wright-Allen Press, 1974) for a description of these two behavior modes.

The principle above implies a convenient rule of thumb for deciding when simply comparing model output with real time series on a point-by-point basis is appropriate as a validity test, or when more elaborate techniques such as Kalman-Bucy filtering are necessary.\*14\* If the model is capable of showing probabilistic entrainment, with systems similar to itself it should be capable of showing probabilistic entrainment with the real system, and comparing model output with real time series is valid as a validity test. If such a test is performed, then, as remarked in Section 5.1, phase differences between model output and time series could be attributed to a difference in natural period between the model and the real system. If the model cannot show probabilistic entrainment with systems similar to itself, then a more elaborate whole-model test of validity, such as tests derived from Kalman-Bucy filter theory, is necessary.

---

\*14\*This thesis is not the appropriate place to elaborate on Kalman-Bucy filtering. Suffice it to say that it can give whole-model validity tests (as opposed to tests of individual model relationships) even for systems that are unable to show probabilistic entrainment. For an explanation of Kalman-Bucy filtering, see Schweppe, op. cit., Chapter 6. For an exposition of validity tests that can be derived from Kalman-Bucy filtering theory, see Peterson, D.W. "Hypothesis, Estimation, and Validation of Dynamic Social Models." Unpublished Ph.D. dissertation (Cambridge, Mass.: MIT, 1975).

Observations on Mathematical Techniques. As the present section contains the only extensive example of mathematical analysis in this thesis, it is appropriate here to reflect on the choice between mathematical analysis and simulation methods. The two are not alternatives, nor substitutes for one another. They differ in the generality and the certainty of their conclusions. They differ in the level of effort required to achieve results, and they provide different kinds of opportunities for unexpected insights. The choice between mathematical analysis and simulation analysis really depends on the purpose and current status of the analysis, and not upon the predilections of the analyst.

Mathematical analysis and simulation analysis produce results that apply to different classes of systems, and with different degrees of certainty. Simulation analysis produces very general, intuitive concepts that, at least in principle, apply to many similar systems. The ability to apply a basic set of concepts about system behavior to wide classes of systems usually appears under the rubric of "transferability of structure." However, the general concepts of system behavior are not certainties for each system being studied; even if the concepts are correct, simulation testing is necessary to confirm and refine them. (One example is the simple principle on the origin of oscillations of the production sector given in Section 2.1. Many simulations were needed to be able to assert the causes of such system behavior with confidence. In contrast, mathematical analysis provides

very certain and very exact conclusions about particular properties of one system structure (or any different system structures which result from only parameter changes.)\*14\*

Mathematical analysis says little or nothing about systems similar to the system being analyzed. For example, the mathematical analysis in this section gives no confidence that the results apply to, for example, nonlinear oscillatory systems. (The heuristics in Section 5.1, however, indicate that the results do transfer to other linear or smoothly-nonlinear oscillatory systems, but the mathematical analysis cannot say.) As another example, formal sensitivity analysis can yield sensitivities for every parameter of the system in the neighborhood of the nominal set of parameter values. However, the formal analysis provides no indication of the consequences of either large parameter changes, multiple parameter changes, or parameter changes under different structural assumptions. In contrast, a general concept of system behavior (such as the identification of the dominant loops of the system) provides at least tentative indications.

-----

\*14\*The generality of mathematical results should not be overstated, however. Most useful and realistic models seem to be complex enough that closed-form analytical solutions do not yield much tangible information. Therefore, in even the most sophisticated mathematical analysis, one must eventually resort to numerical methods to yield tangible results. For example, the equations that describe entrainment of two general linear systems provide very little description of the class of systems which can be so entrained.

Mathematical analysis and simulation analysis can differ dramatically in the level of effort required to obtain results. This is not to say that the greater certainty of mathematical analysis necessarily involves more effort in all cases. In this section, the mathematical analysis handled the probabilistic nature of the problem much more cleanly and simply than could a corresponding simulation analysis. In some large systems, such as power systems, the structure of the system is so involved that the sources of behavior within the structure are virtually impossible to identify with simulation. In such cases, the initial overhead involved in setting up a mathematical description of the system is in fact the most effective means of arriving at conclusions about the system's behavior. There are, however, many problems for which mathematical analysis is not the most efficient means. For example, the principles in Chapters 2 through 4 derive from simulation results. In theory, those or similar conclusions could also have been drawn from a mathematical analysis. But the cost would have been high relative to the insights gained.

Mathematical analysis and simulation analysis differ in their fertility for new insights. Each activity provides opportunities for different kinds of insights about the systems. For example, it is unlikely that enough simulations would have led me to discover that the time two sectors require to become entrained is about the same time as the damping time constants of the two systems. In contrast, the form

of the mathematical analysis made the investigation and the conclusion quite natural and straightforward. Similarly, the example of a purely positive loop oscillating followed quite naturally from the analysis of sinusoids propagating around a second-order loop. Simulation and conceptual analysis provide the opportunity for different sorts of insights. Forrester's observations on the characteristics of complex systems are a perfect example of principles that derive from repeated simulations.\*15\* Such insights can in theory derive from any number of techniques of analysis, and obviously each technique contains biases toward some insights (or types of insights) and away from others.

Finally, the appropriateness of mathematical analysis versus more conceptual simulation analysis depends on the current status of the analysis. This chapter provides an example of shifting from conceptual to mathematical analysis and back. The order of presentation in this chapter is quite close to the way in which the research actually progressed: initially, there was no dynamic hypothesis - no perception of probabilistic entrainment - upon which to base a mathematical analysis. Simulation provided an opportunity to observe probabilistic entrainment, and to motivate the search for a heuristic explanation. With an explanation of the "how" of probabilistic entrainment, the next question naturally asked "how much," and the mathematical analysis yielded the

-----

\*15\*Forrester, J.W., Urban Dynamics (Cambridge: MIT Press, 1969), Chapter 6.



answer. That mathematical machinery provided a means of validating and refining the initial heuristic principle, as well as facilitating the insight that the concepts of the principle applied to each oscillatory behavior mode of a multimodal system. Thus, the analysis for this chapter alternately used mathematical analysis and conceptual simulation analysis, gaining certainty and generality with each alternation.

CHAPTER SIX:

CONCLUSION

Chapter 6 reviews the principles developed in the first five chapters, and the concepts out of which the principles are built. The chapter goes on to describe three types of results which could arise from the explication of these principles: their immediate value to a reader of this thesis, the various avenues of research they open up, and their function as focal points for developing an organized and efficient curriculum to teach the relation between system structure and system behavior.

CONTENTS

6.1. Summary of Concepts Introduced	316
6.2. Summary of Principles Developed	320
6.3. Consequences of Thesis	323
Fulfilling the Purpose: Value to the Reader	323
Avenues for Further Development of Principles	324
Role in Curriculum Development	332

## 6.1 SUMMARY OF CONCEPTS INTRODUCED

Below are listed the concepts introduced in this thesis, and their definitions. Some are original, and some are merely more precise and technical specification of normal English definitions.\*1\* Their role here is not to summarize the thesis results (which are the principles relating structure to behavior), but to serve as the framework within which the principles can be stated. The concepts below serve as intermediaries between the specific examples and the general principles: they tie together common features of the examples under one concept. (For example, the structure of the spring-mass system and the structure of the employment-backlog system are tied together by the concept of "loop"; both systems of cause-and-effect relationships form a loop.) The principles provide still more organization and abstraction by tying together several concepts into a specific relationship. ("Oscillations occur when a disturbance propagates around a loop formed of phase-lag subsystems to return the system to a state comparable to the initial disequilibrium state.")

---

\*1\*The list here by no means lists all of the concepts used or defined in the thesis. For the purpose of concluding the thesis, only the relatively unfamiliar concepts actually used in the principles are listed; the remainder await some enterprising pedagogue to compile a glossary of System Dynamics terms.

If one were to create an analogy for the relationship among the examples, the concepts, and the principles, the analogy might be stated in terms of access to a filing system. Specific papers go into one of many folders, which in turn go into one of many locked file drawers. The principle corresponds the key that unlocks the appropriate collection of folders (concepts) within which the relevant papers (relevant experiences) may be accessed. In terms of the analogy, then, the reader is being given a means of classifying and storing specific experiences so that shortly, the principles themselves can give the reader the keys to all of the information in the thesis.

### Subsystem

A collection of rates, levels, and auxiliaries within a system with one input and one output, which, for the purpose of analyzing the behavior of the overall system, is characterized only by its input-output characteristics; internal structure per se is not of major concern in the analysis.

### Phase-lag subsystem

A subsystem (possessing one or more levels) that produces a phase lag between its input and its output, so that when the input reaches its steady-state value, the output reaches its steady-state value only later.

### Loop formed of phase-lag subsystems

Several phase-lag subsystems, each of whose inputs is the output of another phase-lag subsystem, with each output being used as input only once.

### Disturbance

A condition where one or more levels in a system are different from their respective steady-state values. The steady-state values can be either constant, or varying as a function of another mode of behavior, such as exponential growth, or oscillations of a longer period.

Disturbance propagating around a loop formed of phase-lag subsystems

Behavior in which a disturbance away from steady-state values in one subsystem disturbs the next subsystem around the loop away from its steady-state values, and so on, so that a pattern of disturbances moves around the loop.

Comparable disequilibrium states

Two total states of a system (possibly at different points in time) where the disturbances of each of the levels from their steady-state values are proportional to one another. For systems that go to equilibrium in steady-state, comparable disequilibrium states in state space lie on the same ray from the equilibrium point.

Damping

A type of behavior, where the magnitude of oscillations diminishes over time.

Gain element

A subsystem whose output at any given time depends only on its input at that moment of time.

Effectively-first-order system

A system in which the response time of one level significantly exceeds (perhaps by a factor of ten) those of other phase-lag subsystems in the system, which thus effectively become gain elements with respect to the movements of the remaining level.

Cross-link

A cause-and-effect relationship that connects two subsystems that otherwise would be distant from one another, in terms of the number of intervening levels and other cause-and-effect relationships.

Minor loop

A loop containing only one level. Thus, when one creates a minor loop by adding structure around an existing level, the added structure contains no more levels.

Minor loop with a delay

A loop containing two levels, one in the form of a delay. Thus, when one creates a minor loop with a delay by adding structure around an existing level, the added structure contains one level, in the form of a delay.

