ELEMENTARY SYSTEM DYNAMICS STRUCTURES

by

MICHAEL ROLFE GOODMAN

S.B., Purdue University
(1970)

SUBMITTED IN
PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF
TECHNOLOGY

January, 1972

Signature of Author..
Department of Mechanical Engineering, January 21, 1972

Certified by..
Thesis Supervisor

Certified by.... ....
Thesis Supervisor

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

ABSTRACT

ELEMENTARY SYSTEM DYNAMICS STRUCTURES

by

Michael Rolfe Goodman

Submitted to the Department of Mechanical Engineering on January 21, 1972 in partial fulfillment of the requirements for the degree of Master of Science

System Dynamics is a simulation modeling approach capable of yielding insight into the performance of physical as well as social systems. System Dynamics models are composed of interconnected, non-linear feedback loops involving material and information flows (rates) and the accumulations of flows (levels). Many times a system cannot be grasped as a whole and requires that investigation be made of the individual loops or simple structures comprising the model. Frequently, simple structures are encountered which are common to many systems. This thesis begins an effort to collect and organize fundamental structures into an anthology of these generic structures. This anthology should assist the serious System Dynamics student.

Two first order structures are examined. The negative feedback structure involves a decision process intervening between the level and rate, attempting to reach and maintain a goal. The goal-oriented structure is typical of man-made as well as biological control processes. The linear negative feedback structure produces exponential decay behavior over time. The equilibrium value reached when a constant input is applied to the linear structure will not be the desired goal value. A nonlinear negative feedback process shares many of the goal-directed characteristics of the linear process. However, a nonlinear structure involving a saturation phenomenon is unable to maintain a goal in the presence of a constant input whose value exceeds the maximum compensatory rate of the structure.

The elementary structure capable of producing S-shape growth, exponential growth followed by exponential decay, adequately represents many real world growth cycles. Population growth and diffusion phenomena are classic examples. S-shape growth, in its simplest form, requires a nonlinear rate-level relationship which begins with positive feedback and concludes with negative feedback.

Thesis Supervisors: Jay W. Forrester, Professor of Management

Thomas B. Sheridan, Professor of Mechanical Engineering

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

The Importance of Simple Feedback Structures

Frequently, a feedback structure emerges which seems to be pertinent to many different systems. A feedback structure having this characteristic is classified as a simple or generic structure. Generic structures are the building blocks of System Dynamics models. The study of generic structures, the intent of this thesis, has many important advantages.

Very often a complex system can be adequately represented by a simple structure. A system whose behavior is dominated by a particular structure is a likely candidate. In a similar vein, all higher order (complex) structures are composed of interconnected simple structures. Focusing attention on an elementary structure often yields insight into the dynamics of the overall system, insight that might be obscured otherwise.

Simple structures are generally easy to comprehend. A simple structure is useful for communicating fundamental ideas of structure and behavior to decision-makers. They have significant value as teaching vehicles. An intuitive grasp of why a particular type of behavior eventuates from a given structure is one by-product. The ease with which one simple structure can be used to understand a whole host of phenomenon previously housed under separate academic disciplines is another educational by-product.

Having a repertoire of basic structures at one's disposal facilitates the modeling process. Seemingly unrelated phenomenon having a common mode of behavior might share the same basic structure. The elementary relationships underlying the dynamics of an ecological system might apply to an economic system for example. Knowledge of the elementary feedback structure capable of producing a given behavior alerts the modeler to the kinds of structural relationships that could prove to be important. The modeler has at least a starting point for his modeling activity.

This thesis is the beginning effort aimed at cataloging and investigating elementary System Dynamics structures. Two generic structures are developed. Chapter 2 contains the most fundamental feedback loop, the goal-seeking negative feedback loop producing exponential decay. Chapter 3 utilizes the structure in Chapter 2 and combines with it the positive loop associated with exponential growth to form the mechanism responsible for generating S-shape growth over time.

Each chapter is self-contained. The behavioral phenomenon is introduced followed by a detailed analysis of the underlying rate-level relationship of the generic structure. Three specific adaptations of the structure to interesting and varied occurrences of the behavior mode complete each chapter. It is assumed only that the reader has had a brief introduction to the modeling approach of System Dynamics, including Dynamo notation.*

---

*For the reader interested in obtaining a background, Forrester's Principles of Systems (Forrester 1968) is recommended.

Chapter 2
First Order Negative Feedback

Introduction

Negative feedback as used in System Dynamics is defined as goal directed or goal oriented behavior. Any process that is governed by an implicit or explicit objective qualifies as negative feedback. Frequently, systems dominated by negative feedback are characterized by terms such as self-governing, self-regulating, self-equilibrating, homeostatic, or adaptive which imply the presence of a goal. A few examples should illustrate the goal seeking nature of negative feedback systems.

The thermostatically controlled heating/air conditioning system responsible for regulating much of man's indoor environment is a common self-governing system. A causal diagram of the system is presented in Figure 1.

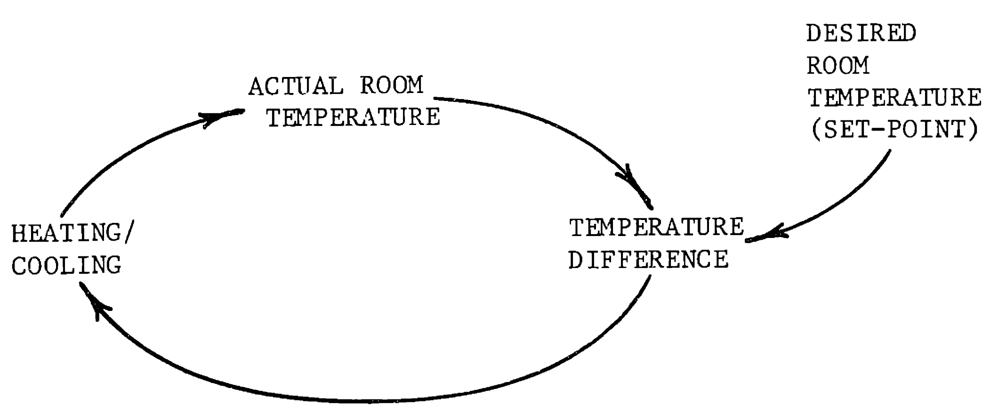


Figure 1

Thermostat Heating/Air-Conditioning System

Involved is an interacting set of electrical, mechanical, and thermal components. The desired room temperature or "set temperature" is the goal that the system attempts to maintain. The human component of the system is not an active factor since once the set temperature is selected, the system is independently self-regulating (the reason for its popularity). When a disparity between the desired and the actual room temperature is sensed by the thermostat, the decision-making unit, the heating or air-conditioning unit is switched on. The heat addition or removal eventually raises or lowers the room temperature. The thermostat automatically shuts off the heater or cooler when the room temperature has reached the desired temperature and a discrepancy no longer exists.

The diagram of the physical system in Figure 1 could easily apply to a thermostatically controlled oven, the electric eye of a camera, the automatic pilot of an airplane, or the speed governor of an engine. These systems all belong to a class known as control systems. Arising from the field of control technology, such systems are specifically designed to achieve and maintain a particular objective. The notion of control is itself expressive of a goal orientation.

Mechanically controlled systems are not without analogy in the biological world. The human body, for instance, is composed of numerous self-regulating physiological processes. The ultimate purpose of such processes is to maintain a relatively constant internal environment necessary for survival and described as homeostasis. The temperature regulation system is a typical example. Its purpose is to maintain the normal body temperature. Continual alteration of the metabolic activities or blood flow rates are the mechanisms involved. Digestion, blood-sugar regulation, and waste removal are additional examples of homeostatic processes.

Homeostatic systems are not restricted to individual organisms. The growth and regulation of a single population as well as a community composed of various populations (i.e., an ecosystem) involves negative feedback. The survival and maintenance of a population depends upon a myriad of ecological checks and balances. Competition, parasitism, predation, food supply, soil, light, and weather, to mention a few, all act to prevent the population from inundating the environment responsible for survival. Ecological control is exercised both in the short term through migration, disease, starvation, and cannibalism as well as in the long term through evolutionary adaptation. There are few elements in the organic world that do not depend on feedback control for survival.

Goal directed action is a fundamental part of man's activities. For purposes of illustration, consider the socialization of the child in a parental environment. Parents transmit their values, attitudes, and expectations to the child largely through a trial and error process, itself an expression of negative feedback. When a discrepancy arises between the behavior they desire and the behavior of the child, the parents take appropriate corrective action in the form of reward and punishment. In turn, by making mistakes, the child learns through trial and error what is expected of him and how he is to behave to appease his parents.

In this chapter, the basic attributes of simple negative feedback will be investigated. The causal and flow diagrams of first order negative feedback will be introduced. The behavior of the system over time will be investigated by considering graphically the rate-level relationship involved in negative feedback. The definition of the time constant will be developed and related to the slope of the rate-level graph. The response of the structure to constant inputs will be discussed, followed by mention of nonlinear negative feedback. Three case studies will be presented: an inventory control system; a coffee cup; and a pollution dissipation system.

Causal Diagram of Simple Negative Feedback

It is feasible to pin down the elements of negative feedback in detail. The diagram of Figure 2 is introduced. Four elements form the structure: the desired state (goal), the discrepancy, the action (rate), and the system state (level). The thermostat system in Figure 1 is a specific example of the feedback loop in Figure 2.

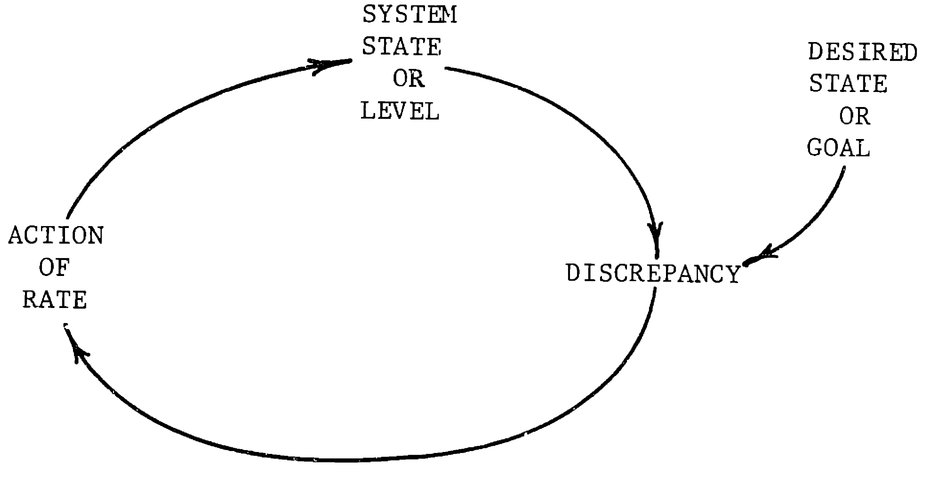


Figure 2

Causal Diagram--Negative Feedback

For a simple negative feedback system, one containing a single level, there are no direct causal links from the system to the goal. The goal is determined externally or exogenously and is not involved in the circular loop except as an input. It serves as a reference or guideline on which the system bases action. The set temperature, the desired speed of an automobile, or the desired inventory of a warehouse are common examples of goals.

The system state or, alternatively, the level, is the object of control. It is the accumulation of all past action. The amount of heat in the room, the inventory, the population, the traditions of a society, are typical system states.

The only way that the state of the system can be modified is through the action element or rate. Action is defined as the activity that is utilized to alter the system state. Heating or cooling a room, births and deaths, sales and production are all activities which result in decrements or increments with respect to time to the level.

The magnitude and direction of the action taken depends upon the discrepancy between the goal and the state of the system. Since the discrepancy element must be able to sense the state of the system and compare it with the goal, it is often the most complex link. In the heating/air-conditioning system this function is performed by the thermostat. A sensor such as a thermocouple monitors the temperature of the room. This might be in the form of a proportional electrical resistance. The resulting voltage is compared to a reference voltage from the thermostat setting. The voltage drop between the two activates the appropriate cooling or heating unit.

For a man attempting to keep his car on the road, the discrepancy function is performed by the brain. Through his visual perception, the driver is able to sense the lateral position (state) of his vehicle relative to the curb. Should he find that the auto is not in the desired position, he initiates corrective action. The resulting mechanical alteration of the steering wheel modifies the position of the automobile.

The sensing, comparison, and decision-making intervening between the level and rate can be viewed as the information, control, or more generally, the decision process sector (Forrester 1968, p. 36):

> ...a decision process is one that controls any system action. It can be a clear explicit human decision. It can be a subconscious decision. It can be the governing processes in biological development. It can be the natural consequences of the physical structure of the system.

The decision process sector, in fact, completes the circularity of the system. The presence of the decision process can be used to differentiate the non-feedback, "open loop", system from the feedback, "closed loop", system.

The closed loop nature of the system in Figure 2 can be illustrated. Suppose the system state was initially the same as the goal. An arbitrary increase in the state, for instance, would then cause a discrepancy to appear between the goal and state of the system. In order to minimize the discrepancy, the system would take action to decrease the level. This decrease, resulting from the internal relationship of the elements in the closed loop, is in opposition to the outside change in the system level. This same scenario would be witnessed in any of the feedback systems mentioned earlier.

An open loop system is one in which the state does not alter the rate. The decision process is explicitly missing. Cyclic machines such as toasters, automatic washing machines, and vending machines (when a human operator is not actively adjusting the machines) are examples of open loop systems.

Flow Diagram

From the dicsussion of the fundamental elements of negative feedback, the generalized flow diagram of Figure 3 can be produced. The valve symbol or rate RT is the action component of the system and is the only means of altering the system state, the level LEV. Information about the magnitude of the level LEV is compared to the goal GL through the discrepancy DISC. The discrepancy DISC when modified by the amount or fraction of the action taken per time unit FPT determines the rate RT.

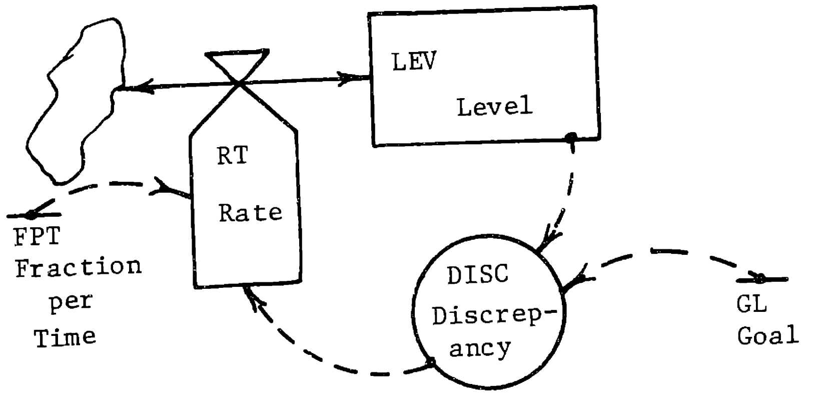


Figure 3

Flow Diagram--First Order Negative Feedback

The Dynamo equations for the first order loop follow:

L	LEV.K	=	LEV.J + (DT)(RT.JK)	(Units)
N	LEV	=	0	(Units)
R	RT.KL	=	FPT*DISC.K	(Units/Time)
C	FPT	=	.1	(Fraction/Time)
A	DISC.K	=	GL - LEV.K	(Units)
C	GL	=	100	(Units)
C	DT	=	1	(Time)

Eliminating the auxilary equation and constants by substituting them into the rate equation yields:

R RT.KL = FPT*(GL-LEV.K)

A useful way of investigating the behavior of the system is to graph the rate RT versus the level LEV for arbitrary values of the constants FPT and GL using the rate equation. This is done in Figure 4.

The relationship in Figure 4 does not explicitly depend on time and is thus a static relationship. The slope of line FPT, the horizontal axis intercept GL, and the vertical axis intercept FPT*GL, are determined by the values of the constants. The importance of these constants will be examined later.

The behavior of the system with respect to time can be graphically simulated by using Figure 4 and constructing a level versus time plot. A simple procedure for doing this is:

1. The rate RT is determined from the most recent level LEV value. For the starting conditions of this system, the

initial value of the level LEV is zero.

2. The product of the latest RT value and the time interval DT is added to the latest level value LEV. This is equivalent to adding an increment to the level LEV and yields a revised level value which is plotted on a level LEV versus time graph one DT to the right of the last calculation.

3. With the new level value LEV, a new rate value RT is computed and so forth until the time span of interest is covered. Or equivalently, one returns to step 1 and step 2 until the level-time plot is computed.

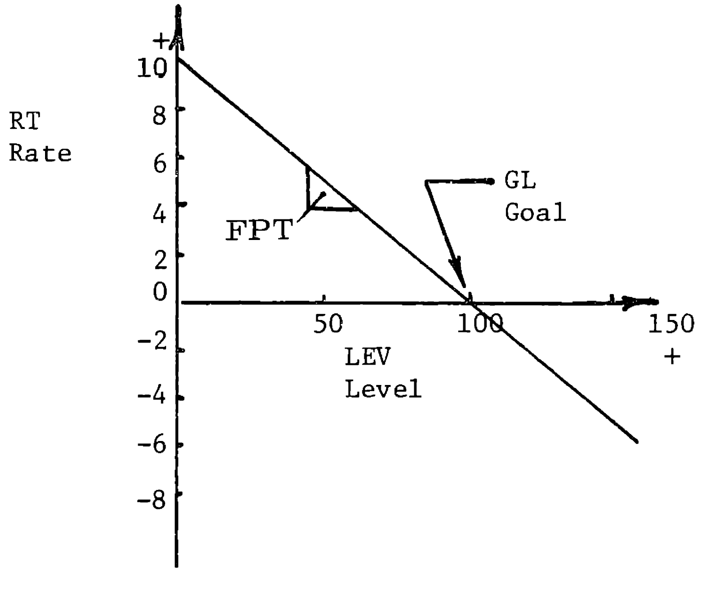


Figure 4
Rate-Level Graph

The algorithm above is followed in Figure 5. The initial value of level LEV(LEV(0) = 0) yields an initial rate RT(RT(0) = 10), as seen in Figure 5(a). The product of DT and RT(0) is added to LEV(0) producing LEV(1) at time equal to one unit (Figure 5(b)). LEV(1) is used in turn to determine RT(1) which when multiplied by DT is added to LEV(1) producing LEV(2) at time equal to two units. Because of the negative slope of rate-level graph, each new increment to the level LEV becomes smaller and smaller. This is because each new rate value RT is FPT*DT*100% or in this case 10% of the previous one. Eventually, the rate RT becomes approximately zero as it approaches LEV = 100 and new increments to the level cease. Once the goal GL is reached, the system enters a steady or equilibrium state since the rate RT is zero. Connecting the points generated in step 2 produces the characteristic curve of a linear, first order negative feedback system (Figure 5(b)).

The behavior visualized in Figure 5(b) can be viewed as containing two distinct regions: the transient and the steady state. These regions are seen clearly in Figure 6, a reproduction and expansion of Figure 5(b). The transient region is characterized by its goal seeking or transitory nature. In this region the level value is different from the goal value. The steady state region is characterized by its goal attainment or stationary behavior. When the time shape of the system state is constant as in the steady state region, the system is said to be in equilibrium.

Mathematical definitions of the transient and steady states can also be useful. The transient region is identified by a varying rate not equal to zero. The steady state region is defined by a rate that is zero. Extending these definitions to Figure 4, it is noted that the steady state region reduces to a point--the goal GL. Any other point on the curve in Figure 4 must lie in the transient region.

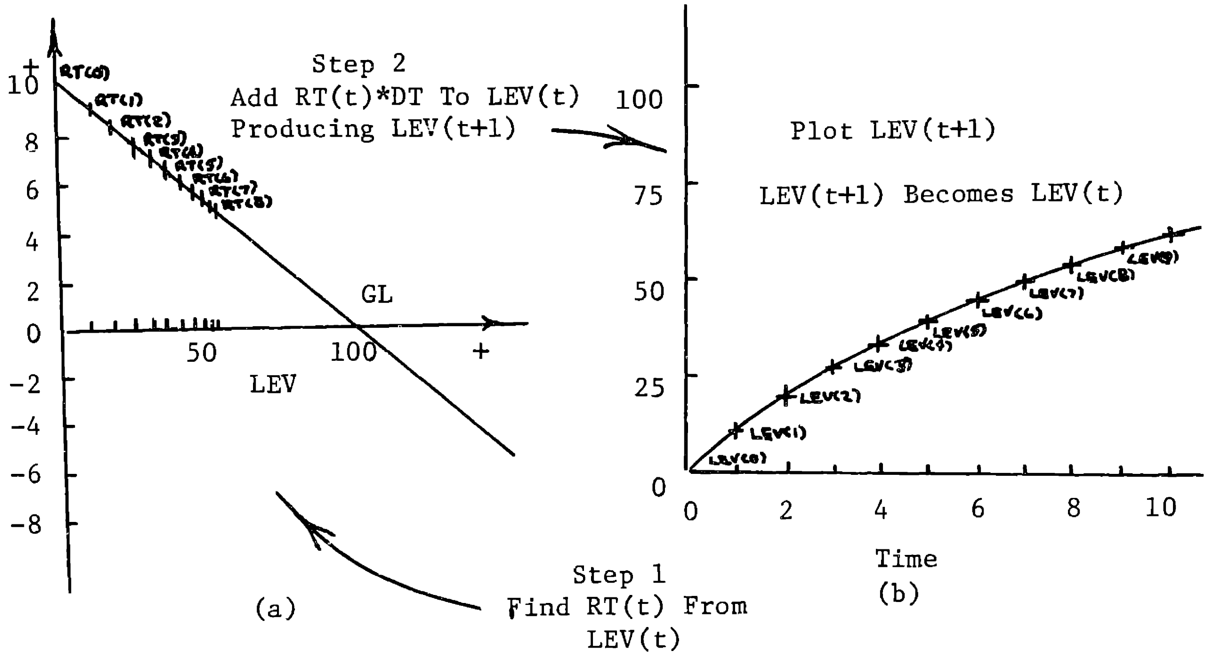


Figure 5

Simulation of Negative Feedback

An analytical expression for the system can be derived which gives the value of the level LEV at any point in time:

$$\text{(1)} \qquad LEV(t) = GL + (LEV(0) - GL)e^{(-FPT \cdot t)}$$

where

LEV(t) = level value at time t

GL = goal

LEV(0) = level initial value

e = exponential function

FPT = slope

t = time

The presence of the minus sign in the exponential power indicates that the system involves exponential decay.

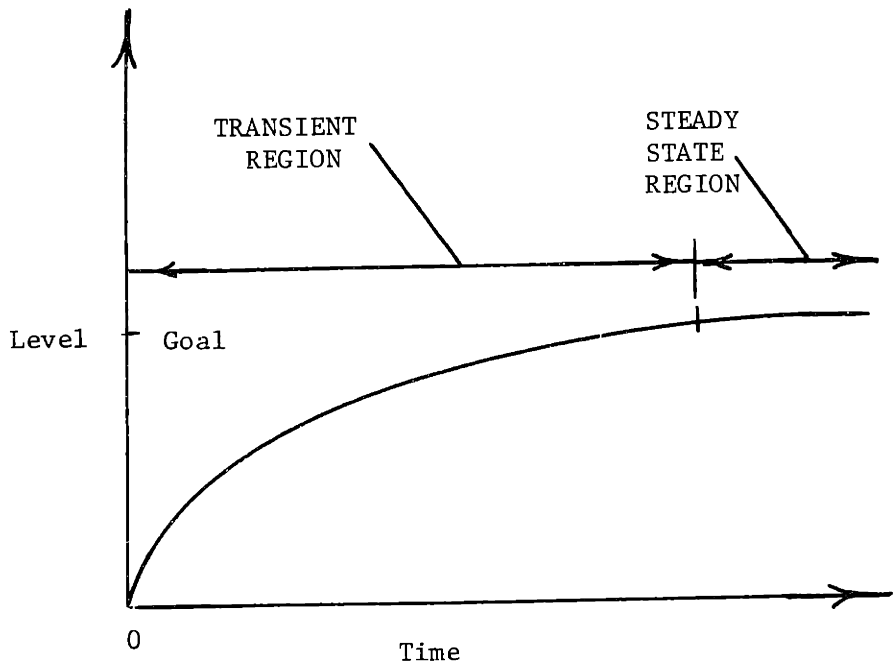


Figure 6
Time Plot of Negative Feedback

Slope and Time Constant

The slope FPT has some interesting properties. When a period of time has elapsed equal to the inverse of the slope FPT, the level LEV has attained 63 percent of the discrepancy between GL and LEV(0). This is demonstrated by the use of equation (1):

$$LEV(t=1/FPT) = GL + (LEV(0)-GL)e^{(-1)}$$

$$= LEV(0) + .632\ (GL-LEV(0))$$

The inverse of the slope is called the time constant T of the system. It allows one to compare the transient responses of various systems. The larger the value of T, the longer it takes for LEV to reach 63 percent of its final value.

The value of the slope FPT does not alter the general behavior of system. The direction (sign) of the slope FPT, however, does. In fact, the direction of the slope for a single level system differentiates a positive feedback system from a negative system. Figure 7 illustrates this point.

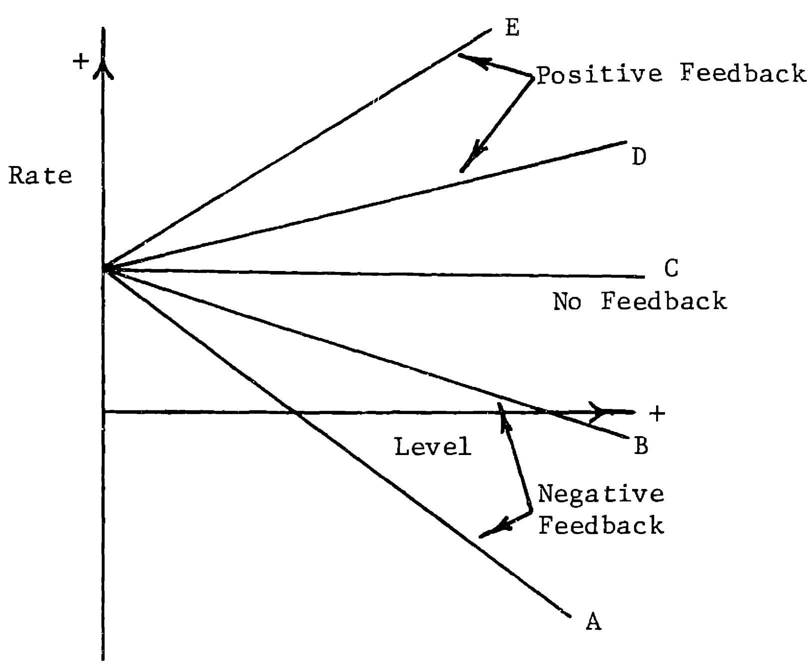


Figure 7

Rate-Level Slopes

Lines A and B, for example, differ only in the magnitude of the slope FPT. They are both negative feedback systems. A system whose rate-level graph yields line D or E is a positive feedback system. The structure producing curve C, a horizontal line having a zero slope, is an open loop system. The rate resulting from curve C is no longer a function of the level.

Goal Values

As seen in Figure 4, in a negative feedback system, the intercept of the rate-level line determines the final or equilibrium value of the system. It was noted that the level intercept was simply the goal or desired level of the system. The question is now asked, "What if the goal were zero?"

A special but not uncommon case of negative feedback is produced by letting the goal become zero. The causal loop, flow diagram and equations for the structure are seen in Figure 8.

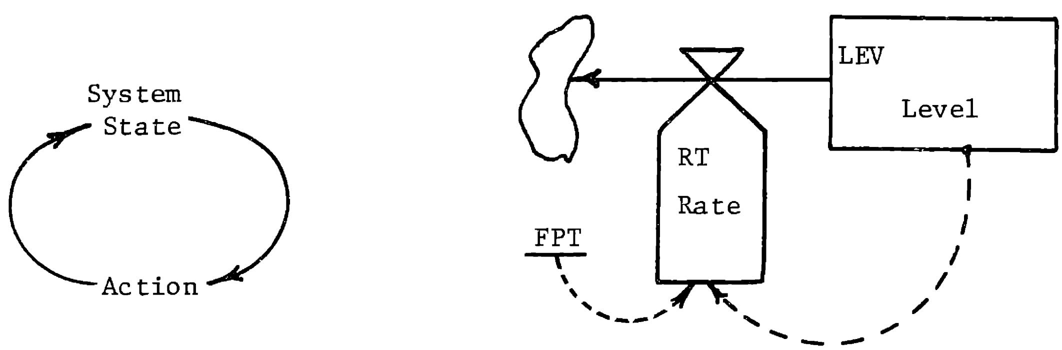
Two of the basic elements of the loop have been made implicit. First, the exogenous goal no longer is required. Second, because the goal is zero, it is no longer necessary to include the discrepancy link. The resulting rate-level graph for this system is seen in Figure 9.[1]

From Figure 9 it is observed that the rate RT is always negative and thus out of the level LEV. Additionally, the equilibrium value of the system is zero. Only when the rate RT is zero can the system achieve a

---

1. Negative level values are assumed to be meaningless.

steady state which in this system must occur at the zero level value. Should the initial value of the level be zero, then the system would be in equilibrium. A non-zero starting value is required for the transient behavior of the system to be observed. Figure 10 is an example of the transient behavior mode.



L	LEV.K	=	LEV.J + (DT)(RT.JK)	(Units)
N	LEV	=	100	(Units)
R	RT.KL	=	-FPT*LEV.K	(Units/Time)
C	FPT	=	.10	(Fraction/Time)

Figure 8

Negative Feedback-Goal Equal Zero

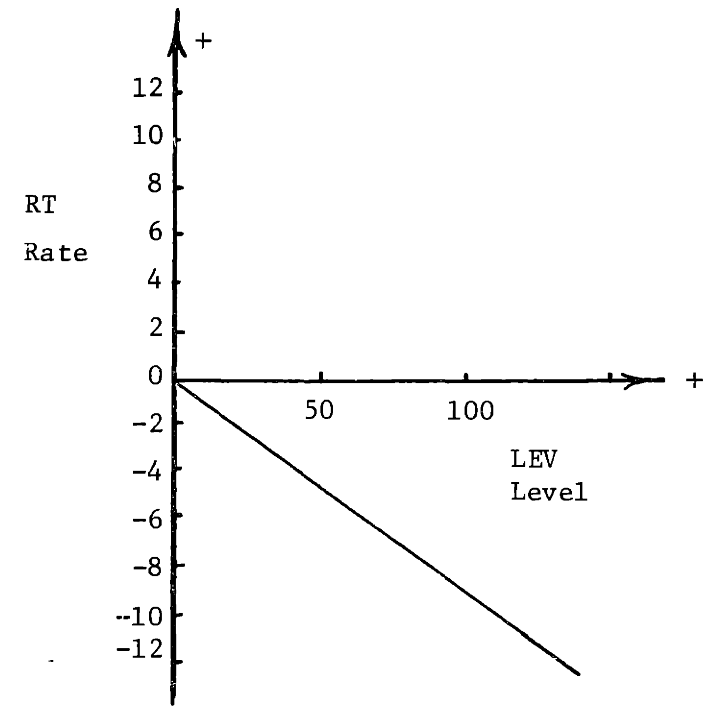


Figure 9

Zero-Goal Rate-Level Graph

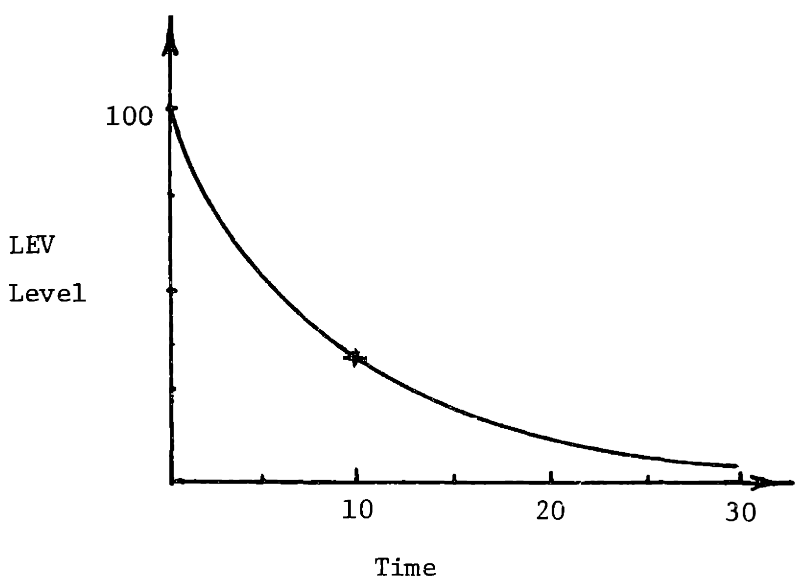


Figure 10

Zero-Goal Level Response

The behavior in Figure 10 is well known in the chemical and biological sciences. Radioactive decay is a prime example. In general, the behavior of any system requiring constant rejuvenation or maintenance seems to display the decay behavior. The depreciation of capital and the decline of a population having a higher death rate than birth rate are other examples.