Stabilize

To cause a system to exhibit fewer oscillations before essentially reaching equilibrium. To stabilize in this sense is not necessarily to decrease the time constant of the exponential envelope around damped oscillations, if whatever change stabilizes the system also increases the period of oscillation.

## 6.2 SUMMARY OF PRINCIPLES DEVELOPED

Below is the list of principles developed in this thesis. These are the culmination of this thesis: in terms of the filing-system analogy, the principles are the keys to the whole affair, which the reader can take away to use at his or her discretion. The principles will allow the reader to recall relevant examples, both from here in the text and from the reader's own experiences, and to recall them at appropriate times in the course of model formulation, analysis, and testing.

This section is here for the benefit of people who have read the text of the thesis, scrutinized the examples, and become familiar with the concepts. To the extent that the reader does not have this preparation, the principles are not useful: the concepts and the principles are just words, with no experiences behind them to make them useful. If the reader is experienced with the dynamics of oscillatory feedback systems, and has elected to start reading here, at least go back and read the concepts, because they are the terms in which the principles are couched, and their meaning is not always the ordinary English meaning. If the reader is not experienced, go back to the beginning of the thesis and get exposure to the material; in terms of the file system analogy, a filing system is useless if there is nothing to go into the folders.



By now, it is quite possible that the reader, especially an experienced reader, will have internalized the principles, and find them completely obvious, if not old hat. This is a totally acceptable result, and is evidence that the thesis has fulfilled its purpose.. A good principle can seem perfectly obvious, with the only real mystery being why someone did not say it before. Here, then, are the hopefully old hat principles connecting system structure to system behavior:

Origin of oscillations

(Section 2.1)

Oscillations occur when a disturbance propagates around a loop formed of two or more phase-lag subsystems to return the system to a state comparable to the initial disequilibrium state.

Damping

(Section 2.2)

A structural or parametric change produces more damping when it realigns the trajectory toward a future state, comparable to a state on the original trajectory, that corresponds to oscillations of smaller magnitude. The simplest structural change that potentially can produce such a realignment is the addition of a minor negative loop. More complex changes can produce damping as well.

Reduction to an effectively-first-order system (Section 3.1)

An oscillatory system can be made not to oscillate by changing it to an effectively-first-order system, so that when the remaining effective level passes through its equilibrium value, the entire system does so, and no further movement occurs.

Adding cross-links

(Section 3.2)

Adding cross-links between subsystems that reduce the ability of the subsystems to move out-of-phase with one another can reduce the tendency of the system to oscillate.

Adding a minor negative loop with a delay (Section 3.3)

If a minor negative loop with a delay is added around a level already on an oscillatory loop, the added loop forms another pathway through which disturbances in the level can propagate back to the level. When the additional disturbance returns to the level, it moves the level more rapidly to and past its steady-state value, which results in a shorter period and less stable oscillation.

Adding a positive loop (Section 4.1)

Adding a positive loop tends to augment whatever disturbance is present in a system's levels, so that the levels move more slowly toward their steady-state values. The system thus takes longer to complete a cycle in an oscillation, or to smoothly approach a steady-state value.

Adding a minor positive loop with a delay (Section 4.2)

If a minor positive loop with a delay is added around a level already on an oscillatory loop, the added loop forms another pathway through which disturbances in the level can propagate back to the level. When the additional disturbance returns to the level, it retards the movement of the level toward its steady-state value, which results in a longer period and more stable oscillations.

Similar systems with a common random input (Section 5.2)

Two similar oscillatory systems will show significantly-similar time-profiles of response to a common random input when the time it takes for the two oscillations to go from in-phase to out-of-phase equals or exceeds four times the geometric mean of the two damping time constants. In such a case, the initial portion of the response of the two systems to each random impulse will dominate their behavior, since the latter portions are diminished by damping. In the initial portion of the impulse response, the two systems oscillate fundamentally in phase, but with the system with the faster response time (usually implying a shorter period) peaking and troughing ahead of the slower system. Thus, differences in period manifest themselves as phase differences, rather than as a "beat" phenomenon.

### 6.3 CONSEQUENCES OF THESIS

At this point, this thesis is complete. All that remains is to be of assistance in considering what to do about or with the material the reader has acquired in the course of reading this thesis. There seem to be three types of future contribution this thesis can make: changing the reader's ability to relate system structure to system behavior (resulting from just reading the thesis itself), opening new avenues of inquiry into the relation between structure and behavior, and serving as a focal point in developing curriculum materials to teach the relation between structure and behavior.

Fulfilling the Purpose: Value to the Reader. The purpose of this thesis, as stated in Section 1.1, is

to effectively communicate to the audience my present experiences with oscillation and entrainment, which are embodied in a series of principles.

At this point, only the reader can judge whether or not the thesis is useful, by observing whether or not the reader utilizes the experiences embodied by the principles in the course of everyday modeling, either in model formulation, testing, refinement, or policy design. As discussed in Section 1.1, there are four possible ways in which these concepts and principles can be used:

- (1) In some situations, the reader of this thesis should be able to use a principle to arrive quickly at hypotheses about the structural features of a system that do or could cause a given behavior.

- (2) In other situations, the reader of this thesis will use the principles to form hypotheses and expectations about the relation between a system's structure and behavior. When those expectations are not met (and the principle breaks down), the modeler is alerted that there is something unusual in the model's structure and behavior.
- (3) The concepts within the principles add to the vocabulary with which systems can be described. If people share the same descriptive vocabulary (and the similar experiences that underlie the vocabulary), they can communicate with one another more effectively than without the descriptive vocabulary.
- (4) By utilizing principles, the reader of this thesis should become more aware of his or her own half-conscious rules of thumb that relate structure to behavior, to the extent that they can be explicated as principles. This both adds to the body of communicable experiences with systems (embodied in the principle), and creates the opportunity for still less-conscious conceptualizations of experiences to emerge.

These, then, describe one form of contribution available to the reader just by the act of reading this document. As valuable as the present body of principles may be, my experience in developing them was that they stimulated more questions than they answered. The following subsection outlines many of these questions.

Avenues for Further Development of Principles. How does developing principles clear the way for developing still more principles? In terms of the filing-system analogy, developing principles is like filing old experiences neatly away in file drawers, each under its appropriate concept folder. This process brings one squarely up against the remaining materials that cannot (as yet) be handled so neatly.

One batch of material that does not really fit into any present principle concerns a more precise specification of the behavioral consequences of closing a feedback loop. The present body of principles is fairly vague in this area. For example, there is much confusion about positive and negative loops, resulting from the ability of negative loops to show divergent behavior (explosive oscillation) and positive loops to show both smoothly-convergent behavior (when the steady-state open-loop gain is between zero and one) or oscillatory behavior (as in Figure 2-11). Thus, there is still a vagueness in the concepts of positive loop, negative loop, convergent behavior, and divergent behavior. This vagueness need not be present, for it is certainly not present in mathematical treatments of closing a feedback loop. For example, thinking in terms of a Nichols chart, a negative loop is characterized as negative only because its steady-state, open-loop gain is negative. If a negative loop produces divergent oscillations, the sinusoidal disturbance is propagating around the loop at zero phase shift and a gain larger than one. The sinusoidal disturbance adds to itself each cycle, resulting in exponentially-growing magnitude. At that frequency,

the loop must be considered a positive loop. Clearly then, this matter of loop polarity is one area in which clarifying concepts and principles are needed. My intuition is that those concepts and principles will be the rough qualitative equivalent of the quantitative Nichols chart. Those concepts and principles should allow one to answer such questions as: "What does it mean for a signal to propagate around a closed loop at other than zero phase shift? What are the consequences of inserting another subsystem into the loop which is itself resonant, or which produces a phase lead?"\*2\* Indeed, how might one characterize the behavior of a loop composed of phase lead systems, or mixed phase lead and phase lag subsystems? Are there many real examples of such systems? Are there any qualitative features of the open loop characteristics (as shown on Bode plots, for example) that provide an indication of the closed-loop behavior? Is there a good way of intuiting a system's response to driving inputs, given its undriven behavior characteristics?

Other questions about the consequences of closing a loop occur in the time domain: Are there any qualitative features of a system's open-loop step/impulse/initial condition response (which is easy to approximate intuitively) that provide an indication of the

---

\*2\*Recall that the principle on the origin of oscillations (Section 2.1) describes oscillations only in terms of phase-lag subsystems. While this principle is of assistance in considering a great many systems, the neglect of phase lead subsystems is a major lacuna in the conceptual framework established by the principles in this thesis.

system's closed loop response? Are there any simple ways of characterizing the effects of changing a gain or delay time in a loop? Can one state any rule of thumb for the natural period of a higher-order loop, corresponding to the second-order  $P = 2\pi$  (Geometric mean of coupling time constants)?

The phenomenon of entrainment begins to be approached in several places in the thesis: Sections 3.3 and 4.3 describe systems with two loops each of which is capable of independent oscillation, and Sections 5.1 and 5.2 of course describe the probabilistic entrainment of two separate but similar systems. The whole area of adding complex dynamic structure to an already-oscillatory system is filled with uncertainty. Is it significant if the added structure is capable of independent oscillation? What determines how easy it is for added structure to modify the damping or period, possibly pulling the period toward its own resonant frequency? In Section 3.3, activating a loop, which in isolation had a relatively long natural period, actually shortened the period of the total system. Is there some kind of continental divide effect, where if the periods were closer together, activating the loop would lengthen the period?

The three-level systems in Sections 3.3 and 4.2 are quite similar: two second-order loops with one level in common. That structure seems as if it is susceptible to mathematical and conceptual analysis, in a format similar to the treatment of the two separate second-order loops in Chapter 5. The dynamics of the third-order systems from Sections 3.3 and 4.2 could be solved analytically, and then the

natural period and damping of the total system could be computed as a function of the natural period and damping of each of the two loops in isolation. These analytical and numerical results could probably be simplified into a principle far more general than those in Section 3.3. and 4.2, just as the numerical results in Section 5.2 were simplified.\*3\* Another means of approaching the area of entrainment might come from extending the analysis of Chapter 5 to two similar systems subject to sinusoidal inputs: under what set of combinations of natural periods, damping ratios, and driving frequencies will the two systems show approximately entrained steady-state or transient behavior? Once that question is answered, one could proceed to the special case where the two systems are productive sectors in a national economy, the driving input is demand for their respective

-----

\*3\*Many of the questions above may seem to imply that I am advocating doing control theory to System Dynamics models. That implication is not quite accurate. I see the principles that would emerge from answering the questions above as an independently-useful body of knowledge, which does not depend (as control-theoretic knowledge so often does) on performing quantitative calculations for an exactly-specified model structure. I would very much favor, however, concepts and principles that closely relate to control-theoretic concepts and calculation. For example, if a principle could be formulated that describes in the frequency domain the consequences of closing a loop, it will almost surely be the qualitative equivalent of the quantitative description provided by a Nichols chart. Indeed, one way of formulating such a principle might be to attempt to explain in conceptual, intuitive terms why a Nichols chart works. That explanation undoubtedly is already possessed nonconsciously by control engineers experienced with classical techniques; the purpose of developing a principle is to furnish explicit concepts and relationships to make that experience communicable to people without spending ten years doing classical control theory.



products, and the output is orders for factors of production such as capital or labor. Are there conditions under which the two sectors are close in phase with each other, both outputs (orders for factors of production) are in phase with the input (orders for the sector's product), and fluctuations in the output are greater in magnitude than fluctuations of the input? If there were such conditions, and if it could be argued that those conditions typified real economies, then one could go a long way toward explaining entrainment into a business cycle in terms of each sector's in-phase amplification of economic transactions between sectors. Even if these conditions did not obtain, the principles that would fall out of the analysis would illuminate the whole subject of driven oscillations.

An area that might be termed the theory of complex systems is another area in which the present principles expose many additional areas of uncertainty. For example, several of the principles speak of adding a loop around a level and oscillatory or dominant loop. Why should one loop dominate others in the first place? Section 2.1 begins to address this question, by characterizing the dominant oscillatory loop as the path that is in some sense the easiest for a disturbance to traverse. Still, this explanation is less than satisfying, and does not really explain the dominance of a positive loop producing smooth exponential growth at all. Also, most of the principles are predicated on the assumption that the behavior mode of the basic system will not change much as a result of making the structural change. Why should system behavior be so insensitive to structural and parametric changes

in so many cases?\*4\* And is there any way of detecting instances when the system is sensitive to its parameters, or instances when the behavior is not a smooth, continuous function of parameter values?

A remarkable number of System Dynamics models of socioeconomic behavior have a large number of coupled positive and negative loops. A surprising number of these systems are insensitive to most of their parameters, and, around equilibrium at least, show moderately-damped oscillatory behavior. Is there any connection between these facts? It seems like the question of the origin of moderately-damped oscillations might be susceptible to mathematical analysis, similar to the way in which theorems about steady-state undriven responses for (continuous) general nonlinear systems have been proven.\*5\* This is not to say, however, that a theorem about damping could be proven with the same mathematical tools, or even that I have any intuition for how one would go about proving such a result.

The final avenue of investigation is most embarrassing, and therefore last: weakness in the present principles. As has been noted

---

\*4\*The beginnings of an answer are given in Graham, Alan K., "Parameter Formulation and Estimation in System Dynamics Models," System Dynamics Group Working Paper D-2349-1, Alfred P. Sloan School of Management, (Cambridge, Mass.: MIT, 1976), page 29. That explanation, however, requires considerable amplification and support before it can become either convincing or useful.

\*5\*In essence, the results say that a very general class of nonlinear systems must eventually reach a steady state of either equilibrium or behavior that is for, most practical purposes, periodic.

previously, all of the principles concerning structural additions are predicated on the assumption that the additions will only modify the previous behavior mode, and do not create a new behavior mode. Section 4.3 illustrates that creating a new behavior mode is a very real possibility, but none of the principles implies any clear guidelines for when such changes occur.\*6\* Characterizing this area of uncertainty in a more positive light, there are no principles that allow one to identify the parameters and structures at which policy changes could have the most impact, and actually alter the behavior mode of the system.

The other major weakness in the present principles is the vagueness of the principle in Section 3.2 on adding crosslinks. What characterizes a stabilizing cross-link versus a destabilizing cross-link? Will the stabilizing effect of a cross-link depend on the specific parameter values of the original system? Answers to such questions could prove quite useful, especially in industrial modeling, where less-than-richly-interconnected chains of production and distribution are common.

\*6\*Another example of an apparent change in behavior mode occurs in Forrester, Jay W., Industrial Dynamics (Cambridge, Mass.: MIT Press, 1961), Appendix M, in which a single sign change initiates behavior quite different from the original behavior (explosive instead of convergent oscillations with a seventy percent reduction in period.)

Role in Curriculum Development. The conclusion has thus far characterized the desirable consequences of this thesis in terms of the value to a reader, and the opening up of new avenues for further research. There is a final area to discuss, the area of curriculum development and education. The process of educating people in the art and science of System Dynamics is usually long and difficult, due in part to the intuitive nature of virtually every phase of System Dynamics modeling. Very little of the process is mechanical enough that one can get by with just following rules.

Perhaps the most difficult facet of a System Dynamics education is learning to relate system structure to system behavior. Prior to this thesis, there were four developed bodies of knowledge to which one could turn to establish conceptual relationships between system structure and system behavior: First-order systems (as taught in Principles of Systems or Study Notes in System Dynamics), characteristics of complex systems (as enumerated in Urban Dynamics), characteristics of particular models (in Industrial Dynamics or Introduction to Urban Dynamics), and whatever concepts could be gleaned from a computationally-oriented control theory.\*7\*

-----

\*7\*Forrester, J.W., Principles of Systems (Cambridge, Mass.: Wright-Allen Press, 1968). Goodman, M.R., Study Notes in System Dynamics (Cambridge, Mass.: Wright-Allen Press, 1974). Forrester, J.W., Urban Dynamics (Cambridge, Mass.: MIT Press, 1969). Forrester, J.W., Industrial Dynamics (Cambridge, Mass.: MIT Press, 1961). Alfeld, L.A. and A.K. Graham, Introduction to Urban Dynamics (Cambridge, Mass.: Wright-Allen Press, 1976). There are numerous texts on control theory; for one of the best see Bryson, A.E. and Y.C. Ho, Applied Optimal Control (Waltham, Mass.: Ginn and Company, 1969).

Unfortunately, these four bodies do not cover all that one needs to know about structure and behavior to do System Dynamics. Usually, this missing knowledge has had to be accumulated intuitively, by repeated exposure to many, many systems. Principles relating system structure and system behavior (like those developed in this thesis) offer a means of structuring that missing body of knowledge.

More specifically, there are a variety of ways in which principles function as a focal point for designing a curriculum that explicitly and effectively teaches the relationship between structure and behavior. First, because the concepts and principles are explicit, they provide a goal toward which the curriculum can aim. For example, without concepts and principles about oscillation, all one can teach is "more stuff" about oscillation. With the principle on the origin of oscillations (Section 2.1), there is a definite ending-place, the principle itself, toward which each piece of prior material must move. Second, and again because concepts and principles are explicit, they allow the material being taught to be broken down into a step-by-step organization of lectures and exercises. For example, the organization of Section 2.1 demonstrates how a principle can be communicated to an experienced dynamicist; students would of course require additional preparation, in the form of actually working through integration of sinusoids, constructing and manipulating a state-space plot, analyzing computer runs, and so on. With a guiding principle, each of these

activities has a clear role in attaining the educational objectives.

A third way in which principles can aid curriculum organization arises from the generality of well-chosen concepts and principles: they can serve as a bridge between verbal description of specific models and the mathematical machinery of quantitative analysis. For example, the concept (developed from several examples in Section 2.1) of a disturbance propagating around a loop provides an intuitive basis for going on to learn about phase and gain, transfer functions, and so on. Similarly, the concepts involving movements through state space provide students with a simple means of understanding the (potentially very arbitrary and mechanical) manipulations of matrix exponentials and Liapunov stability theory.

A fourth and final way in which principles can serve as focal points for curriculum development comes about because principles are a simple, compact way of summarizing and "carrying away" a considerable amount of learning: principles can integrate both specific case studies and technical knowledge into the mainstream of System Dynamics modeling. Without principles relating system structure to system behavior, there is little likelihood that a student can detect opportunities for appropriately using mathematical analysis, or making an analogy to a system previously studied. A principle causes one to analyze the system with concepts that naturally lead to the mathematical tools or previous examples needed. Thus, for example, after studying a unit on oscillations, the student can work subsequent exercises using the

principles that emerged from the unit, and carrying the analysis into computation of gain, phase shift, eigenvalues when appropriate.

In brief then, principles relating system structure to system behavior serve as focal points for developing educational goals, structuring lectures, exercises, cases, and technical material to achieve those goals, and continually integrating the material learned into the ongoing practice of System Dynamics.