From equation (1) it is possible to derive an exact solution for the system by simply letting GL equal zero:

$$LEV(t) = LEV(0)*e^{(-1/T)(t)}$$

The equation above has the exact form of the exponential increase equation of positive feedback except for the negative sign. When $t = T$,

$$LEV(t=T) = .368\ LEV(0)$$

Hence, in a time interval equal to T, 63 percent of the initial value of the level has been removed.

A useful description of the rate of decay is the half-life. The half-life is defined as the amount of time that elapses for half the initial contents of the level to be lost. Mathematically, the half-life may be computed by solving for t in the level equation above by setting $LEV(t) = .5\ LEV(0)$. The doubling time of exponential growth is the analog of the half-life in exponential decay.

Initial Conditions

The role of initial conditions can be discerned from the rate-level graph (Figure 4). For a starting level value less than the goal, the system will respond by flow into the rate. As was demonstrated in the graphical simulation, the flow into the level will cease as the goal is reached.

When the initial value of the level is equal to goal, the system is already at its equilibrium point.

An initial value greater than the goal produces a system response exactly similar to the zero-goal system above. The initial value specifies a flow that must be out of the level. This action decreases the level until the goal is reached and equilibrium is obtained. The initial condition only determines from which direction the goal will be reached.

System Compensation

There are many circumstances when the desired goal is not reached. For example, consider what the response and equilibrium value of the system would be for a step input rate. A constant rate such as curve C in Figure 7 that is applied at some arbitrary point in time and persists indefinitely is denoted as a step input.

Curve (a), RT1, in Figure 11 is the rate-level graph for the system without a constant input rate. Curve (b), RT2, is the exogenous constant input alone. The equations for the two rates are:[2]

R RT1.KL = FPT(GL-LEV.K)

---

2. Omission of a numerical value for a constant in the Dynamo equations in this paper means that any arbitrary number may be used.

R RT2.KL = CONST

C CONST =

If RT1 and RT2 are combined into a net rate NTRT, the following results:

R NTRT.KL = FPT(GL-LEV.K) + CONST

NTRT is plotted as curve (c) in Figure 11.

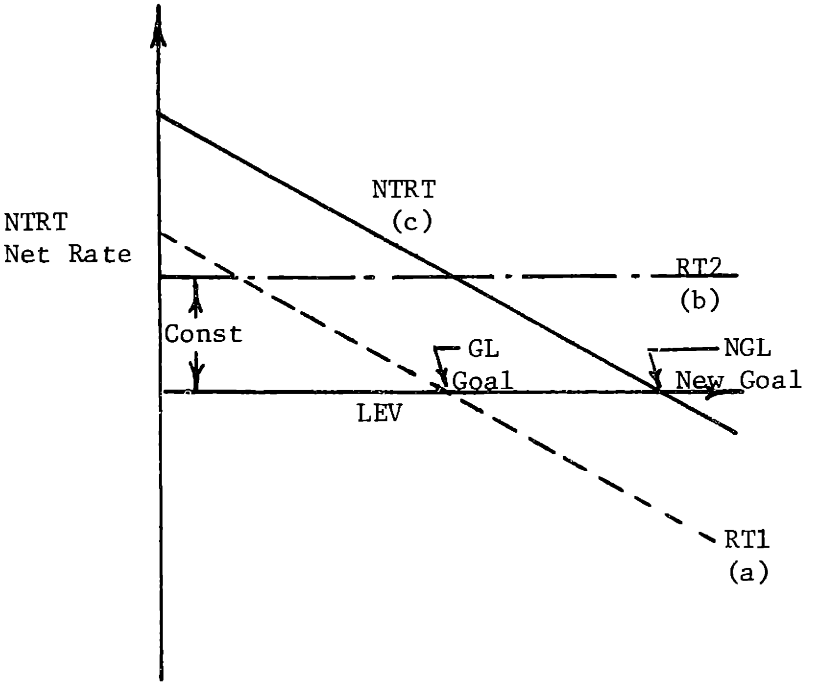


Figure 11

Constant Input to System--
Rate-Level Graph

For equilibrium to occur RT1 must be equal to RT2 or

-CONST = FPT(GL-LEV)

and the new goal NGL becomes

$$LEV = GL + CONST/FPT = NGL$$

In terms of the time constant T:

$$NGL = GL + (T)(CONST)$$

The new equilibrium value of the system is larger than the desired goal GL by the product of the time constant T and magnitude of the step input CONST. A system with a very rapid response (i.e., a small T) will have a new equilibrium value closer to the desired value for a given input than a system with a large T. For an input that is negative (i.e., a constant removal rate) the new equilibrium will be smaller than the goal by the same product.

Intuitively, what occurs when a constant input is applied can be seen in Figure 11. The input rate RT2 causes an initial increase in the level LEV. The new level value produces a negative outflow rate RT1 since LEV is now greater than GL. But RT1 is smaller than the input rate RT2 and there is a further net gain in the level LEV. LEV continues to increase but more slowly since NTRT is being reduced each time increment. Finally, the outflow rate RT1 compensates for the inflow rate RT2 and equilibrium is established. In reaching equilibrium, the system, however, has accumulated a net inflow rate NTRT and hence a higher equilibrium value. The level will exhibit the same behavior over time as the case of an initial value less than the goal without a step input. Conversely, a constant removal rate would produce behavior similar to an initial condition greater than the goal without a step input.

Nonlinear Negative Feedback

The discussion above has so far pertained to linear systems. It is appropriate to briefly touch upon the more general case of simple non-linear negative feedback. The rate-level curve is extremely helpful in analyzing the behavior of nonlinear relationships.

Two of many possible nonlinear relationships are shown in Figure 12. The only restrictions that are imposed are done so to preserve the goal seeking definition of negative feedback. It is assumed that negative values of the level are meaningless and the curve intersects the horizontal axis at least once. The curves shown might result from a varying time constant.

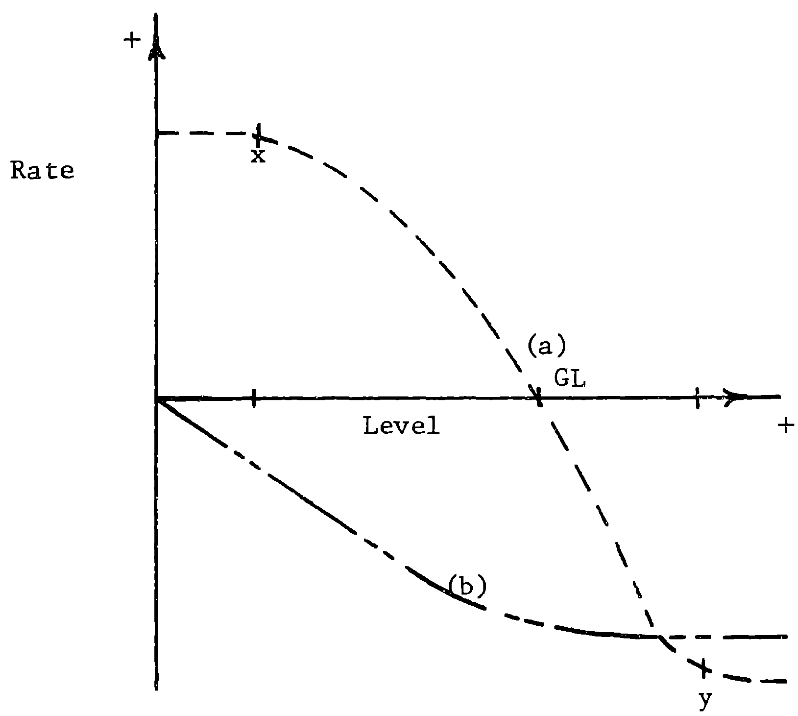


Figure 12
Nonlinear Negative Feedback

Curve (a) is characterized by flat portions representing threshold or saturated regions. Before point x is reached, a constant rate persists. This results in the linear growth of the level with respect to time. When x is reached, the rate begins decreasing as in the linear system until GL. For initial values greater than y, the system behavior would be exactly opposite. A constant flow out of the level would eventuate until y was reached. A decreasing outflow rate from y would bring the level back to GL.

Curve (b) is similar to curve (a) except that the rate curve passes through zero indicating a zero-goal value. Forcing the level into the saturated regions of either curve (a) or (b) such as by a constant input rate can convert the system from a goal seeking to a goal diverging system. This will be explored in detail in Example 3.

Summary

Negative feedback is a process tending to keep a given system at equilibrium. Discrepancies that occur are offset by internal pressures or forces which attempt to restore the system to its status quo. A simple model of a negative feedback system is found in the flow diagram of Figure 3. Information about the level is compared to the exogenously determined desired goal. The inflow or outflow rate is adjusted appropriately until the level has reached the desired value and the discrepancy is zero. The point at which the rate-level relationship intercepts the level axis is the point of stable equilibrium. The time constant, the inverse of the slope of the rate-level plot, is a measure of the rapidity with which the system reacts to external changes. The initial conditions determine whether the initial flow will be inward or outward to achieve the goal.

Three examples of simple negative feedback systems will now be examined.

Example 1--Inventory Control System

Inventory control is an example of a simple feedback system. Consider a dealer who would like to maintain a desired level of inventory. When his stock of goods falls below the desired level, he places orders to the distributor to replenish his supply. He ceases ordering when his stock has once again been built up. Facing a condition of too much inventory, if the option is available, he would send the excess back to the distributor. Determining how much to order, and when, may be completely automatic or it may be sporadic as the need arises. In this scenario a widely accepted ordering policy is recognized which contains the basic elements of a negative feedback system. The flow diagram for the inventory control system is produced in Figure 13.

Sales, which deplete the inventory INV, are dependent on market conditions. It is assumed that the dealer has little or no influence on the demand for his goods. Depending on the nature of the product, this assumption may or may not be valid. For example, a dealer might promote sales by reducing the price of his goods. While such causal links could be accommodated in the model, we shall simply assume sales are exogenously determined.

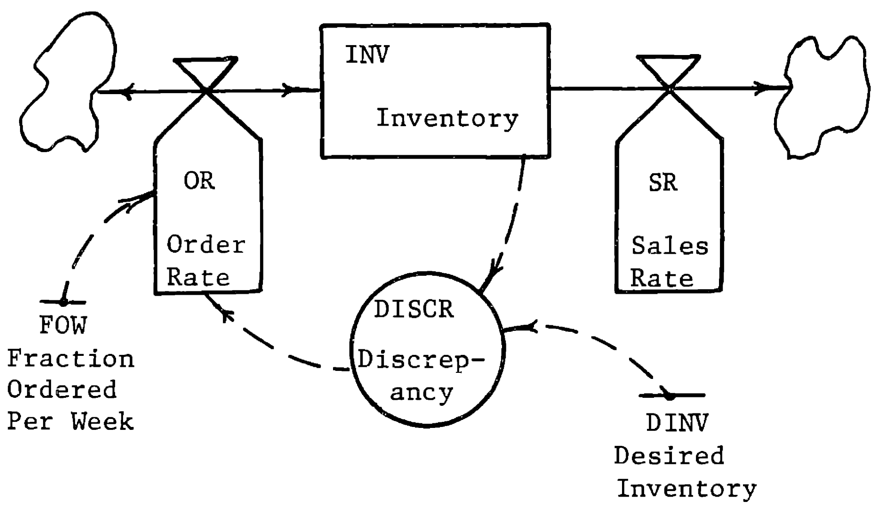


Figure 13
Simple Inventory Control System

The order rate OR replenishes the inventory stock. It is dependent on the ordering policy of the dealer. Any number or combination of policies could be used. The flow diagram represents one fairly simple ordering scheme. The desired inventory DINV is determined from such factors as space limitations and overhead costs. When the inventory INV falls below the desired level, the dealer orders an amount equal to a fraction of the difference of the two. In the model the time required to make up an inventory shortage is two weeks, the time constant of the system. Or equivalently, on a weekly basis, half the shortage is made up the first week and half the second week. The fraction ordered per week FOW then becomes .5.

The simple model assumes that there are no material delays--delays in shipping or handling inventory goods. That is, once an order is placed it is immediately filled. Such would be the case where the dealer has ready and easy access to his supplier. A further assumption is the exclusion of information delays involved in determining the status of the inventory at any time. In other words, the dealer is assumed to know at all times the exact amount of his inventory. Such might be the case for a company keeping a running inventory (e.g., an automated accounting system). The assumption of no material delays and no information delays need not be made. It is employed in the model to provide a "best" case or ideal situation and to preserve the simplicity of the model.

The equations for the model in Figure 13 are given below:

```
L  INV.K    =  INV.J + (DT)(OR.JK-SR.JK)    (Items)
N  INV      =  DINV                         (Items)
C  DINV     =  200                          (Items)
R  OR.KL    =  FOW*DISCR.K                  (Items/Week)
C  FOW      =  .5                           (Fraction/Week)
A  DISCR.K  =  (DINV-INV.K)                 (Items)
R  SR.KL    =  STEP(20,4)                   (Items/Week)
```

The system is initially at equilibrium. The inventory INV is equal to the desired inventory DINV which is set at 200 units. Through use of the step function, a sudden rise in sales from zero to 20 items/week after four weeks is simulated. The performance of the simple ordering policy under these circumstances is desired. In particular, how well does the dealer do in maintaining his desired inventory?

From the simulation run in Figure 14, surprising behavior is witnessed in the level of inventory. The actual inventory INV is 20 percent less than the desired inventory DINV. The simple discrepancy ordering policy proves not to be adequate for maintaining a desired level of inventory. Figure 14 is a classic example of the general behavior described in the section on system compensation.[3] In that section it was demonstrated that for a constant input rate into the level, the final equilibrium value will not be the desired equilibrium. The final value will be greater or less than the desired value by the quantity equal to the product of the time constant and the input. For the example in Figure 13, the value is

$$\begin{aligned} INV &= -(1/FOW)(SR) + DINV \\ &= -2*20 + 200 \\ &= 160 \end{aligned}$$

From Figure 14, it is possible to see why this should occur. The difference between the sales rate SR and the order rate OR between week 4 and week 8 means that there is a net flow of goods <u>out</u> of the inventory. The net flow is reduced to zero once the order rate OR compensates for the sales rate SR. The total loss in inventory is represented by the cross-hatched area in Figure 14. The cross-hatched area is in fact equal to the product of the time constant (1/FOW) and the input (SR).

---

3. It is well known in the engineering control literature that a control scheme based on the error between the desired state and actual state of a physical system will never be adequate in the presence of a constant input. An error or discrepancy will always exist.