APPENDIX A

DOCUMENTED EQUATIONS FOR THE SPRING-MASS SYSTEM

$P.K = P.J + (DT) (V.J)$	1, L
$P = IP$	1.1, N
$IP = .5$	1.2, C
P - POSITION (FEET)	
V - VELOCITY (FEET/SECOND)	
IP - INITIAL POSITION (FEET)	
$V.K = V.J + (DT) (A.JK)$	2, L
$V = IV$	2.1, N
$IV = 0$	2.2, C
V - VELOCITY (FEET/SECOND)	
A - ACCELERATION (FEET/SECOND/SECOND)	
IV - INITIAL VELOCITY (FEET/SECOND)	
$A.KL = (FS.K + FF.K + FD.K) / (W/G)$	3, R
$W = 160$	3.1, C
$G = 32$	3.2, C
A - ACCELERATION (FEET/SECOND/SECOND)	
FS - FORCE FROM SPRING (POUNDS)	
FF - FORCE FROM FRICTION (POUNDS)	
FD - FORCE FROM DRIVING (POUNDS)	
W - WEIGHT (POUNDS)	
G - GRAVITATIONAL ACCELERATION (FEET/SECOND/SECOND)	
$FS.K = -SC * P.K$	4, A
$SC = 5$	4.1, C
FS - FORCE FROM SPRING (POUNDS)	
SC - SPRING CONSTANT (POUNDS/FOOT)	
P - POSITION (FEET)	
$FF.K = -FC * V.K$	5, A
$FC = 0$	5.1, C
FF - FORCE FROM FRICTION (POUNDS)	
FC - FRICTION COEFFICIENT (POUNDS/FOOT/SECOND)	
V - VELOCITY (FEET/SECOND)	
$FD.K = DC * V.K$	6, A
$DC = 0$	6.1, C
FD - FORCE FROM DRIVING (POUNDS)	
DC - DRIVING CONSTANT (POUNDS/FOOT/SECOND)	
V - VELOCITY (FEET/SECOND)	



APPENDIX B

ANALYZER LISTING FOR THE SPRING-MASS SYSTEM

NAME	NO	T	DEFINITION
A	3	R	ACCELERATION (FEET/SECOND/SECOND)
	V,L,2/PLOT,6.3/SAVE,6.4		
DC	6.1	C	DRIVING CONSTANT (POUNDS/FOOT/SECOND)
	FD,A,6		
FC	5.1	C	FRICTION COEFFICIENT (POUNDS/FOOT/SECOND)
	FF,A,5		
FD	6	A	FORCE FROM DRIVING (POUNDS)
	A,R,3		
FF	5	A	FORCE FROM FRICTION (POUNDS)
	A,R,3		
FS	4	A	FORCE FROM SPRING (POUNDS)
	A,R,3		
G	3.2	C	GRAVITATIONAL ACCELERATION (FEET/SECOND/SECOND)
	A,R,3		
IP	1.2	C	INITIAL POSITION (FEET)
	P,N,1.1		
IV	2.2	C	INITIAL VELOCITY (FEET/SECOND)
	V,N,2.1		
LENGTH			
	SPEC,6.5		
P	1	L	POSITION (FEET)
	1.1 N		
	FS,A,4/PLOT,6.3/SAVE,6.4		
PLTPER			
	SPEC,6.5		
SAVPER			
	SPEC,6.5		
SC	4.1	C	SPRING CONSTANT (POUNDS/FOOT)
	FS,A,4		
V	2	L	VELOCITY (FEET/SECOND)
	2.1 N		
	P,L,1/FF,A,5/FD,A,6/PLOT,6.3/SAVE,6.4		
W	3.1	C	WEIGHT (POUNDS)
	A,R,3		

APPENDIX C

EQUATION LISTING FOR THE SPRING-MASS SYSTEM

```
SM.DYNAMO
00001 *    SPRING-MASS SYSTEM
00010 L    P.K=P.J+(DT) (V.J)
00011 N    P=IP
00012 C    IP=.5
00020 L    V.K=V.J+(DT) (A.JK)
00021 N    V=IV
00022 C    IV=0
00030 R    A.KL=(FS.K+FF.K+FD.K)/(W/G)
00031 C    W=160
00032 C    G=32
00040 A    FS.K=-SC*P.K
00041 C    SC=5
00050 A    FF.K=-FC*V.K
00051 C    FC=0
00052 NOTE FRICITION COEFFICIENT FC = 2.5 FOR DAMPING
00060 A    FD.K=DC*V.K
00061 C    DC=0
00062 NOTE DRIVING CONSTANT DC = 3 FOR DRIVING
00063 PLOT P=P/V=V/A=A
00064 SAVE P,V,A
00065 SPEC DT=.01/LENGTH=10/PLTPER=.2/SAVPER=.2
00066 RUN  FIGURE 2-3
00067 C    FC=2.5
00068 CPLOT P.BASE,P.FRIC,A.BASE,A.FRIC,V.BASE,V.FRIC
00069 C    PRTPER=.4
00071 RUN  FIGURES 2-19, 2-20
00072 C    FC=25
00073 PLOT P=P,V=V,A=A
00074 RUN  FIGURE 3-10
00075 C    DC=3
00076 PLOT P=P/V=V/A=A
00077 RUN  FIGURE 4-4
```

APPENDIX D

DOCUMENTED EQUATIONS FOR THE EMPLOYMENT-BACKLOG SYSTEM

MACRO PKNS (MNSE, SDN, TCN) .2

PKNS - COLORED (PINK) NOISE FUNCTION  
MNSE - MEAN OF NOISE (DIMENSIONLESS)  
SDN - STANDARD DEVIATION OF NOISE (DIMENSIONLESS)  
TCN - TIME TO CORRELATE NOISE (YEARS)

PKNS.K=PKNS.J+(DT/TCN) (\$SFN\*NOISE()+MNSE-PKNS.J) 1, L  
PKNS=MNSE 1.1, N  
\$SFN=SDN\*SQRT(24\*TCN/DT) 1.2, N

PKNS - COLORED (PINK) NOISE FUNCTION  
DT - COMPUTATION INTERVAL (YEARS)  
TCN - TIME TO CORRELATE NOISE (YEARS)  
\$SFN - SCALING FACTOR FOR NOISE (DIMENSIONLESS)  
MNSE - MEAN OF NOISE (DIMENSIONLESS)  
SDN - STANDARD DEVIATION OF NOISE (DIMENSIONLESS)

MEND

E.K=E.J+(DT) (NHR.JK) 1, L  
E=EN 1.1, N  
EN=50 1.2, C

E - EMPLOYMENT (MEN)  
DT - COMPUTATION INTERVAL (YEARS)  
NHR - NET HIRING RATE (MEN/YEAR)  
EN - EMPLOYMENT INITIAL (MEN)

NHR.KL=(DE.K-E.K)/TCE 2, R  
TCE=.5 2.1, C

NHR - NET HIRING RATE (MEN/YEAR)  
DE - DESIRED EMPLOYMENT (MEN)  
E - EMPLOYMENT (MEN)  
TCE - TIME TO CORRECT EMPLOYMENT (YEARS)

DE.K=DOUT.K/PROD 3, A  
PROD=30 3.1, C

DE - DESIRED EMPLOYMENT (MEN)  
DOUT - DESIRED OUTPUT (UNITS/YEAR)  
PROD - PRODUCTIVITY (UNITS/YEAR/MAN)

$DOUT.K = EAO.K + CB.K$	4, A
DOUT - DESIRED OUTPUT (UNITS/YEAR)	
EAO - EXPECTED AVERAGE OUTPUT (UNITS/YEAR)	
CB - CORRECTION FOR BACKLOG (UNITS/YEAR)	
$CB.K = (B.K - DB.K) / TCB$	5, A
$TCB = .5$	5.1, C
CB - CORRECTION FOR BACKLOG (UNITS/YEAR)	
B - BACKLOG (UNITS)	
DB - DESIRED BACKLOG (UNITS)	
TCB - TIME TO CORRECT BACKLOG (YEARS)	
$DB.K = EAO.K * DBC$	6, A
$DBC = .5$	6.1, C
DB - DESIRED BACKLOG (UNITS)	
EAO - EXPECTED AVERAGE OUTPUT (UNITS/YEAR)	
DBC - DESIRED BACKLOG COVERAGE (YEARS)	
$B.K = B.J + (DT) (OR.JK - OUT.JK)$	7, L
$B = BN$	7.1, N
$BN = 660$	7.2, C
B - BACKLOG (UNITS)	
DT - COMPUTATION INTERVAL (YEARS)	
OR - ORDER RATE (UNITS/YEAR)	
OUT - OUTPUT (UNITS/YEAR)	
BN - BACKLOG INITIAL (UNITS)	
$OUT.KL = E.K * PROD * (1 - WTO) + TOUT.K * WTO$	8, R
$WTO = 0$	8.1, C
OUT - OUTPUT (UNITS/YEAR)	
E - EMPLOYMENT (MEN)	
PROD - PRODUCTIVITY (UNITS/YEAR/MAN)	
WTO - WEIGHTING ON TEST OUTPUT (DIMENSIONLESS)	
TOUT - TEST OUTPUT (UNITS/YEAR)	
$OR.KL = (AOR.K) (MDO.K)$	9, R
OR - ORDER RATE (UNITS/YEAR)	
AOR - ADJUSTED ORDER RATE (UNITS/YEAR)	
MDO - MULTIPLIER FROM DELAY ON ORDERING (DIMENSIONLESS)	
$AOR.K = DOR.K + CBO.K$	10, A
AOR - ADJUSTED ORDER RATE (UNITS/YEAR)	
DOR - DESIRED ORDER RATE (UNITS/YEAR)	
CBO - CORRECTION FOR BACKLOG ON ORDERING (UNITS/ YEAR)	

CBO.K=(DBO.K-B.K)/TCBO 11, A  
TCBO=1E11 11.1, C  
CBO - CORRECTION FOR BACKLOG ON ORDERING (UNITS/  
YEAR)  
DBO - DESIRED BACKLOG FOR ORDERING (UNITS)  
B - BACKLOG (UNITS)  
TCBO - TIME TO CORRECT BACKLOG FOR ORDERING  
(YEARS)

DBO.K=PDBA.K\*DOR.K 12, A  
DBO - DESIRED BACKLOG FOR ORDERING (UNITS)  
PDBA - PERCEIVED DELAY FOR BACKLOG ADJUSTMENT  
(YEARS)  
DOR - DESIRED ORDER RATE (UNITS/YEAR)

DOR.K=1320\*PKNS(1,SDN,TCN) 13, A  
SDN=0 13.1, C  
TCN=.7 13.2, C  
DOR - DESIRED ORDER RATE (UNITS/YEAR)  
PKNS - COLORED (PINK) NOISE FUNCTION  
SDN - STANDARD DEVIATION OF NOISE (DIMENSIONLESS)  
TCN - TIME TO CORRELATE NOISE (YEARS)

MDO.K=TABLE(TMDO,PDO.K/DBC,0,4,.5) 14, A  
TMDO=1.3/1.2/1/.75/.5/.3/.2/.15/.1 14.1, T  
MDO - MULTIPLIER FROM DELAY ON ORDERING  
(DIMENSIONLESS)  
TMDO - TABLE FOR MULTIPLIER FROM DELAY ON ORDERING  
PDO - PERCEIVED DELAY FOR ORDERING (YEARS)  
DBC - DESIRED BACKLOG COVERAGE (YEARS)

PDO.K=PDO.J+(DT/TPDO)(DD.J-PDO.J) 15, L  
PDO=DBC 15.1, N  
TPDO=1E11 15.2, C  
PDO - PERCEIVED DELAY FOR ORDERING (YEARS)  
DT - COMPUTATION INTERVAL (YEARS)  
TPDO - TIME TO PERCEIVE DELAY FOR ORDERING  
DD - DELIVERY DELAY (YEARS)  
DBC - DESIRED BACKLOG COVERAGE (YEARS)

DD.K=B.K/OUT.JK 16, A  
DD - DELIVERY DELAY (YEARS)  
B - BACKLOG (UNITS)  
OUT - OUTPUT (UNITS/YEAR)

TOUT.K=1320+STEP(TOSH,TOST)+SINA\*SIN((6.28/SINP)\*  
TIME.K) 17, A  
TOSH=0 17.2, C  
TOST=1 17.3, C  
SINA=0 17.4, C  
SINP=3 17.5, C  
TOUT - TEST OUTPUT (UNITS/YEAR)  
TOSH - TEST OUTPUT STEP HEIGHT (UNITS/YEAR)  
TOST - TEST OUTPUT STEP TIME (YEAR)  
SINA - SINE AMPLITUDE (UNITS/YEAR)  
SINP - SINE PERIOD (YEARS)  
TIME - ELAPSED TIME (YEARS)

EAO.K=EAO.J+(DT/TAO)(OUT.JK-EAO.J) 18, L  
EAO=DOR 18.1, N  
TAO=1E11 18.2, C  
EAO - EXPECTED AVERAGE OUTPUT (UNITS/YEAR)  
DT - COMPUTATION INTERVAL (YEARS)  
TAO - TIME TO AVERAGE OUTPUT (YEARS)  
OUT - OUTPUT (UNITS/YEAR)  
DOR - DESIRED ORDER RATE (UNITS/YEAR)

PDBA.K=PDBA.J+(DT/TPDBA)(DD.J-PDBA.J) 19, L  
PDBA=DBC 19.1, N  
TPDBA=1E11 19.2, C  
PDBA - PERCEIVED DELAY FOR BACKLOG ADJUSTMENT  
(YEARS)  
DT - COMPUTATION INTERVAL (YEARS)  
TPDBA - TIME TO PERCEIVE DELAY FOR BACKLOG  
ADJUSTMENT (YEARS)  
DD - DELIVERY DELAY (YEARS)  
DBC - DESIRED BACKLOG COVERAGE (YEARS)

FLTPER.K=CLIP(PLTINT,0,TIME.K,PLTST) 20, A  
PLTINT=.2 20.1, C  
PLTST=-1 20.2, C  
FLTPER - PLOT PERIOD (YEARS)  
PLTINT - PLOTTING INTERVAL (YEARS)  
TIME - ELAPSED TIME (YEARS)  
PLTST - PLOT STARTING TIME (YEAR)

PRTPER.K=CLIP(0,PRTINT,TIME.K,PRTKT) 21, A  
PRTINT=0 21.1, C  
PRTKT=1E11 21.2, C  
DT=.05 21.3, C  
LENGTH=10 21.4, C  
    PRTPER - PRINTING PERIOD (YEARS)  
    PRTINT - PRINTING INTERVAL (YEARS)  
    TIME - ELAPSED TIME (YEARS)  
    PRTKT - PRINTING KILL TIME (YEAR)  
    DT - COMPUTATION INTERVAL (YEARS)  
    LENGTH - LENGTH OF SIMULATION (YEARS)

SAVE B,E,DOR 21.5  
    B - BACKLOG (UNITS)  
    E - EMPLOYMENT (MEN)  
    DOR - DESIRED ORDER RATE (UNITS/YEAR)

PLOT DB=D,B=B/NHR=N/E=E,DE=J/OUT=O,OR=R,DOUT=P, 21.6  
    EAO=V  
    DB - DESIRED BACKLOG (UNITS)  
    B - BACKLOG (UNITS)  
    NHR - NET HIRING RATE (MEN/YEAR)  
    E - EMPLOYMENT (MEN)  
    DE - DESIRED EMPLOYMENT (MEN)  
    OUT - OUTPUT (UNITS/YEAR)  
    OR - ORDER RATE (UNITS/YEAR)  
    DOUT - DESIRED OUTPUT (UNITS/YEAR)  
    EAO - EXPECTED AVERAGE OUTPUT (UNITS/YEAR)

PLOT DB=D,B=B/NHR=N/E=E,DE=J/OUT=O,OR=R,DOUT=P, 21.7  
    EAO=V  
    DB - DESIRED BACKLOG (UNITS)  
    B - BACKLOG (UNITS)  
    NHR - NET HIRING RATE (MEN/YEAR)  
    E - EMPLOYMENT (MEN)  
    DE - DESIRED EMPLOYMENT (MEN)  
    OUT - OUTPUT (UNITS/YEAR)  
    OR - ORDER RATE (UNITS/YEAR)  
    DOUT - DESIRED OUTPUT (UNITS/YEAR)  
    EAO - EXPECTED AVERAGE OUTPUT (UNITS/YEAR)

PRINT E,B 21.8  
    E - EMPLOYMENT (MEN)  
    B - BACKLOG (UNITS)

APPENDIX E

ANALYZER LISTING FOR THE EMPLOYMENT-BACKLOG SYSTEM

NAME	NO	DEFINITION	WHERE USED
AOR	10	A ADJUSTED ORDER RATE (UNITS/YEAR)	OR,R,9
B	7	I BACKLOG (UNITS)	7.1 N
BN	7.2	C BACKLOG INITIAL (UNITS)	B,N,7.1
CB	5	A CORRECTION FOR BACKLOG (UNITS/YEAR)	DOUT,A,4
CBO	11	A CORRECTION FOR BACKLOG ON ORDERING (UNITS/YEAR)	AOR,A,10
DB	6	A DESIRED BACKLOG (UNITS)	CB,A,5/PLOT,21.6/PLOT,21.7
DBC	6.1	C DESIRED BACKLOG COVERAGE (YEARS)	DB,A,6/MDO,A,14/PDO,N,15.1/PDBA,N,19.1
DBO	12	A DESIRED BACKLOG FOR ORDERING (UNITS)	CBO,A,11
DD	16	A DELIVERY DELAY (YEARS)	PDO,T,15/PDBA,L,19
DE	3	A DESIRED EMPLOYMENT (MEN)	NHR,R,2/PLOT,21.6/PLOT,21.7
DOR	13	A DESIRED ORDER RATE (UNITS/YEAR)	AOR,A,10/DBO,A,12/EAO,N,18.1/SAVE,21.5
DOUT	4	A DESIRED OUTPUT (UNITS/YEAR)	DE,A,3/PLOT,21.6/PLOT,21.7
DT	21.3	C COMPUTATION INTERVAL (YEARS)	PKNS,T,1/SSFN,N,1.2/E,L,1/B,L,7/PDO,T,15/EAO,T,18/PDBA,L,19
E	1	I EMPLOYMENT (MEN)	1.1 N
EAO	18	I EXPECTED AVERAGE OUTPUT (UNITS/YEAR)	18.1 N
EN	1.2	C EMPLOYMENT INITIAL (MEN)	DOUT,A,4/DB,A,6/PLOT,21.6/PLOT,21.7



LENGTH 21.4 C LENGTH OF SIMULATION (YEARS)  
MDO 14 A MULTIPLIER FROM DELAY ON ORDERING  
(DIMENSIONLESS)  
OR,R,9  
MNSE MEAN OF NOISE (DIMENSIONLESS)  
MACRO,.2/PKNS,L,1/PKNS,N,1.1  
NHR 2 R NET HIRING RATE (MEN/YEAR)  
E,L,1/PLOT,21.6/PLOT,21.7  
OR 9 R ORDER RATE (UNITS/YEAR)  
B,L,7/PLOT,21.6/PLOT,21.7  
OUT 8 R OUTPUT (UNITS/YEAR)  
B,L,7/DD,A,16/EAO,L,18/PLOT,21.6/PLOT,21.7  
PDBA 19 L PERCEIVED DELAY FOR BACKLOG ADJUSTMENT  
19.1 N (YEARS)  
DBO,A,12  
PDO 15 L PERCEIVED DELAY FOR ORDERING (YEARS)  
15.1 N  
MDO,A,14  
PKNS 1 L COLORED (PINK) NOISE FUNCTION  
1.1 N  
MACRO,.2/DOR,A,13  
PLTINT 20.1 C PLOTTING INTERVAL (YEARS)  
PLTPER,A,20  
PLTPER 20 A PLOT PERIOD (YEARS)  
PLTST 20.2 C PLOT STARTING TIME (YEAR)  
PLTPER,A,20  
PROD 3.1 C PRODUCTIVITY (UNITS/YEAR/MAN)  
DE,A,3/OUT,R,8  
PRTINT 21.1 C PRINTING INTERVAL (YEARS)  
PRTPER,A,21  
PRTKT 21.2 C PRINTING KILL TIME (YEAR)  
PRTPER,A,21  
PRTPER 21 A PRINTING PERIOD (YEARS)  
SDN 13.1 C STANDARD DEVIATION OF NOISE (DIMENSIONLESS)  
MACRO,.2/\$SFN,N,1.2/DOR,A,13  
SINA 17.4 C SINE AMPLITUDE (UNITS/YEAR)  
TOUT,A,17  
SINP 17.5 C SINE PERIOD (YEARS)  
TOUT,A,17  
TAO 18.2 C TIME TO AVERAGE OUTPUT (YEARS)  
EAO,L,18  
TCB 5.1 C TIME TO CORRECT BACKLOG (YEARS)  
CB,A,5  
TCBO 11.1 C TIME TO CORRECT BACKLOG FOR ORDERING  
(YEARS)  
CBO,A,11