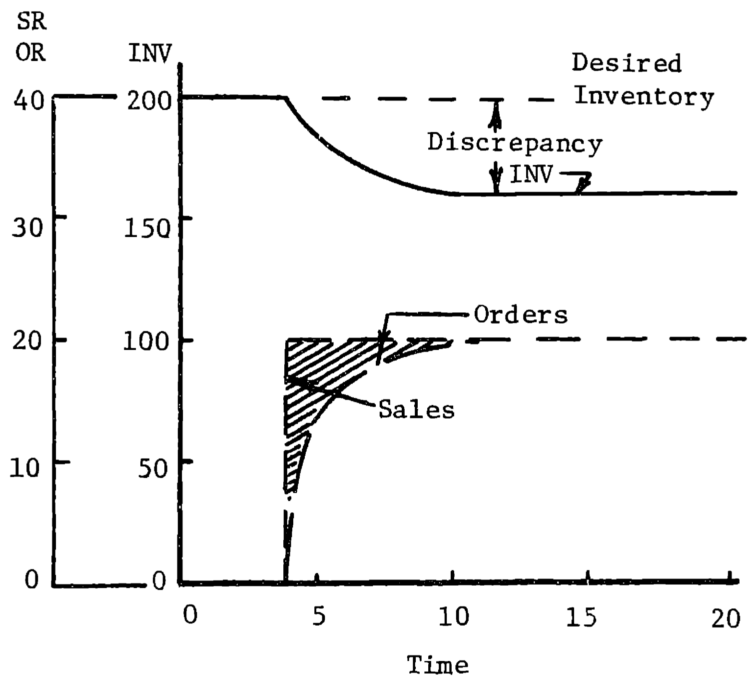


Figure 14
Inventory Control--Step Input in Sales

Unless the time constant of the system is reduced to zero (a highly unlikely situation), the dealer's ordering policy will never be able to maintain a desired level of inventory under constant or fluctuating sales conditions. Further, if informational and material delays were included, the disparity between the final inventory and the desired inventory would be even greater.

In this example the general first order negative feedback structure has been applied to a fairly common inventory maintenance system. Though certain assumptions were made concerning delays in the system, the simple structure yielded considerable insight. In particular, the compensatory nature of negative feedback was demonstrated for a system involving a constant input. The order rate rose to meet sales but at the expense of a lower equilibrium inventory.

Example 2--The Coffee Cup

An everyday first order feedback process with which most people are familiar is the cooling of a cup of hot coffee. This behavior belongs to a general class of thermal phenomena in which an object at one temperature is inserted or "dunked" into an environment at a different temperature. Involved are all the elements of a goal seeking system.

Heat flows between the environment (room) and the coffee cup. The net accumulation of the heat transferred to or from the cup determines the amount of heat contained in the coffee cup at any time. The temperature of the coffee is a measure of the amount of heat. The direction and amount of heat transfer is governed by a well-known physical law: the heat transfer rate HTR is proportional to the difference DISC between the ambient (room) temperature RTP and the coffee temperature CTP. The constant of proportionality C2 is dependent upon the physical properties of the coffee cup such as the volume and the insulation material involved. A model of the system is given in Figure 15.

The Dynamo equations are contained below. The initial temperature of the coffee is 200°F, while the room temperature is 78°F.

```
L  HT.K  =  HT.J + (DT)(HTR.JK)          (BTU)
N  HT    =  TI/C1                        (BTU)
C  TI    =  200                          (°F)
```

C	C1	=	1	(°F/BTU)
A	CTP.K	=	C1*HT.K	(°F)
A	DISC.K	=	RTP - CTP.K	(°F)
C	RTP	=	78	(°F)
R	HTR.KL	=	C2*DISC.K	(BTU/MIN)
C	C2	=	.1	(BTU/°F MIN)

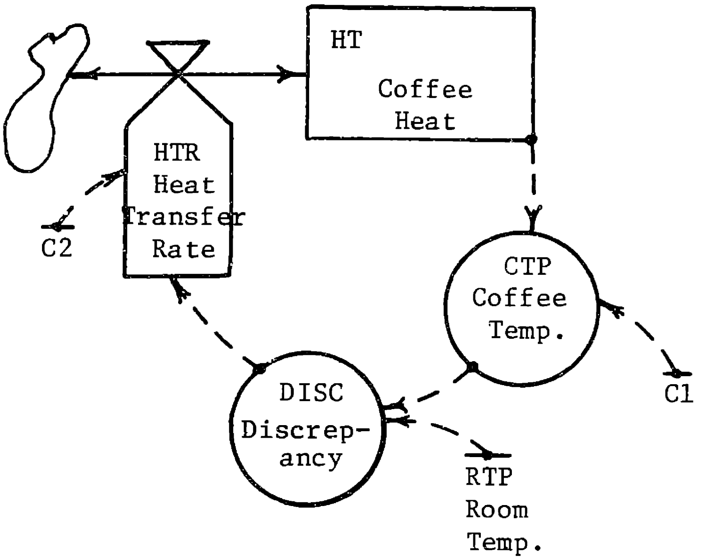


Figure 15

Flow Diagram
Coffee Cup

It is convenient to plot HTR as a function of CTP as in Figure 16. Since the coffee temperature CTP is proportional to the coffee heat HT in this example, this can be done easily. From Figure 16 it is seen that the initial condition of the coffee specifies that the heat flow must be out of the cup in order that equilibrium can be established between the cup and the environment. The coffee temperature CTP will exhibit exponential decay as it approaches the goal, the room temperature RTP. The simulation run of the temperature over time is contained in Figure 17.

The time constant of the coffee cup is the reciprocal of C2 or ten minutes in this example. Within ten minutes, the coffee temperature CTP will be

$$
\begin{aligned}
CTP(10) &= C1*HT(10) \\
&= 200 + .63(78-200) \\
&= 123°F
\end{aligned}
$$

Introducing a highly insulated coffee cup is tantamount to increasing the time constant by decreasing C2 in the model. A perfectly insulated coffee cup implies that the time constant is infinite and heat is prevented from dissipating from the cup. The initial temperature would be maintained indefinitely.

The coffee cup is an excellent example of a physical system that involves a decision process sector. The simple negative feedback structure yields considerable insight into the nature of a whole host of physical behavior involving a change in environmental conditions such as temperature.

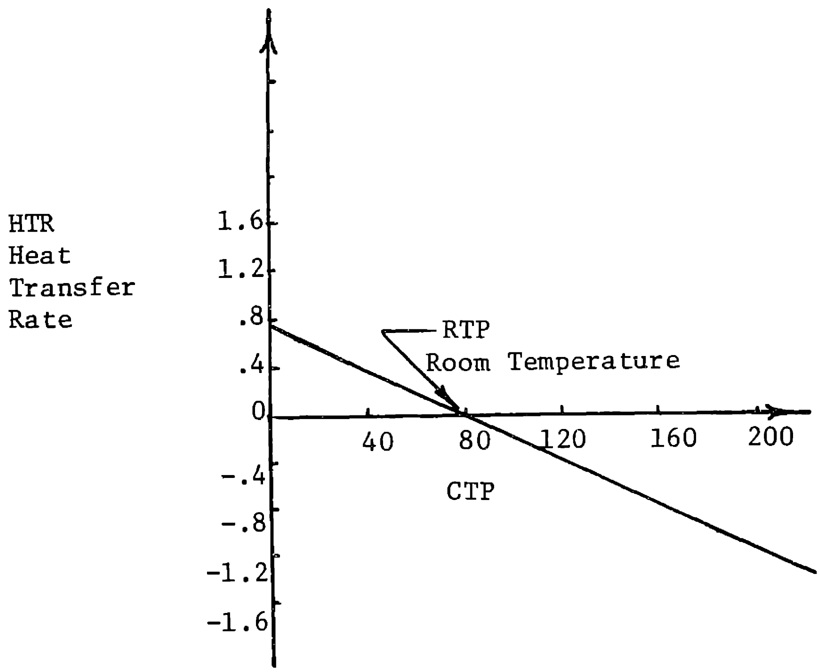


Figure 16

Rate-Level Graph

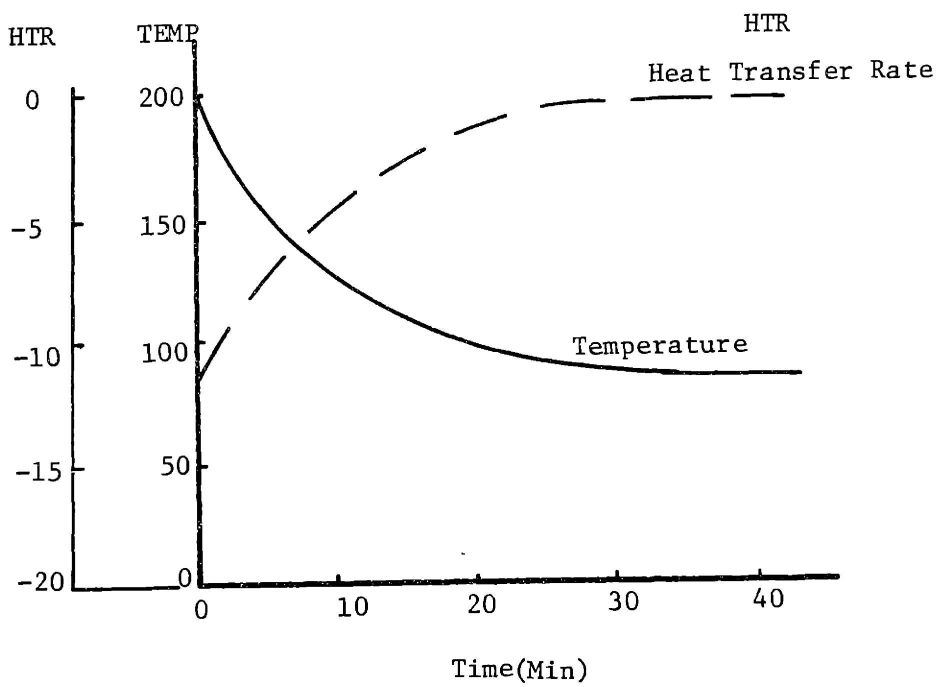


Figure 17

Coffee Dynamics

Example 3--Pollution Absorption

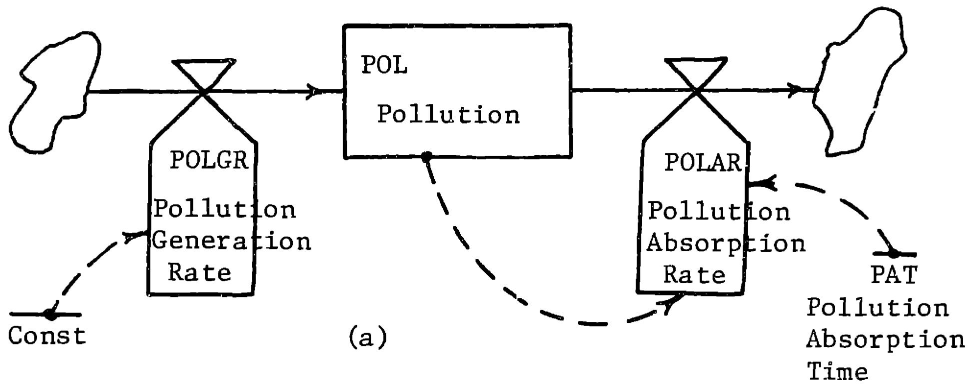
There are many simple feedback processes involving nonlinearities which are capable of converting a negative feedback process to a positive feedback process or vice-versa. In the next chapter, the latter case--a shift in dominance from positive to negative feedback--will be dealt with. In this example attention will turn to a special case of the former, pollution absorption.[4]

The basic pollution system as shown in Figure 18(a) contains the pollution level POL which is increased by pollution generation and decreased by pollution absorption. The pollution generation rate POLGR, like the sales rate in Example 1, is exogenously determined. The pollution absorption rate POLAR is a function of the amount of pollution present in the environment at any point in time. It is completely analogous to the negative feedback loop having a zero-goal common to many biological decay processes. The fixed time constant, in this case, the pollution absorption time PAT, yields a linear pollution absorption rate POLAR which is a function of the pollution level POL as in Figure 18(b).

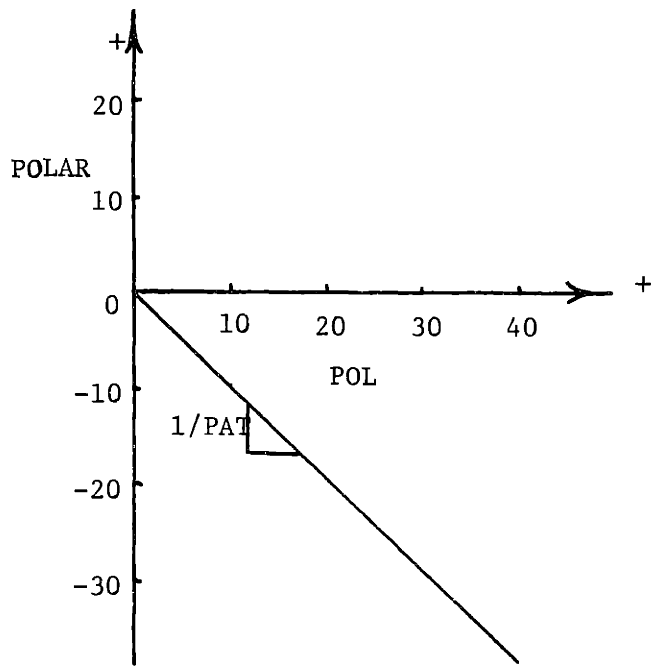
In the absence of pollution generation, any value of POL greater than zero would yield a proportionate value of POLAR according to the magnitude of the slope 1/PAT. In a recursive manner, POL and POLAR would be reduced until the entire initial amount of pollution were dissipated.

---

4. This example is based on the pollution sector of World Dynamics (Forrester 1971).



(a)



(b)

L	POL.K	=	POL.J + (DT)(POLGR.JK-POLAR.JK)	(Tons)
N	POL	=	0	(Tons)
R	POLGR.KL	=	CONST	(Tons/Yr)
C	CONST	=	5	(%/Yr)
R	POLAR.KL	=	POL.K/PAT	(Tons/Yr)
C	PAT	=	1	(Yrs)

Figure 18

Basic Pollution Model

Response of Basic Model to a Constant POLGR

If POLGR was active and equal to a constant, what would the final level of pollution equal? An arbitrary POLGR equal to CONST is combined with POLAR in Figure 19 generating a net pollution rate NPR versus POL graph. The equilibrium pollution level POL is established as seen in Figure 19 when POLAR is equal to POLGR.

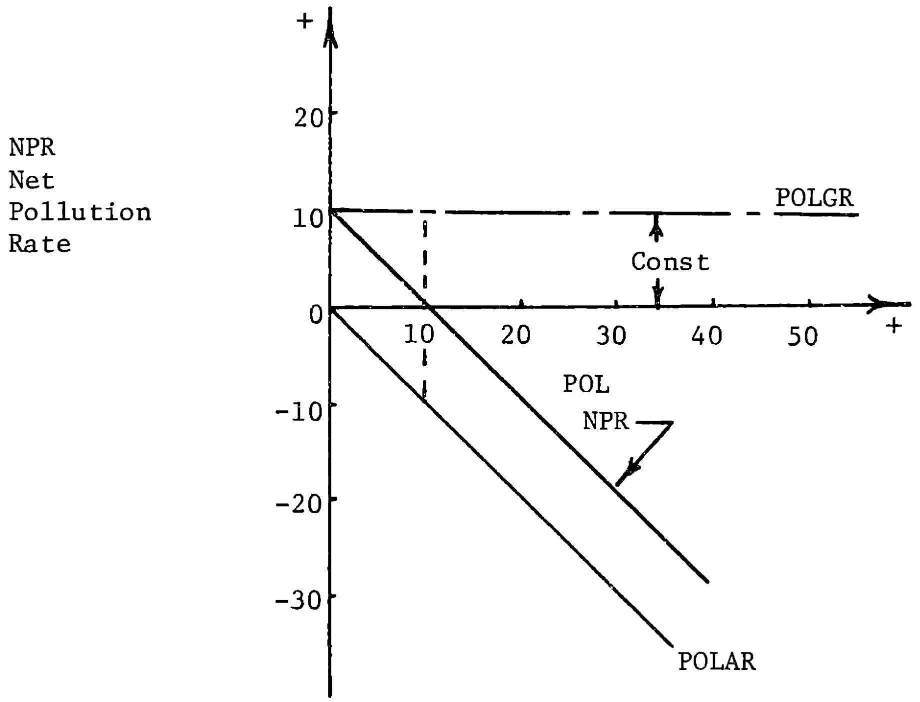


Figure 19

Net Pollution Rate
vs
Pollution

The equilibrium value can be found by setting POLAR to POLGR or

$$CONST = (1/PAT)(POL)$$

and

$$POL = (PAT)(CONST)$$

The result is not unexpected. As demonstrated previously (e.g., Example 1), applying a constant input rate to a nonzero goal structure produced a new goal value greater (less) than the original goal by an amount equal to the product of the time constant and input rate magnitude. Regardless of the magnitude of the constant input rate to the structure in Figure 18, however, equilibrium will be attained. The behavior of POL over time will emulate that of a nonzero goal structure as evidenced in Figure 20.

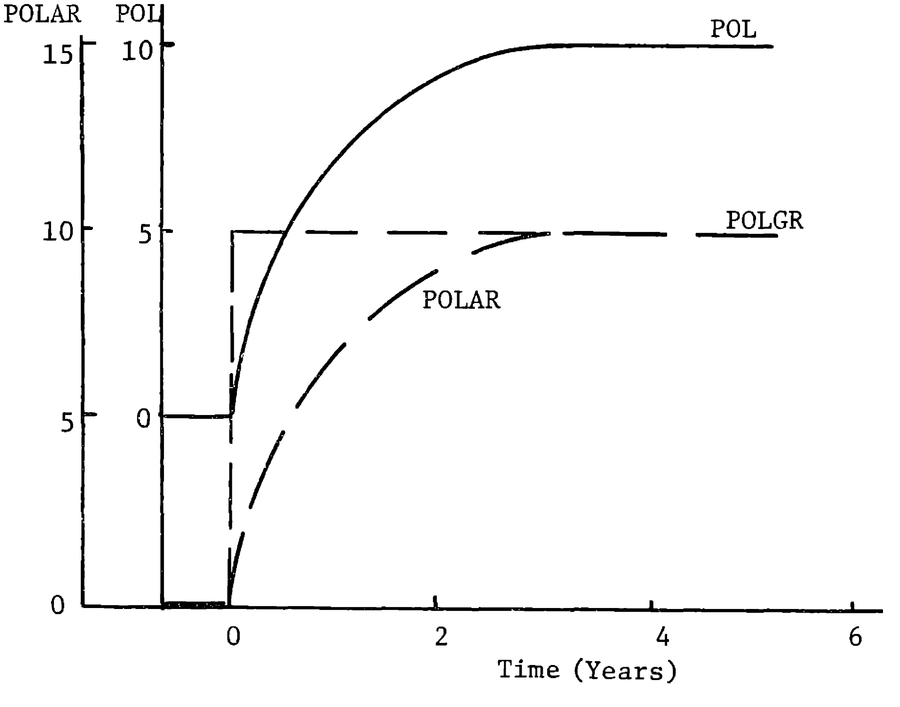


Figure 20

Response of POL to a Constant POLGR

Nonlinear Pollution Absorption Model

At high levels of pollution, the natural environmental clean-up processes are often inhibited or destroyed, depressing the amount of pollution that can be dissipated over a unit of time. The assumption of a constant pollution absorption time PAT does not seem to be valid at elevated levels of pollution. Hence making PAT a function of POL captures the "overloading" effect of high pollution on the dissipative capacity of the environment. A table function relating PAT to a pollution ratio POLR, the actual pollution level POL divided by a standard or definitional amount of pollution POLS (equal to one is this example), is given below in Figure 21. It is similar to the nonlinear relationship found in World Dynamics (Forrester 1971). The accompanying verbal description follows (Forrester 1971, pp. 57-58).

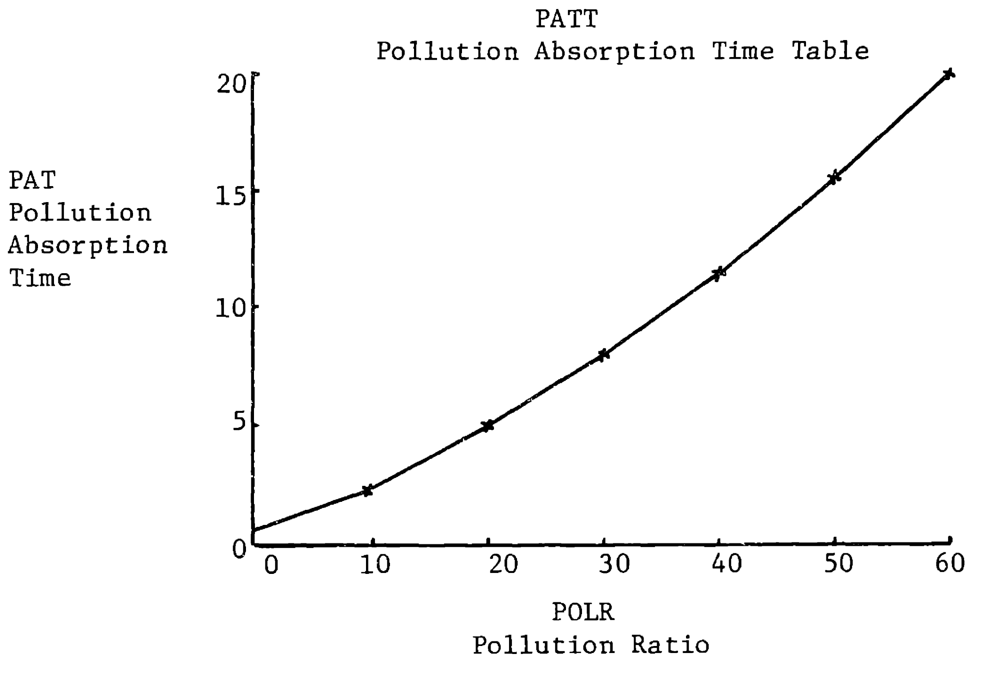
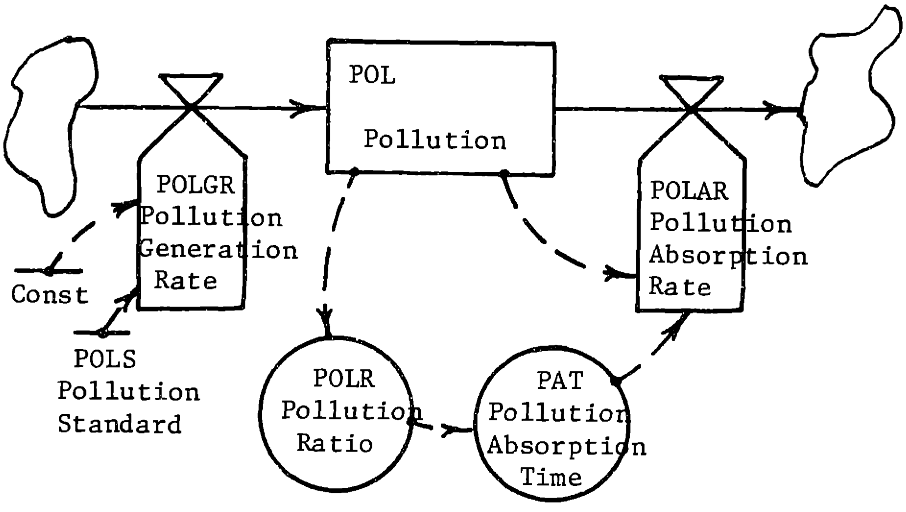


Figure 21

Pollution Absorption vs Pollution Ratio

> A pollution ratio POLR of 1 represents the conditions existing in 1970. A value for PAT of 1 year is taken for 1970. This means an assumption that under present conditions a year would be needed to dissipate about two-thirds of the existing pollution if all new pollution generation were to stop. For some of the polluting materials, that is too slow. On the other hand, one sees estimates that 90% of all DDT that has ever been manufactured is still in the environment. Certainly many kinds of pollution, probably including the more serious kinds, take longer than a year to disappear. A year is here used as an average. But as the amount of pollution increases, the pollution-absorption time is assumed to increase. This represents the poisoning and destroying of the pollution-cleanup mechanisms. Small amounts of pollution are dissipated quickly. But large amounts can have a cumulative effect by interfering with the natural processes of dissipation. Figure 3-15 [Figure 21] suggests that the decay time for two-thirds of existing pollution rises to 5 years for pollution levels, 20 times the 1970 values, to 10 years for a pollution increase of about 40 times, and to 20 years for 60 times the 1970 pollution. Such delay times are already observed. Many lakes may have become irreversible in their pollution or would recover only after times as long as shown in Figure 3-15 [Figure 21]. Estimates in the newspapers after the strike of sewerage-plant workers in England in 1970 gave estimates of 10 years for river life to recover to the condition it had before the excessive load of pollution.

The variable PAT is integrated into the basic pollution structure as shown in Figure 22(a). Plotting POLAR as a function of POL yields the curve in Figure 22(b). It has the same saturated form as curve (b) in Figure 12 followed by a decline in the magnitude of the absorption rate with increasing pollution. The direction of the slope changes from negative to positive. The negative feedback system is converted to a positive feedback system because of the increasing PAT. In the limit, PAT becomes infinitely large causing POLAR to become infinitely small (i.e., zero).



L	POL.K	=	POL.J + (DT)(POLGR.JK-POLAR.JK)	(Tons)
N	POL	=	0	(Tons)
R	POLGR.KL	=	CONST	(Tons/Yr)
C	CONST	=		(%/Yr)
R	POLAR.KL	=	POL.K/PAT.K	(Tons/Yr)
A	PAT.K	=	TABLE(PATT,POLR,0,60,10)	(Yrs)
A	POLR.K	=	POL.K/POLS	(Dimensionless)
C	POLS	=	1	(Tons)
T	PATT	=	.6/2.5/5/8/11.5/15.5/20	

Figure 22(a)

Complete Pollution Model

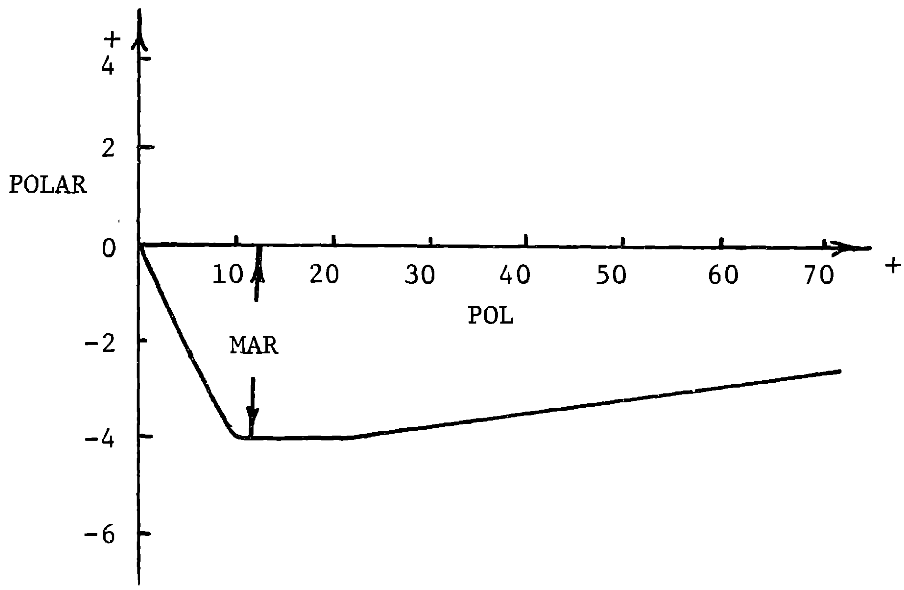


Figure 22(b)
Complete Pollution Model

Response of Complete Model to a Constant POLGR

How does the behavior of the nonlinear pollution model compare to the linear model? As shown in the net pollution rate NPR graph of Figure 23, much depends upon the size of the pollution generation rate POLGR. It is assumed that the initial value of the pollution level POL is zero. The net pollution rate NPR will cross the abscissa in a manner that is consistant with a negative feedback system provided CONST is less than the maximum absorption rate MAR. A value of CONST greater than MAR will yield a net pollution rate NPR that does not intersect the horizontal axis.

That is, equilibrium can never be achieved. POL will exhibit an initial amount of exponential decay followed by linear growth. As POLAR passes through the saturated region and begins to decrease in magnitude with increasing POL, exponential growth will result. Eventually, POLAR becomes zero and POL will continue to grow linearly without limit. The simulation runs in Figure 24 verify the analysis.