TCE 2.1 C TIME TO CORRECT EMPLOYMENT (YEARS)  
NHR,R,2

TCN 13.2 C TIME TO CORRELATE NOISE (YEARS)  
MACRO,.2/PKNS,L,1/\$SFN,N,1.2/DOR,A,13

TIME ELAPSED TIME (YEARS)  
TOUT,A,17/PLTPER,A,20/PRTPER,A,21

TMDO 14.1 T TABLE FOR MULTIPLIER FROM DELAY ON ORDERING  
MDO,A,14

TOSH 17.2 C TEST OUTPUT STEP HEIGHT (UNITS/YEAR)  
TOUT,A,17

TOST 17.3 C TEST OUTPUT STEP TIME (YEAR)  
TOUT,A,17

TOUT 17 A TEST OUTPUT (UNITS/YEAR)  
OUT,R,8

TPDBA 19.2 C TIME TO PERCEIVE DELAY FOR BACKLOG  
ADJUSTMENT (YEARS)  
PDBA,L,19

TPDO 15.2 C TIME TO PERCEIVE DELAY FOR ORDERING  
PDO,L,15

WTO 8.1 C WEIGHTING ON TEST OUTPUT (DIMENSIONLESS)  
OUT,R,8

\$SFN 1.2 N SCALING FACTOR FOR NOISE (DIMENSIONLESS)  
PKNS,L,1

APPENDIX F

EQUATION LISTING FOR EMPLOYMENT-BACKLOG SYSTEM

EB.DYNAMO  
00001 \* EMPLOYMENT-BACKLOG SYSTEM  
00002 MACRO PKNS (MNSE,SDN,TCN)  
00010 L  $PKNS.K = PKNS.J + (DT/TCN) (\$SFN * NOISE() + MNSE - PKNS.J)$   
00011 N PKNS=MNSE  
00012 N  $\$SFN = SDN * SQRT(24 * TCN / DT)$   
00013 MEND  
10010 L  $E.K = E.J + (DT) (NHR.JK)$   
10011 N E=EN  
10012 C EN=50  
10020 R  $NHR.KL = (DE.K - E.K) / TCE$   
10021 C TCE=.5  
10030 A  $DE.K = DOUT.K / PROD$   
10031 C PROD=30  
10040 A  $DOUT.K = EAO.K + CB.K$   
10050 A  $CB.K = (B.K - DB.K) / TCB$   
10051 C TCB=.5  
10060 A  $DB.K = EAO.K * DBC$   
10061 C DBC=.5  
10070 L  $B.K = B.J + (DT) (OR.JK - OUT.JK)$   
10071 N B=BN  
10072 C BN=660  
10080 R  $OUT.KL = E.K * PROD * (1 - WTO) + TOUT.K * WTO$   
10081 C WTO=0  
10090 R  $OR.KL = (AOR.K) (MDO.K)$   
10100 A  $AOR.K = DOR.K + CBO.K$   
10110 A  $CBO.K = (DBO.K - B.K) / TCBO$   
10111 C TCBO=1E11  
10112 NOTE TCBO=.3 FOR BACKLOG CORRECTION  
10120 A  $DBO.K = PDBA.K * DOR.K$   
10130 A  $DOR.K = 1320 * PKNS(1, SDN, TCN)$   
10131 C SDN=0  
10132 C TCN=.7  
10140 A  $MDO.K = TABLE(TMDO, PDO.K / DBC, 0, 4, .5)$   
10141 T  $TMDO = 1.3 / 1.2 / 1 / .75 / .5 / .3 / .2 / .15 / .1$   
10150 L  $PDO.K = PDO.J + (DT / TPDO) (DD.J - PDO.J)$   
10151 N PDO=DBC  
10152 C TPDO=1E11  
10153 NOTE TPDO=2 FOR MDO RESPONDING TO DELIVERY DELAY

```
10160 A      DD.K=B.K/OUT.JK
10170 A      TOUT.K=1320+STEP(TOSH,TOST)+SINA*SIN(
10171 X          (6.28/SINP)*TIME.K)
10172 C      TOSH=0
10173 C      TOST=1
10174 C      SINA=0
10175 C      SINP=3
10180 L      EAO.K=EAO.J+(DT/TAO)(OUT.JK-EAO.J)
10181 N      EAO=DOR
10182 C      TAO=1E11
10183 NOTE   TAO=1 FOR FLOATING GOAL
10190 L      PDBA.K=PDBA.J+(DT/TPDBA)(DD.J-PDBA.J)
10191 N      PDBA=DBC
10192 C      TPDBA=1E11
10193 NOTE   TPDBA=1 FOR DESIRED BACKLOG RESPONDING TO DELIVERY
10194 NOTE   DELAY
10195 NOISE  1234567
10200 A      PLTPER.K=CLIP(PLTINT,0,TIME.K,PLTST)
10201 C      PLTINT=.2
10202 C      PLTST=-1
10210 A      PRTPER.K=CLIP(0,PRTINT,TIME.K,PRTKT)
10211 C      PRTINT=0
10212 C      PRTKT=1E11
10213 C      DT=.05
10214 C      LENGTH=10
10215 SAVE  B,E,DOR
10216 PLOT  DB=D,B=B/NHR=N/E=E,DE=J/OUT=O,OR=R,DOUT=P,EAO=V
10217 PLOT  DB=D,B=B/NHR=N/E=E,DE=J/OUT=O,OR=R,DOUT=P,EAO=V
10218 PRINT E,B
10219 RUN   FIGURES 2-14, 3-21
10221 C    PRTINT=.2
10222 RUN  FIGURE 2-15
10223 C    TCBO=.3
10224 C    PRTINT=.2
10225 PLOT  DBO=D,B=B/CBO=C/E=E,DE=J/OR=R,OUT=O,DOUT=P,EAO=V
10226 PLOT  DBO=D,B=B/CBO=C/E=E,DE=J/OR=R,OUT=O,DOUT=P,
10227 X      EAO=V(1300,1340)
10228 RUN  FIGURES 2-24,2-25,4-12
10229 C    TCE=.1
10231 PLOT  DB=D,B=B/CB=I/E=E,DE=J/OUT=O,OR=R,DOUT=P,EAO=V
10232 C    PRTINT=.05
10233 C    PRTKT=2
10234 RUN  FIGURES 3-1,3-3
10235 C    TCB=.1
10236 PLOT  DB=D,B=B/CB=I/E=E,DE=J/OUT=O,OR=R,DOUT=P,EAO=V
10237 RUN  FIGURE 3-2
10238 PLOT  DB=D,B=B/NHR=N/E=E,DE=J/OUT=O,OR=R,DOUT=P/MDO=M/DD
10239 C    TPDO=2
10241 RUN  FIGURE 3-20
```

10242 C WTO=1  
10243 C TOSH=132  
10244 C SAVPER=.2  
10245 C PLTINT=0  
10246 RUN OPEN  
10247 C TPDO=2  
10248 C WTO=1  
10249 C TOSH=132  
10251 CPLOT OUT=O,OR=R/DD=D,PDO=P/MDO=M/B=B,DBO=S,B.OPEN=E(300,700  
10252 RUN FIGURE 3-22  
10253 C PLTINT=.1  
10254 C PLTST=17.99  
10255 C LENGTH=23  
10256 C WTO=1  
10257 C SINA=132  
10258 C TPDO=2  
10259 PLOT OUT=O,OR=R/DD=D,PDO=P/MDO=M/B=B,DBO=S  
10261 RUN FIGURE 3-23  
10262 C PLTINT=.1  
10263 C PLTST=17.99  
10264 C LENGTH=23  
10265 C WTO=1  
10266 C SINA=132  
10267 C TPDO=1  
10268 PLOT OUT=O,OR=R/DD=D,PDO=P/MDO=M/B=B,DBO=S  
10269 RUN FIGURE 3-24  
10271 PLOT DB=D,B=B/NHR=N/E=E,DE=J/OUT=O,OR=R,DOUT=P,EAO=V  
10272 C PLTST=1.999  
10273 C LENGTH=4.5  
10274 C PLTINT=.05  
10275 RUN FIGURE 3-25  
10276 C TPDBA=1  
10277 C TCBO=.3  
10278 PLOT DBO=D,B=B/CBO=C,CB=I/PDBA=V,DD=L/E=E/OUT=O,OR=R,DOUT=P  
10279 RUN FIGURE 4-11  
10281 C TCB=1.5  
10282 PLOT DB=D,B=B/CBO=C,CB=I/E=E,DE=J/OUT=O,OR=R,DOUT=P,EAO=V  
10283 RUN FIGURE 4-16  
10284 C TCB=1.5  
10285 C TAO=.5  
10286 PLOT DB=D,B=B/CBO=C,CB=I/E=E,DE=J/OUT=O,OR=R,DOUT=P,EAO=V  
10287 RUN FIGURE 4-17  
10288 C TCB=1.5  
10289 C TAO=.5  
10291 C TCE=.1  
10292 PLOT DB=D,B=B/CBO=C,CB=I/E=E,DE=J/OUT=O,OR=R,DOUT=P,  
10293 X EAO=V(1300,1340)  
10294 RUN FIGURE 4-18

```
10295 C SAVPER=.2
10296 C PLTINT=0
10297 RUN BASE
10298 C TCB=1
10299 C PLTINT=.2
10301 CPLOT E.BASE,E/B.BASE,B/DOR.BASE,DOR(1000,2600)
10302 RUN FIGURE 5-3
10303 C SAVPER=.4
10304 C PLTINT=0
10305 C SDN=.1
10306 C EN=44
10307 C LENGTH=20
10308 RUN BASE
10309 C TCB=1
10311 C PLTINT=.4
10312 C SDN=.1
10313 C EN=44
10314 C LENGTH=20
10315 CPLOT E.BASE,E/B.BASE,B/DOR.BASE,DOR(1000,2600)
10316 RUN FIGURE 5-4
```