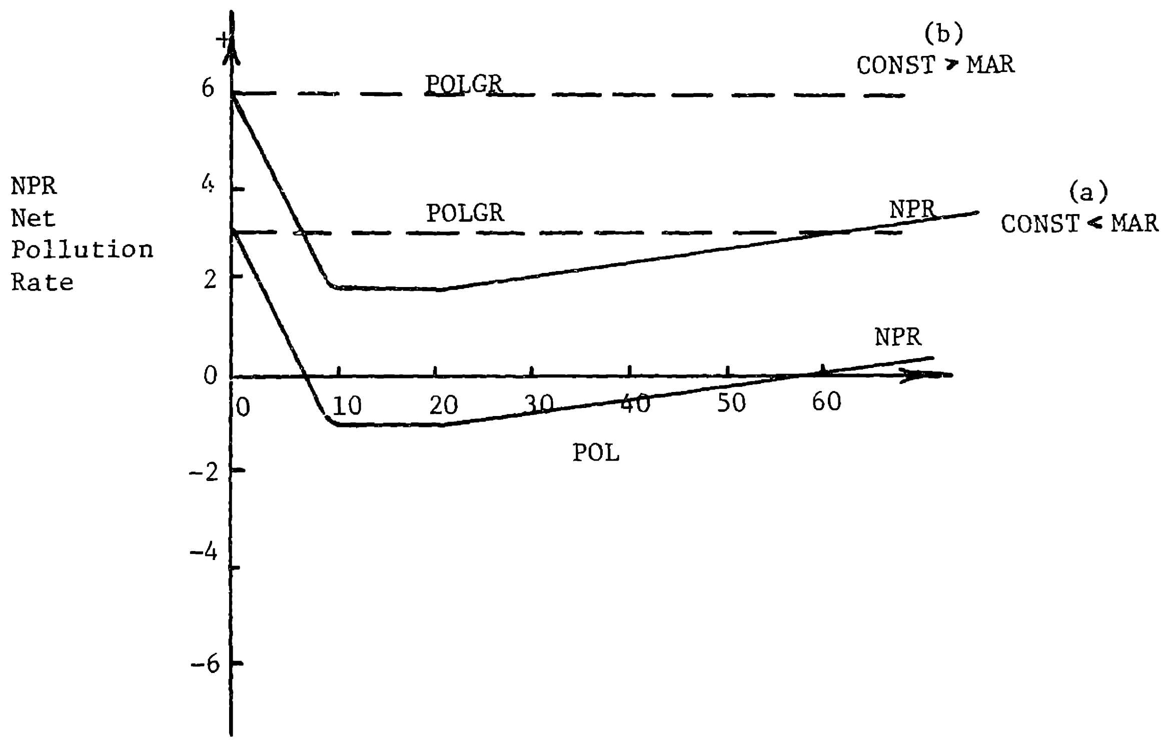


Figure 23

Net Pollution Rate vs
Pollution Level

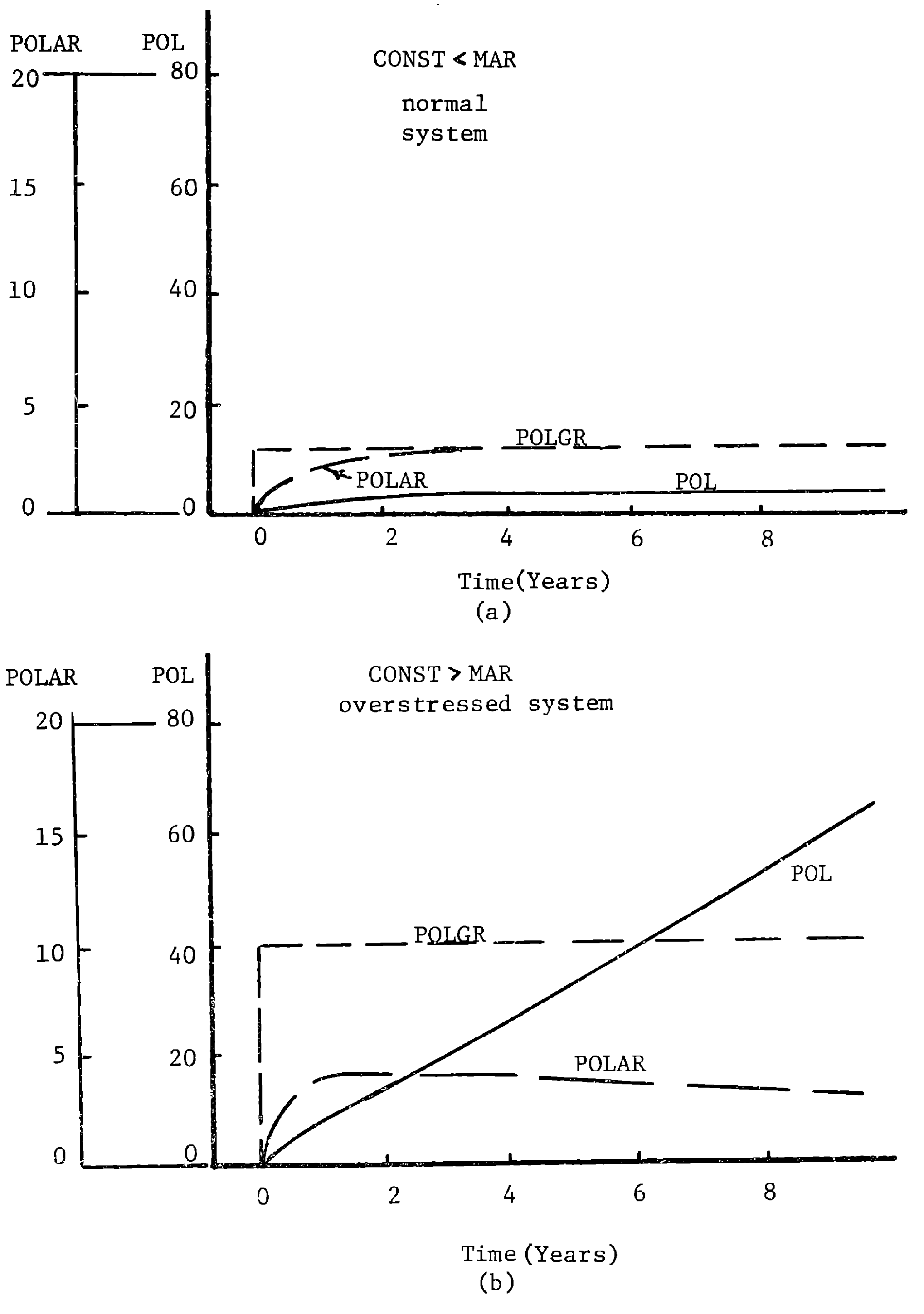


Figure 24

Response of POL to a
Constant POLGR for Complete Model

The complete model can produce two different modes of behavior because of the nonlinear rate-level structure. Run (a) in Figure 24 is comparable to the run in Figure 20. Run (b), the overstressed system, substantially differs in behavioral response to the same value of the constant input used in the run of Figure 20.

The pollution dissipation system illustrates the effect a variable time constant (and hence a nonlinear rate) has on the behavior of a zero-goal, negative feedback structure. A constant pollution absorption time PAT implied that regardless of the magnitude of the constant input, an equilibrium level of pollution resulted. The simulation run of Figure 20 affirmed this. The linear model was altered to account for the saturation and decline of the pollution absorption rate at high pollution levels. At low input rates, pollution accumulated but eventually equilibrated as in the linear system. However, at high input rates, the system's dissipative capacity was exceeded. The inability of the system to compensate for the input and the resultant unrestrained linear growth of the level is typical of many systems stressed beyond their tolerance.

Chapter 3
S-shape Growth Structure

Introduction

A large class of behavioral phenomena seems to combine both exponential growth and exponential decay in a way such that a third type of growth results. The S-shape curve in Figure 25 is an example of this.

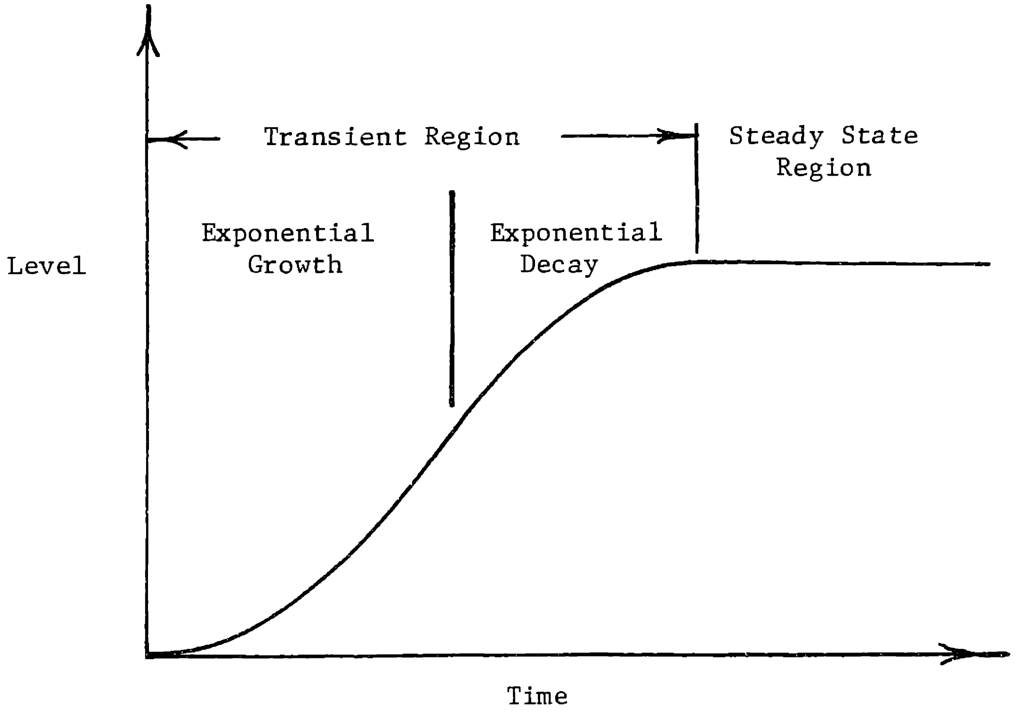


Figure 25
S-shape Growth

S-shape growth, often referred to as logistic growth or sigmoidal growth, is typified by three distinct phases--two in the transient region, followed by a steady state region. The first phase involves exponential growth. It is followed by exponential decay, eventually leading to the last or equilibrium phase when growth has ceased. The reader should note the qualitative difference between the growth shape in Figure 25 and the pure exponential decay shape developed in Chapter 2, that is, the additional exponential growth phase in the transient region.

Many growth patterns are similar to the curve in Figure 25. The most commonly found are the population growth curves of various plants and animals. In fact, the strong resemblance of the curve to laboratory and field experiments involving population growth has given rise to an analytical model, the logistic equation, as the fundamental mathematical description of population growth in the ecology literature (Odum 1971).

Many ecologists concerned with the human population are urging that the present trend of exponential growth be diverted into an equilibrium phase (Meadows 1972). Ideally, they would like to see mankind adopt a growth strategy similar to the logistic growth observed in the rest of the biological world, as sketched in Figure 26.

In Figure 26, human population is plotted as a function of time. The solid line denotes past and projected growth, while the dotted line is the advocated equilibrium growth shape.

The sigmoidal trend is also characteristic of the growth of the individual within a population. For the human being as well as a number of higher vertebrates this includes both physical as well as mental growth and development. For example, the S-shape curve is proposed as the form typical of certain types of "learning curves."

A whole host of social activities seem to resemble the sigmoidal shape in part if not in whole. Urbanization is a classic example. The land area growth of four empires has been observed as following the sigmoidal shape (Taagepera 1968). The life cycles of diffusion phenomena in a fixed population, such as the spread of riots, rumors, news, and epidemics, are further examples.

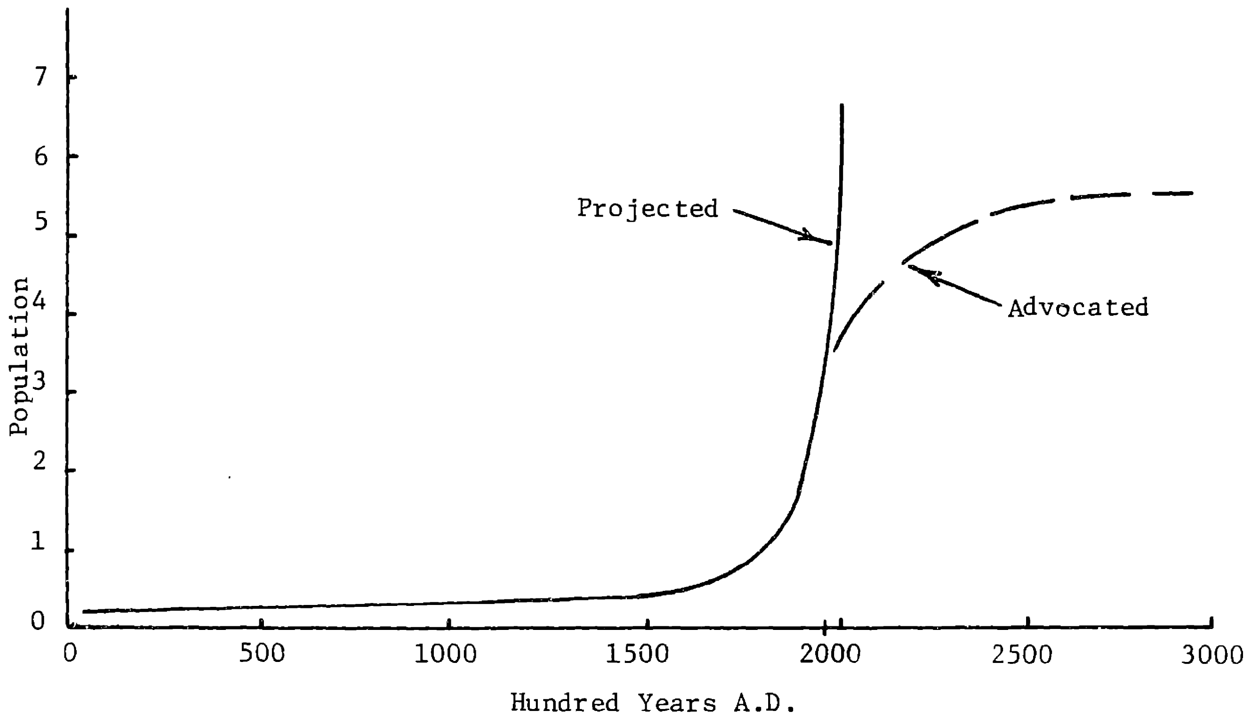


Figure 26
Human Population Growth

This chapter synthesizes from the previous development of positive and negative feedback the structure involved in S-shape growth. The question that will be addressed is, "What must occur in a feedback system for the transition from exponential growth to exponential decay to be made?" This chapter will also demonstrate how one utilizes and builds upon simple structures to produce more complex behavior. Doing such will aid the modeler in the reverse process as well: synthesizing simple structures from real world phenomena. For example, given a behavioral mode such as logistic growth, what are the important structural relationships in the actual system capable of producing the behavior? Knowledge of the simple structure enables the modeler to focus his attention on the strategic elements in the real world system while filtering out the less important ones.

S-shape Growth Structure

As mentioned above, the growth in Figure 25 seems to combine both positive feedback and negative feedback. A possible combination of the two is represented in the causal diagram of Figure 27, where A is the variable of interest.

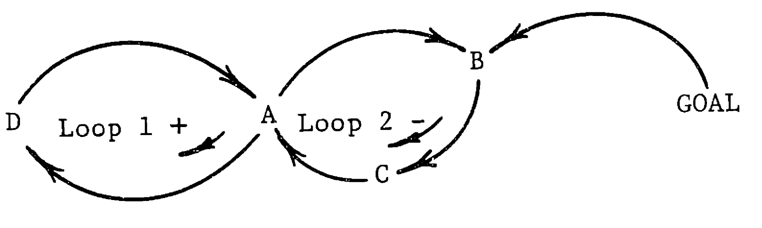


Figure 27

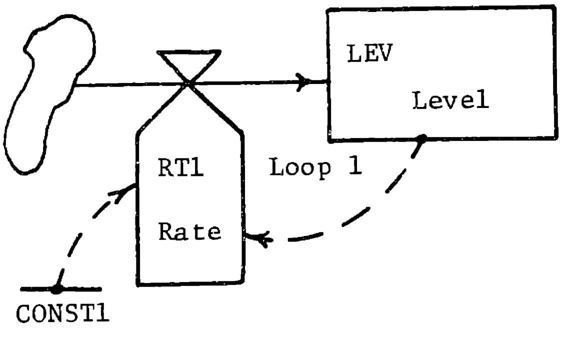
Causal Diagram

Loop 1, the positive feedback loop, attempts to force A to continually increase consistent with positive feedback. A goal is assumed not to exist in the loop. Loop 2 is equivalent to the general causal loop of negative feedback. Its purpose is to attempt to maintain variable A at a desired value, which could be any value. The diagram above, however, does not contain enough information to indicate the behavior of the system.

Flow Diagram of S-shape Growth

One obvious approach to modeling the structure involved in S-shape behavior would be to go directly from the causal diagram above to the flow diagram based on the linear rate and level relationships of positive and negative feedback. A brief review of those relationships is presented in Figures 28 and 29.

Plotting the rate equations in the figures with respect to values of the level LEV, yields the graphs of Figure 30. In Figure 30(a) as the level LEV increases the rate RT increases proportionately, producing, when integrated over time, a still higher value of the level LEV. This in turn produces a larger rate value which eventually yields the exponential growth curve. The opposite occurs in negative feedback. Because of the presence of a goal GL in the formulation, the rate RT is at its maximum value when the level LEV equals 0. As the level LEV increases, the rate RT decreases until the level LEV reaches the goal value GL and the rate RT becomes zero. No further growth is possible. Notice that for both Figures 30(a) and 30(b) the slopes or constants of proportionality CONST1 and CONST2 do not change.



L LEV.K = LEV.J + (DT)(RT1.JK)

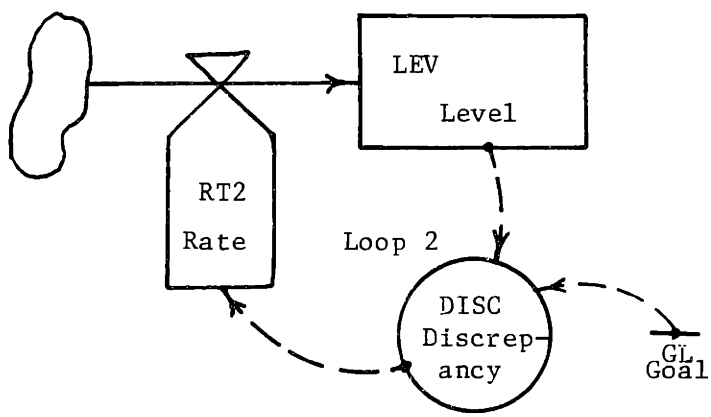
N LEV =

R RT1.KL = CONST1*LEV.K

C CONST1 =

Figure 28

Positive Feedback Structure




```
L  LEV.K    =  LEV.J + (DT)(RT2.JK)
N  LEV      =  0
R  RT2.KL   =  CONST2*DISC.K
A  DISC.K   =  GL - LEV.K
C  GL       =
C  CONST2   =
```

Figure 29

Negative Feedback Structure

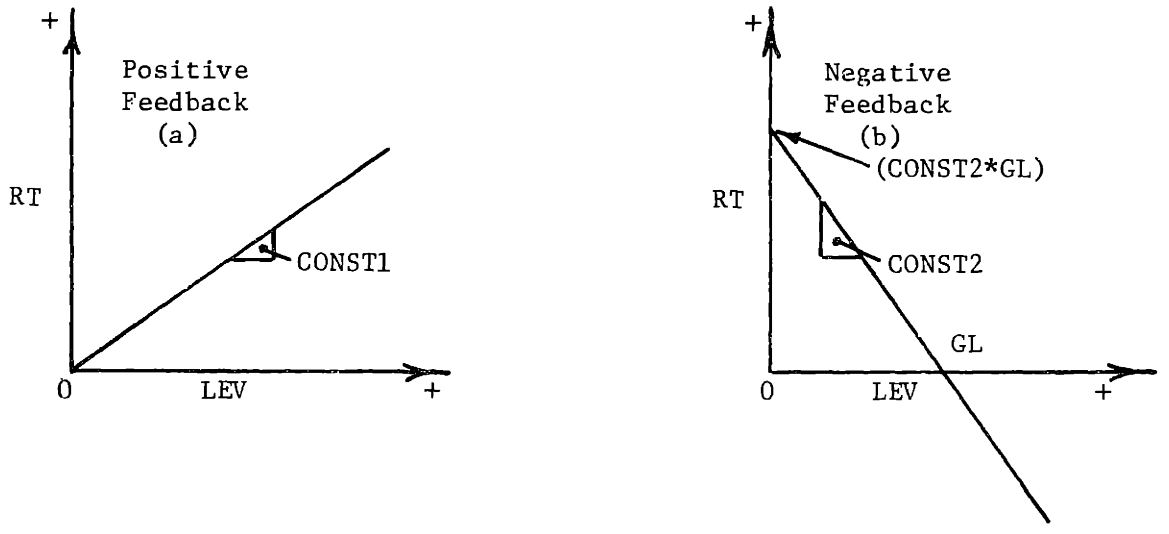
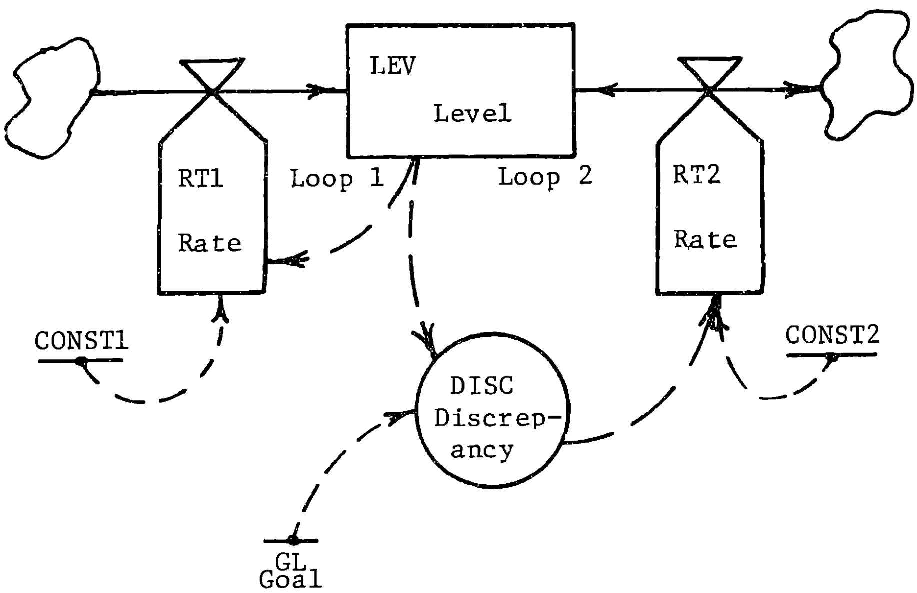


Figure 30
Rate-Level Graphs

Since S-shape growth seems to combine both the mechanisms of positive and negative feedback, one could hypothesize a rate-level structure as seen in Figure 31. The assumption that the slopes CONST1 and CONST2 do not vary will be preserved.

An approach to analyzing the behavior of the system in Figure 31 is to combine RT1 and RT2 algebraically, forming a net rate NTRT. The plots of the net rate NTRT versus the level LEV are seen in Figure 32. Curves (a), (b), and (c) result for goal GL values not equal to zero. Three possible modes of behavior can result, depending upon the relative values of the constants. For CONST1 values greater than CONST2, the resulting level values would follow the exponential growth shape. For CONST1 values less than CONST2, negative feedback prevails and the result is the same decay behavior seen in Chapter 2. When CONST1 is equal to CONST2, the NTRT is constant and equal to GL*CONST2. A constant NTRT value yields the ramp shape also discussed in Chapter 2. Figure 33 summarizes the performance of LEV over time for GL values not equal to zero.




```
L  LEV.K    =  LEV.J + (DT)(RT1.JK+RT2.JK)
N  LEV      =  0
R  RT1.KL   =  CONST1*LEV.K
R  RT2.KL   =  CONST2(DISC.K)
A  DISC.K   =  GL - LEV.K
C  CONST1   =
C  CONST2   =
```

Figure 31

Additive Positive-Negative Feedback

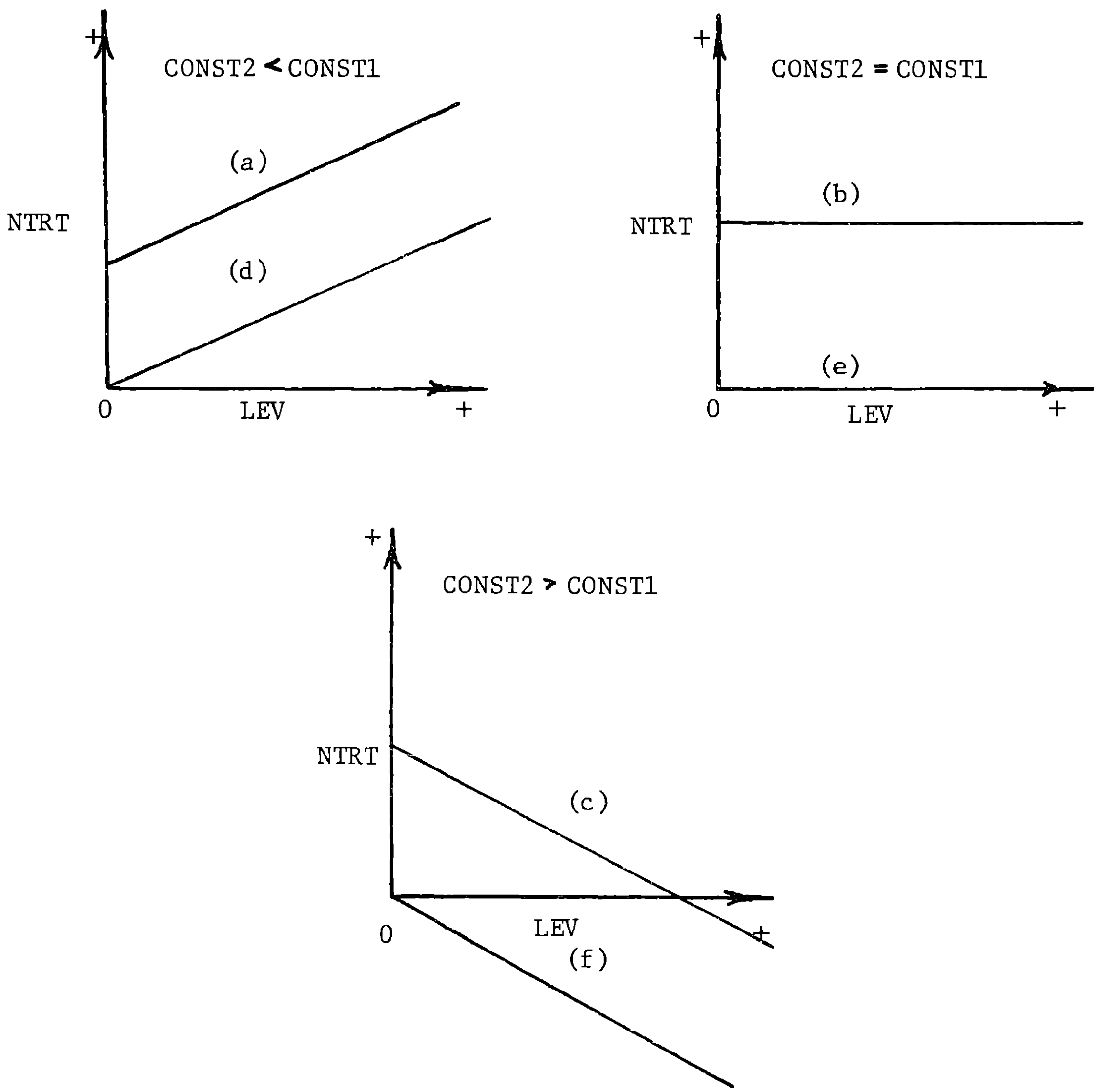


Figure 32

Additive Rate-Level Graphs

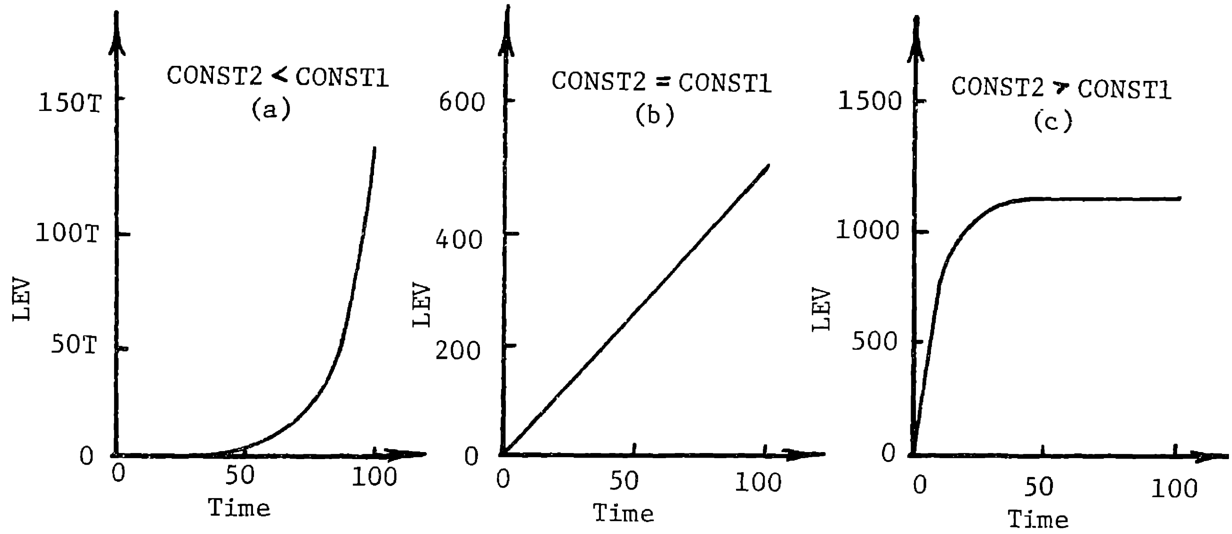


Figure 33
Combined Positive-Negative Feedback Output

The NTRT curves for a goal GL equal to zero are shown in Figure 32 (curves 32(d), (e), (f)). The difference between curves (a), (b), and (c) and (d), (e), and (f) are that the latter must pass through the origin. It is evident that the resultant time shape of the level LEV from curve (d) will be similar to that in Figure 33(a). For curve 32(e) instead of linear growth, the level LEV will remain at its initial value. For curve 32(f) equilibrium will occur when the level LEV is at zero in a fashion similar to first order negative feedback.

No matter what the values of the slopes CONST1 and CONST2 or the goal GL are, the structure in Figure 31 is unable to produce the desired sigmoidal behavior. There is in fact, no way that a linear relationship can produce S-shape growth for a first order system.

A possible alternative is to discard the assumption of constant slopes and investigate nonlinear rate-level relationships. Figure 34 is an example of a nonlinear association between the net rate NTRT and the level LEV. The slope of the curve in Figure 34 between zero and IL is equal to CONST1 while it is equal to CONST2 after IL. At IL the slope is undefined, or discontinuous.

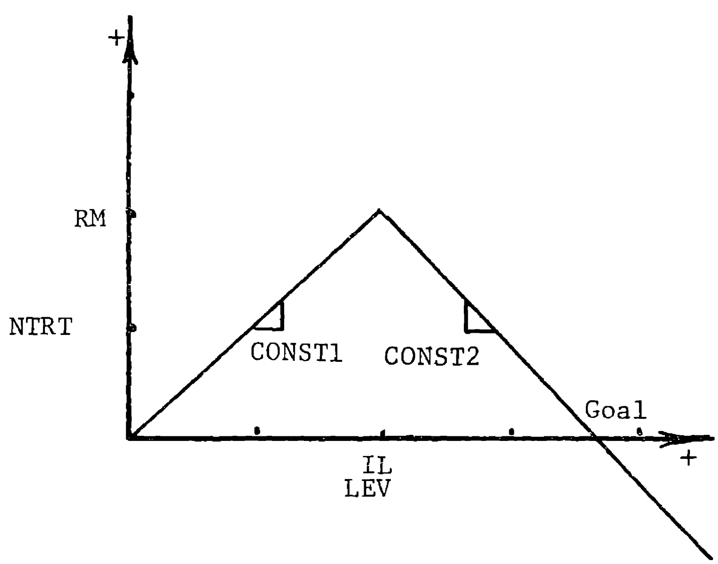
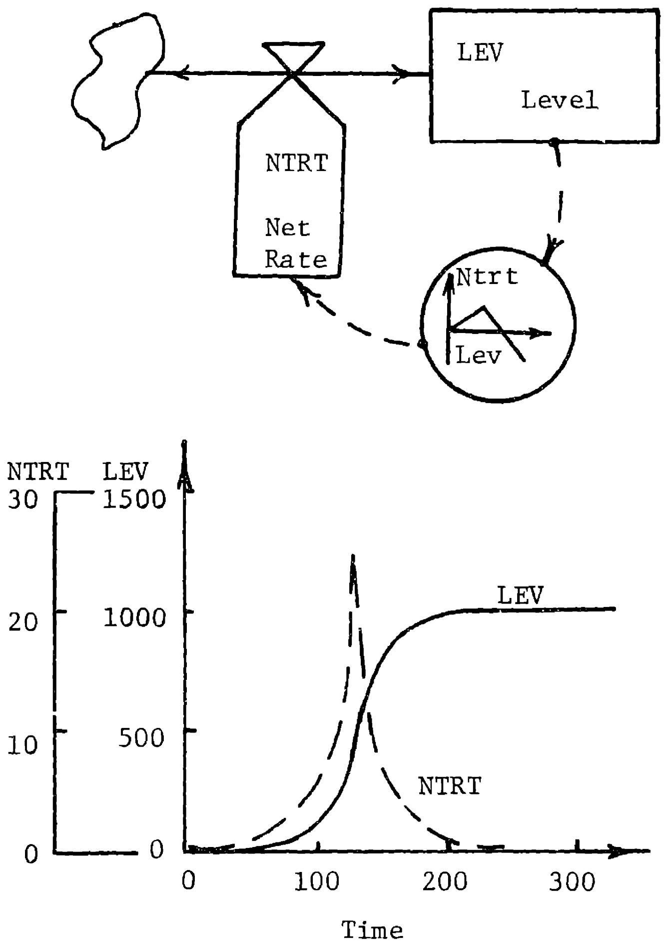


Figure 34
Nonlinear Rate-Level Graph

From Figure 34, it is seen that an initial value of the level LEV between zero and IL would yield a proportionate value of the net rate NTRT. The NTRT value when accumulated over a small interval of time and added to the initial value of the level LEV yields a new and larger value of the level LEV. This new value of LEV produces a still higher value of NTRT. Exponential growth will occur until inflection level IL is reached and the maximum rate RM is achieved. The next value of LEV produces a proportionately smaller value of NTRT. A transition from positive to negative feedback has occurred.

As in a purely negative feedback system, when the values of the NTRT eventually reach zero, the level will obtain its equilibrium value, goal GL. The simplified flow diagram is seen in Figure 35, together with the simulation run of the level based on a table function similar to Figure 34. Sigmoidal growth has been modeled.

The rate-level relationship in Figure 34 is an abstraction and merits further discussion. An abrupt change in slope at IL does not occur often in reality. More continuous shapes as in Figure 36 are found, as will be illustrated later. Notice that the slopes in Figure 36 are varying. The slope in 36(a) between L1 and L2 rapidly changes values. In Figure 36(b) the slope is constant only until L2 is reached at which time it begins to change continuously. In Figure 36(c), the slope continuously varies. Regardless of the exact shape, if a general trend exists similar to those found in Figures 34 and 36, the resultant level behavior will be sigmoidal.