APPENDIX G

EQUATION LISTING FOR NOFPER

NOFPER.DYNAMO

00010 \*       ENTRAINMENT MEASURES  
00020 A       L2RP.K=IL2RP+TIME.K  
00030 C       IL2RP=-2  
00040 A       L2P1.K=L2RP.K+L2P2  
00050 C       L2P2=2  
00060 A       P2.K=4  
00070 A       P1.K=EXP(.693\*L2P1.K)  
00080 A       A.K=- (39.478/(P1.K\*P1.K)) +C.K\*C.K/4  
00090 A       B.K=- (39.479/(P2.K\*P2.K)) +D.K\*D.K/4  
00100 A       C.K=-2/DTC1  
00110 C       DTC1=1  
00120 A       D.K=-2/DTC2  
00130 C       DTC2=1  
00140 A       AL.K=Q/(2\*A.K\*C.K)  
00150 C       Q=1  
00160 A       BE.K=-Q/(2\*A.K)  
00170 A       GA.K=(Q) ((C.K\*C.K-A.K)/(2\*A.K\*C.K))  
00180 A       ET.K=Q/(2\*B.K\*D.K)  
00190 A       LA.K=(Q) ((D.K\*D.K-B.K)/(2\*B.K\*D.K))  
00200 A       SI.K=-Q/(2\*B.K)  
00210 A       DE.K=Q\*(C.K+D.K)/K.K  
00220 A       XI.K=-Q\*(B.K-A.K+C.K\*D.K+D.K\*D.K)/K.K  
00230 A       K.K=(A.K\*D.K+B.K\*C.K)(C.K+D.K)-(A.K-B.K)(A.K-B.K)  
00240 A       RH.K=Q\*(B.K-A.K-C.K\*D.K-C.K\*C.K)/K.K  
00250 A       NU.K=Q\*(C.K\*C.K\*D.K-A.K\*C.K+C.K\*D.K\*D.K-B.K\*D.K)/K.K  
00260 A       ID.K=1/(AL.K\*GA.K-BE.K\*BE.K)  
00270 A       T11.K=ID.K\*(DE.K\*GA.K-RH.K\*BE.K)  
00280 A       T12.K=ID.K\*(-DE.K\*BE.K+RH.K\*AL.K)  
00290 A       T21.K=ID.K\*(XI.K\*GA.K-NU.K\*BE.K)  
00300 A       T22.K=ID.K\*(-XI.K\*BE.K+NU.K\*AL.K)  
00310 A       TTH1.K=T21.K/T11.K  
00320 A       TTH2.K=-T12.K/T22.K  
00330 A       GT11.K=ET.K-AL.K\*T11.K\*T11.K-GA.K\*T12.K\*T12.K-  
00340 X       2\*BE.K\*T11.K\*T12.K  
00350 A       GT22.K=LA.K-AL.K\*T21.K\*T21.K-GA.K\*T22.K\*T22.K-  
00360 X       2\*BE.K\*T21.K\*T22.K  
00370 A       GT12.K=SI.K-AL.K\*T11.K\*T21.K-BE.K\*  
00380 X       (T21.K\*T12.K+T11.K\*T22.K)-GA.K\*T22.K\*T12.K  
00390 A       GT21.K=GT12.K

```
00400 A      GTIDET.K=MAX(1E-15,GT11.K*GT22.K-GT12.K*GT21.K)
00410 A      GTI11.K=GT22.K/GTIDET.K
00420 A      GTI12.K=-GT12.K/GTIDET.K
00430 A      GTI21.K=-GT21.K/GTIDET.K
00440 A      GTI22.K=GT11.K/GTIDET.K
00450 A      TERM3.K=T11.K*GTI11.K+T21.K*GTI21.K
00460 A      TERM4.K=T11.K*GTI12.K+T21.K*GTI22.K
00470 A      TERM5.K=T12.K*GTI11.K+T22.K*GTI21.K
00480 A      TERM6.K=T12.K*GTI12.K+T22.K*GTI22.K
00490 A      M11.K=TERM3.K*T11.K+TERM4.K*T21.K
00500 A      M12.K=TERM3.K*T12.K+TERM4.K*T22.K
00510 A      M21.K=TERM5.K*T11.K+TERM6.K*T21.K
00520 A      M22.K=TERM5.K*T12.K+TERM6.K*T22.K
00530 A      N.K=M11.K*AL.K+M12.K*BE.K+M21.K*BE.K+M22.K*GA.K
00540 A      NH.K=KNH*TOP.K/GMDTC
00550 C      KNH=.917
00560 N      GMDTC=SQRT(DTC1*DTC2)
00570 A      TOP.K=P1.K*P2.K/CLIP(P1.K-P2.K,P2.K-P1.K,P1.K,P2.K)
00580 CPLOT  N.BASE,N.TWO,N.THREE,N.FOUR,N.FIVE,N.SIX,N.TEN(0,8)
00590 CPLOT  TTH1.BASE,TTH1.TWO,TTH1.THREE,TTH1.FOUR,TTH1.FIVE,
00600 X      TTH1.SIX,TTH1(-4,4)
00610 CPLOT  TTH2.BASE,TTH2.TWO,TTH2.THREE,TTH2.FOUR,TTH2.FIVE,
00620 X      TTH2.SIX,TTH2(-4,4)
00630 CPLOT  NH.BASE,NH.TWO,NH.THREE,NH.FOUR,NH.FIVE,NH.SIX,
00640 X      NH(0,8)
00650 C      DT=.1
00660 N      TIME=TIMEN
00670 C      TIMEN=0
00680 C      LENGTH=4
00690 C      PLTPER=0
00700 C      PRTPER=0
00710 C      SAVPER=.1
00720 SAVE  N,TTH1,TTH2,NH
00730 RUN   BASE
```



00740	C	DTC1=2
00750	C	DTC2=2
00760	RUN	TWO
00770	C	DTC1=3
00780	C	DTC2=3
00790	RUN	THREE
00800	C	DTC1=4
00810	C	DTC2=4
00820	RUN	FOUR
00830	C	DTC1=5
00840	C	DTC2=5
00850	RUN	FIVE
00860	C	DTC1=6
00870	C	DTC2=6
00880	RUN	SIX
00890	C	DTC1=10
00900	C	DTC2=10
00910	C	PLTPER=.1
00920	RUN	FIGURES 5-7, 5-9, 5-10, 5-12
00930	C	DTC1=1.18325
00940	C	DTC2=.845178
00950	RUN	BASE
00960	C	DTC1=2.3364
00970	C	DTC2=1.6902
00980	RUN	TWO
00990	C	DTC1=3.5496
01000	C	DTC2=2.5354
01010	RUN	THREE
01020	C	DTC1=4.7328
01030	C	DTC2=3.3805
01040	RUN	FOUR
01050	C	DTC1=5.916
01060	C	DTC2=4.2257
01070	RUN	FIVE
01080	C	DTC1=7.0992
01090	C	DTC2=5.0708
01100	RUN	SIX
01110	C	DTC1=11.832
01120	C	DTC2=8.4514
01130	C	PLTPER=.1
01140	RUN	FIGURE 5-8

APPENDIX H

EQUATION LISTING FOR NOFTIME

NOFTIME.DYNAMO

00010 \*       ENTRAINMENT MEASURES  
00020 C       P2=4  
00030 C       P1=2.713  
00040 N       A=- (39.478/(P1\*P1)) +C\*C/4  
00050 N       B=- (39.479/(P2\*P2)) +D\*D/4  
00060 N       C=-2/DTC1  
00070 C       DTC1=2  
00080 N       D=-2/DTC2  
00090 C       DTC2=2  
00100 L       AL.K=AL.J+(DT) (CAL.JK)  
00110 N       AL=CIC\*Q/(2\*A\*C)  
00120 C       Q=1  
00130 C       CIC=1  
00140 R       CAL.KL=2\*(C\*AL.K+BE.K)  
00150 L       BE.K=BE.J+(DT) (CBE.JK)  
00160 N       BE=-CIC\*Q/(2\*A)  
00170 R       CBE.KL=C\*BE.K+GA.K+A\*AL.K  
00180 L       GA.K=GA.J+(DT) (CGA.JK)  
00190 N       GA=(CIC\*Q) ((C\*C-A)/(2\*A\*C))  
00200 R       CGA.KL=2\*A\*BE.K+1  
00210 L       ET.K=ET.J+(DT) (CET.JK)  
00220 N       ET=CIC\*Q/(2\*B\*D)  
00230 R       CET.KL=2\*(D\*ET.K+SI.K)  
00240 L       SI.K=SI.J+(DT) (CSI.JK)  
00250 N       SI=-CIC\*Q/(2\*B)  
00260 R       CSI.KL=D\*SI.K+LA.K+B\*ET.K  
00270 L       LA.K=LA.J+(DT) (CLA.JK)  
00280 N       LA=(CIC\*Q) ((D\*D-B)/(2\*B\*D))  
00290 R       CLA.KL=2\*B\*SI.K+1  
00300 L       DE.F=DE.J+(DT) (CDE.JK)  
00310 N       DE=CIE\*Q\*(C+D)/K  
00320 C       CIE=1E-10  
00330 R       CDE.KL=C\*DE.K+RH.K+D\*DE.K+XI.K  
00340 L       XI.K=XI.J+(DT) (CXI.JK)  
00350 N       XI=-CIE\*Q\*(B-A+C\*D+D\*D)/K  
00360 R       CXI.KL=C\*XI.K+NU.K+B\*DE.K  
00370 L       RH.K=RH.J+(DT) (CRH.JK)  
00380 N       RH=CIE\*Q\*(B-A-C\*D-C\*C)/K  
00390 R       CRH.KL=A\*DE.K+D\*RH.K+NU.K

```
00400 L      NU.K=NU.J+(DT) (CNU.JK)
00410 N      NU=CIE*Q*(C*C*D-A*C+C*D*D-B*D)/K
00420 R      CNU.KL=A*XI.K+B*RH.K+1
00430 N      K=(A*D+B*C) (C+D) - (A-B) (A-B)
00440 A      ID.K=1/(AL.K*GA.K-BE.K*BE.K)
00450 A      T11.K=ID.K*(DE.K*GA.K-RH.K*BE.K)
00460 A      T12.K=ID.K*(-DE.K*BE.K+RH.K*AL.K)
00470 A      T21.K=ID.K*(XI.K*GA.K-NU.K*BE.K)
00480 A      T22.K=ID.K*(-XI.K*BE.K+NU.K*AL.K)
00490 A      TTH1.K=T21.K/T11.K
00500 A      TTH2.K=-T12.K/T22.K
00510 A      GT11.K=ET.K-AL.K*T11.K*T11.K-GA.K*T12.K*T12.K-
00520 X      2*BE.K*T11.K*T12.K
00530 A      GT22.K=LA.K-AL.K*T21.K*T21.K-GA.K*T22.K*T22.K-
00540 X      2*BE.K*T21.K*T22.K
00550 A      GT12.K=SI.K-AL.K*T11.K*T21.K-BE.K*
00560 X      (T21.K*T12.K+T11.K*T22.K)-GA.K*T22.K*T12.K
00570 A      GT21.K=GT12.K
00580 A      GTIDET.K=MAX(1E-15,GT11.K*GT22.K-GT12.K*GT21.K)
00590 A      GTI11.K=GT22.K/GTIDET.K
00600 A      GTI12.K=-GT12.K/GTIDET.K
00610 A      GTI21.K=-GT21.K/GTIDET.K
00620 A      GTI22.K=GT11.K/GTIDET.K
00630 A      TERM3.K=T11.K*GTI11.K+T21.K*GTI21.K
00640 A      TERM4.K=T11.K*GTI12.K+T21.K*GTI22.K
00650 A      TERM5.K=T12.K*GTI11.K+T22.K*GTI21.K
00660 A      TERM6.K=T12.K*GTI12.K+T22.K*GTI22.K
00670 A      M11.K=TERM3.K*T11.K+TERM4.K*T21.K
00680 A      M12.K=TERM3.K*T12.K+TERM4.K*T22.K
00690 A      M21.K=TERM5.K*T11.K+TERM6.K*T21.K
00700 A      M22.K=TERM5.K*T12.K+TERM6.K*T22.K
00710 A      N.K=M11.K*AL.K+M12.K*BE.K+M21.K*BE.K+M22.K*GA.K
00720 PRINT 1)A,B,C,D
00730 PRINT 2)AL,BE,GA,ET
00740 PRINT 3)SI,LA,DE,XI
00750 PRINT 4)K,RH,NU,ID
00760 PRINT 5)T11,T12,T21,T22
00770 PRINT 6)GT11,GT12,GT21,GT22
00780 PRINT 7)GTI11,GTI12,GTI21,GTI22
00790 PRINT 8)M11,M12,M21,M22
00800 PRINT 9)TTH1,TTH2,N
```

```
00810 C      DT=.01
00820 N      TIME=TIMEN
00830 C      TIMEN=0
00840 C      LENGTH=12
00850 C      PLTPER=0
00860 C      PRTPER=0
00870 C      SAVPER=.3
00880 SAVE   N
00890 CPLOT  N.BASE,N.TWO,N.THREE,N(0,8)
00900 RUN    BASE
00910 C      DTC1=4
00920 C      DTC2=4
00930 C      P1=3.252
00940 RUN    TWO
00950 C      DTC1=4
00960 C      DTC2=4
00970 C      P1=5.19
00980 RUN    THREE
00990 C      DTC1=10
01000 C      DTC2=10
01010 C      P1=3.68
01020 C      PLTPER=.3
01030 RUN    FIGURE 5-11
```

APPENDIX I

DEFINITIONS OF VARIABLES IN CALCULATIONS OF N

NAME	NO	T	DEFINITION
A			PARAMETER CONTROLLING PERIOD IN FIRST SYSTEM
AL			ALPHA -- ELEMENT 1,1 OF GAMMA11
B			PARAMETER CONTROLLING PERIOD IN SECOND SYSTEM
BE			BETA -- OFF-DIAGONAL ELEMENTS OF GAMMA11
C			PARAMETER CONTROLLING DAMPING IN FIRST SYSTEM
CAL			CHANGE IN AL -- ALPHA
CBE			CHANGE IN BE -- BETA
CDE			CHANGE IN DE -- DELTA
CET			CHANGE IN ET -- ETA
CGA			CHANGE IN GA -- GAMMA
CIC			COEFFICIENT ON INITIAL COVARIANCE
CIE			COEFFICIENT ON INITIAL ENTRAINMENT
CLA			CHANGE IN LA -- LAMBDA
CNU			CHANGE IN NU -- NU
CRH			CHANGE IN RH -- RHO
CSI			CHANGE IN SI -- SIGMA
CXI			CHANGE IN XI -- XI
D			PARAMETER CONTROLLING DAMPING IN SECOND SYSTEM
DE			DELTA -- ELEMENT 1,1 OF GAMMA12
DTC1			DAMPING TIME CONSTANT OF FIRST SYSTEM
DTC2			DAMPING TIME CONSTANT OF SECOND SYSTEM
ET			ETA -- ELEMENT 1,1 OF GAMMA22
GA			GAMMA --ELEMENT 2,2 OF GAMMA11
*			CONDITIONAL ERROR COVARIANCE OF ESTIMATED STATE OF SECOND SYSTEM GIVEN THE STATE OF THE FIRST
GAMMA11			UNCONDITIONAL COVARIANCE OF STATE OF FIRST SYSTEM
GAMMA12			UNCONDITIONAL COVARIANCE BETWEEN FIRST SYSTEM AND SECOND SYSTEM
GAMMA22			UNCONDITIONAL COVARIANCE OF STATE OF SECOND SYSTEM

GMDTC	GEOMETRIC MEAN OF TIME CONSTANTS
GTIDET	DETERMINANT OF GAMMATILDE
GTI11	ELEMENT 1,1 OF GAMMATILDE INVERSE
GTI12	ELEMENT 1,2 OF GAMMATILDE INVERSE
GTI21	ELEMENT 2,1 OF GAMMATILDE INVERSE
GTI22	ELEMENT 2,2 OF GAMMATILDE INVERSE
GT11	ELEMENT 1,1 OF GAMMATILDE
GT12	ELEMENT 1,2 OF GAMMATILDE
GT21	ELEMENT 2,1 OF GAMMATILDE
GT22	ELEMENT 2,2 OF GAMMATILDE
ID	INVERSE OF DETERMINANT OF GAMMA11
IL2RP	INITIAL LOG BASE TWO OF THE RATIO OF PERIODS
K	CONSTANT FOR COMPUTATION OF K1
KNH	SCALING CONSTANT FOR NH
K1	CONSTANT TERM FOR ALL ELEMENTS OF T
LA	LAMBDA -- ELEMENT 2,2 OF GAMMA22
L2P1	LOG BASE TWO OF PERIOD ONE
L2P2	LOG BASE TWO OF PERIOD TWO
L2RP	LOG BASE TWO OF THE RATIO OF PERIODS
M	SYMMETRICAL MATRIX USED FOR COMPUTATION OF N
M11	ELEMENT 1,1 OF M
M12	ELEMENT 1,2 OF M
M21	ELEMENT 2,1 OF M
M22	ELEMENT 2,2 OF M
N	ENTRAINMENT INDEX
NH	HEURISTIC ENTRAINMENT INDEX
NU	NU -- ELEMENT 2,2 OF GAMMA12
P1	PERIOD OF FIRST SYSTEM
P2	PERIOD OF SECOND SYSTEM
Q	AMPLITUDE FACTOR OF WHITE NOISE INPUT
RH	RHO -- ELEMENT 2,1 OF GAMMA12
SI	SIGMA --OFF-DIAGONAL ELEMENTS OF GAMMA22
T	MATRIX DEFINING CONDITIONAL STATE OF SECOND SYSTEM GIVEN THE FIRST

TERM3	TERM FOR COMPUTATION OF M11 AND M12
TERM4	TERM FOR COMPUTATION OF M11 AND M12
TERM5	TERM FOR COMPUTATION OF M21 AND M22
TERM6	TERM FOR COMPUTATION OF M21 AND M22
TERM9	TERM FOR COMPUTATION OF TERM5
TIMEN	TIME INITIAL
TOP	TIME TO GET OUT OF PHASE
TTH1	TAN THETA BETWEEN UNIT VECTOR 1 AND THE SECOND STATE VECTOR
TTH2	TAN THETA BETWEEN UNIT VECTOR 2 AND THE SECOND STATE VECTOR
T11	ELEMENT 1,1 OF T
T12	ELEMENT 1,2 OF T
T21	ELEMENT 2,1 OF T
T22	ELEMENT 2,2 OF T
XI	XI -- ELEMENT 1,2 OF GAMMA12

BIOGRAPHICAL NOTE

Alan Karl Graham was born in Iowa City, Iowa on September 29, 1949 to Arlene and J. Robert Graham. He graduated from Robbinsdale High School in Minneapolis, Minnesota in 1967. Admitted to M.I.T. as an M.I.T. National Scholar, he was a teaching assistant in 1969 for 15.571, an introductory laboratory course in feedback systems. In 1970, he was in charge of the course. He was selected in 1969 to participate in the Sloan School's Undergraduate Studies Program, a two-year program of independent study and research leading to a bachelor's degree in management. He obtained two bachelor's degrees in 1973 in management and electrical engineering.

Entering the Doctoral Program of the Department of Electrical Engineering in 1973 as a National Science Foundation Fellow, he earned a Master of Science degree in Electrical Engineering with his thesis "Feedback Processes Underlying the Onset of Puberty in Males." In 1977, he renovated and gave the second-semester core course in System Dynamics, 15.873. In 1977 he completed the requirements of the doctoral program with his thesis "Principles on the Relationship between Structure and Behavior of Dynamic Systems."

He is a member of Eta Kappa Nu, an electrical engineering honorary, Sigma Xi, a research honorary, and the Institute of Electrical and Electronics Engineers (IEEE).

His work experience includes consultation to the Cummins Engine Company, a manufacturer of diesel engines (1969), research on urban problems under contract with the Department of Housing and Urban Development (1970-72), and research on economic dynamics as a part of the System Dynamics National Model Project (1974-77).

His publications include:

"Modeling City-Suburb Interaction in the Urban Dynamics Model," IEEE Transactions on Systems, Man, and Cybernetics, April 19, 1972.

"Understanding Urban Dynamics: An Analysis of Garn's Interpretation," in Mass, Nathaniel J., Editor, Readings in Urban Dynamics, Volume 1 (Cambridge, Mass.: Wright-Allen Press, 1974).

"Modeling City-Suburb Interactions," in Mass, op. cit.

"Corporate Finance in a National Socio-Economic Model," Proceedings of the Summer Computer Simulation Conference, San Francisco, 1975.



"Non-Statistical Selection of Parameter Values in System Dynamics Models,"  
Proceedings of the Summer Computer Simulation Conference, Washington,  
D.C., 1976

"Parameter Formulation and Estimation in System Dynamics Models,"  
The System Dynamics Method: Proceedings of the Fifth International  
Conference on System Dynamics, Geilo, Norway, 1976.

Alfeld, Louis E., and Alan K. Graham, Introduction to Urban Dynamics  
(Cambridge, Mass.: Wright-Allen Press, 1976, 322 pages).