L LEV.K = LEV.J + DT(NTRT.JK)

N LEV = 0

R NTRT.KL = TABLE(NTRTT,LEV.K,0,1200,100)

T NTRTT = 0/5/10/15/20/25/20/15/10/5/0/-5/-10

Figure 35

S-shape Growth Structure

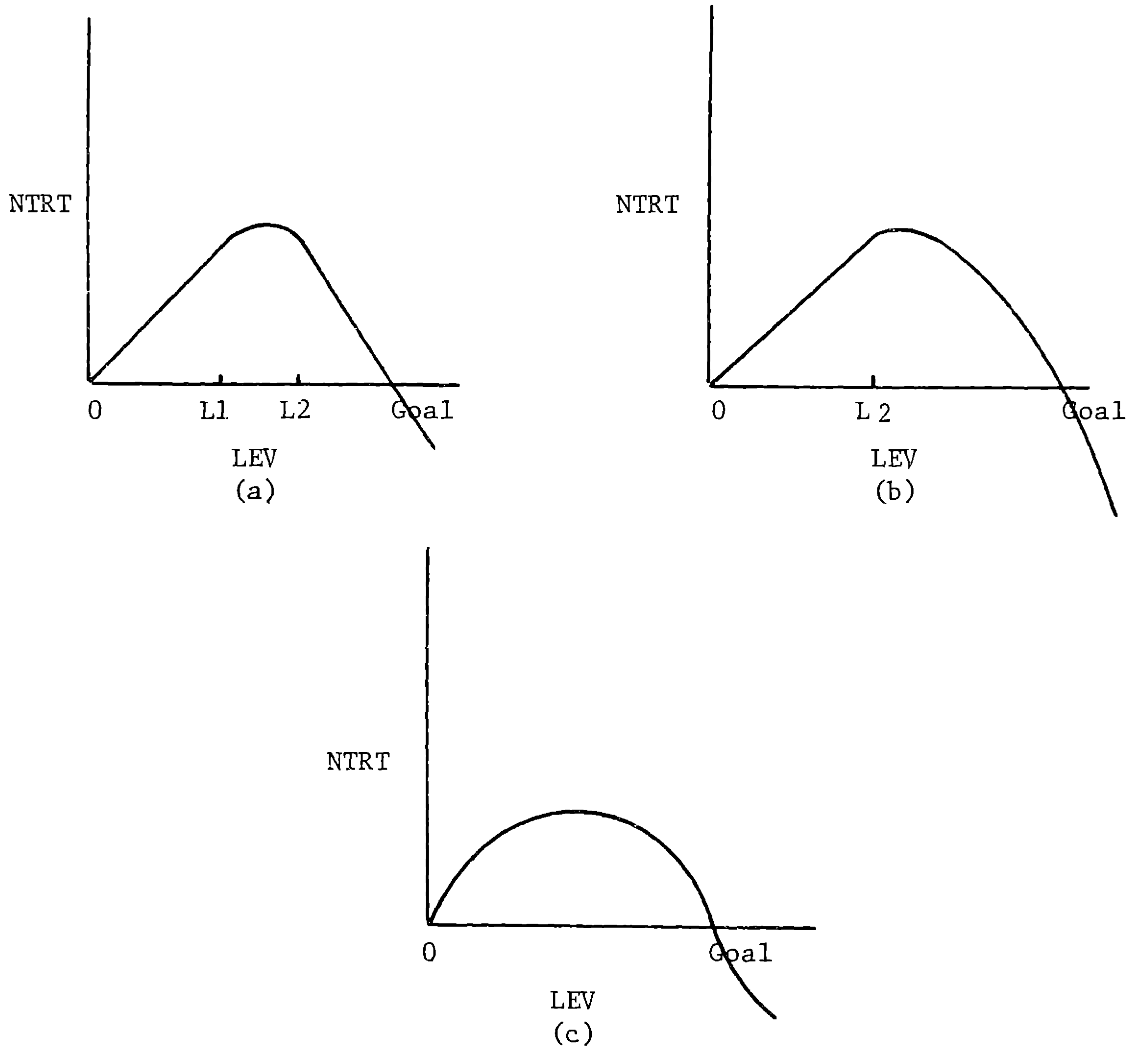


Figure 36

Continuous Rate-Level Graphs

Stable and Unstable Equilibrium

At this point it is feasible to take up the question of the equilibrium value that the level LEV obtains in sigmoid growth. The meaning of stable and unstable equilibrium will be explored.

The goal GL, termed stable equilibrium, will persist indefinitely. No further increase or decrease in the level LEV will occur when it is reached. If the value of the level LEV in Figure 34 were increased by some outside influence such that LEV exceeded GL, for example, a negative NTRT would return the level LEV to its stable equilibrium point. If the value of the level LEV were altered such that LEV fell short of the goal GL growth would occur until GL was once again reached. Seen below are simulation runs where the level LEV is exogenously changed from its equilibrium state by an addition of ten units (Figure 37(a)) and a decrease of ten units (Figure 37(b)).[1] The characteristic goal directed behavior of negative feedback is observed.

Figures 37(a) and 37(b) are graphic representations of the internal pressures (or forces) which arise within the system through the processes of negative feedback in order to restore desired steady-state conditions.

---

1. Exogenous alteration of the level is accomplished by use of the pulse function added to the rate equation where:

   R NTRT.KL = TABLE(NTRTT,LEV.K,0,1200,100) + PULSE (HT,TIME.K, INTVL)

   HT = height of pulse

   TIME.K = time of first pulse

   INTVL = interval between pulses

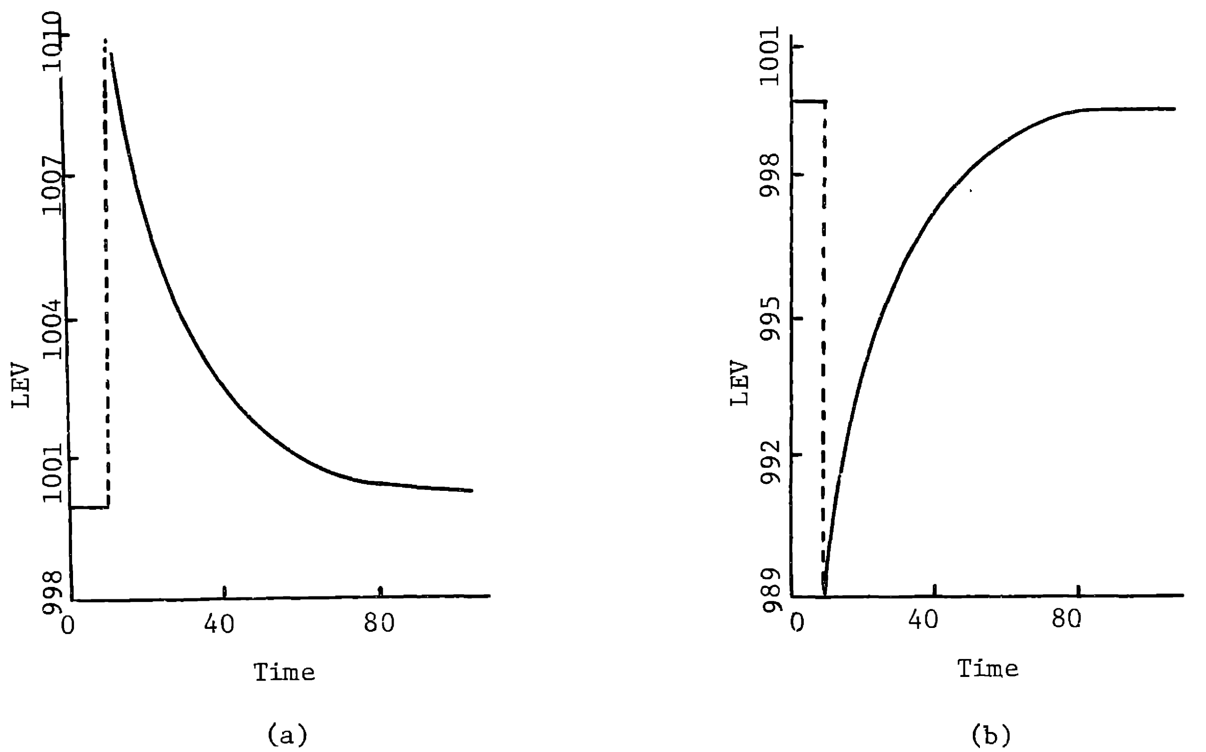


Figure 37
Ten Unit Impulse Response

On the other hand, prior to growth--that is, when the level LEV value is zero--the system is said to be in "unstable equilibrium." If the value of LEV were increased only minimally, such as by an initial condition greater than zero, the net rate NTRT would no longer be zero.[2] The system level would increase as seen in the general S-shape. The system's inability to maintain equilibrium at zero for the slightest perturbation is the reason for the name unstable equilibrium.

---

2. We assume that only positive values of the level LEV are possible.

It is appropriate to investigate what would happen if the slope CONST1 in Figure 34 were changed so that the curve did not pass through the origin, but instead crossed the vertical axis at a value greater than zero (i.e., the curve has a new slope CONST1' which is less than CONST1 and similar to the one in Figure 32(a)). Essentially, the possibility of unstable equilibrium is eliminated, since the NTRT can never be zero except at stable equilibrium. The system is able to generate the S-shape behavior even if the initial value of the level LEV is zero. Figure 38 shows the simulation run with the new slope CONST1'.

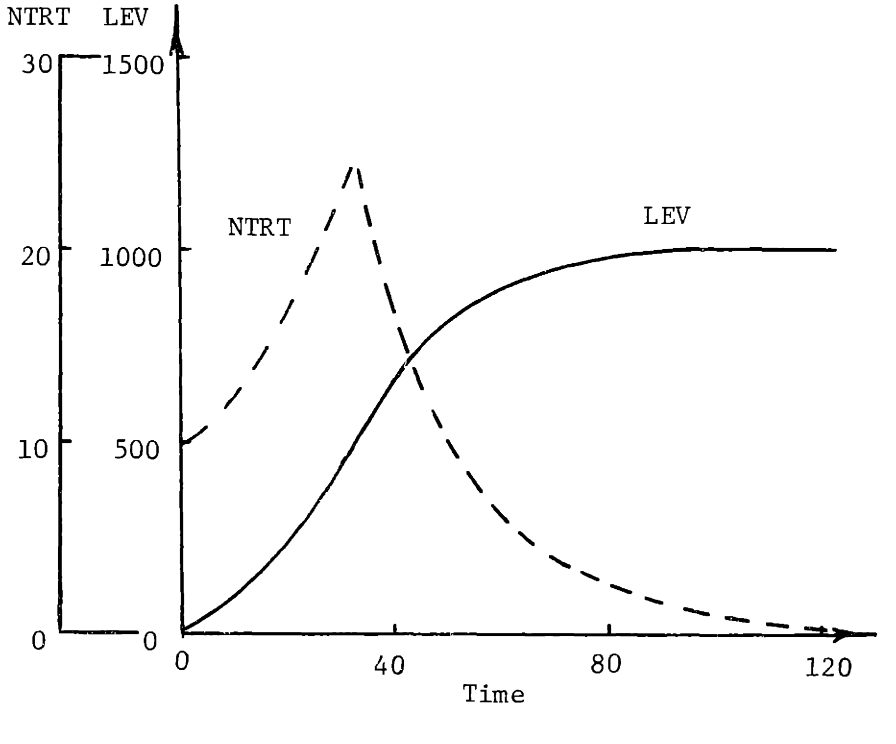


Figure 38
Simulation with Slope CONST1'

Summary

In general, a simple rate-level structure that exhibits rate increases with increasing level values, and then exhibits rate decreases with further level increases, will produce common S-shape growth. The desired or equilibrium value reached will exhibit the stable equilibrium characteristics of the first order negative feedback structure.

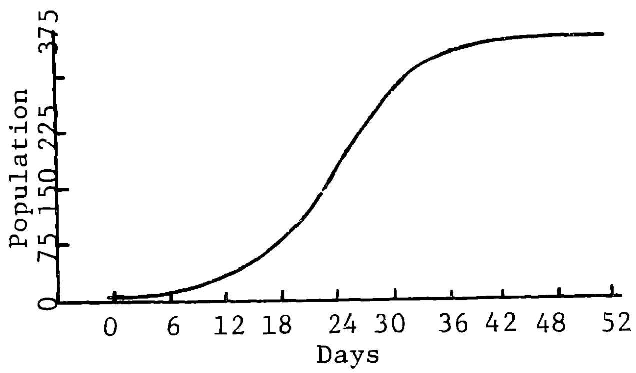
It is important to note that the causal diagram of Figure 27 does not by itself contain enough information to determine what the behavior of the system will be. A net rate based on a combination of the linear Loops 1 and 2 is incapable of producing sigmoid growth. The structure can only display exponential growth, decay, or linear growth depending upon whether Loop 1, Loop 2, or neither Loop 1 nor Loop 2 predominates. Sigmoidal growth requires that the dominance *shift* from Loop 1 (positive feedback) to Loop 2 (negative feedback) through a nonlinear relationship.

Application of the S-shape Growth Structure

Real world occurrences of sigmoid growth can be modeled utilizing the underlying feedback structure above. Knowledge of simple feedback relationships should facilitate the modeling of a diverse collection of often observed but infrequently explained S-shape phenomena. The cases investigated are: the growth pattern of a population of rats, the propagation of a contagious disease, and the behavior of a damped pendulum.

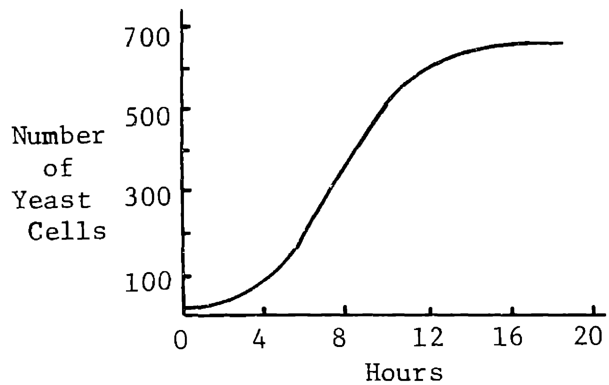
Example 1--Population Growth

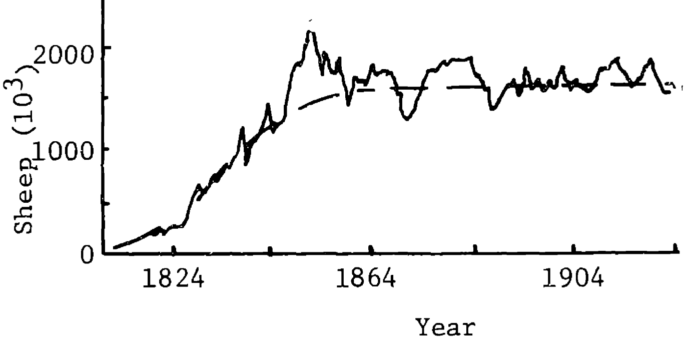
The growth and stabilization of the populations of various organisms are a phenomena observed by many population biologists. Experiments using fruit flies, bacteria, and sheep, for example, have resulted in the curves shown in Figure 39.



Growth of a Population of Drosophilia (Fruit Flies) under controlled experimental conditions, according to Pearl and Parker. Source: Lotka 1956

The growth curve of yeast cells in the laboratory. Source: Kormondy 1969





The growth curve of sheep subsequent to their introduction in Tasmania showing an intial sigmoidal pattern followed by semi-equilibrium. Source: Kormondy 1969.

Figure 39

Population Growth Examples

E. J. Kormondy (Kormondy 1969, p. 67) points out:

> Although they are difficult to come by, there are enough studies on a spectrum of different kinds of plants and animals to permit the statement that most species show a sigmoidal pattern during the initial stages of their population growth. There is, in such cases, an initial slow rate of growth, in absolute numbers, followed by an increase in rate to a maximum, at which point the curve begins to be deflected downward; it terminates in a rate that gradually lessens to zero, as the population more or less stabilizes itself with respect to its environment.

The stable equilibrium values of the population are achieved when the population is at the "carrying capacity" of the environment (i.e., the limit at which the environment can support the population). If the environmental conditions were shifted chemically or physically, a different equilibrium level would likely result (Kormondy 1969, p. 66).

Biotic factors, biological relationships within species (intraspecific) and between species (interspecific), together with abiotic factors, environmental characteristics, seem to be the regulatory forces involved in population growth and stability. Kormondy (Kormondy 1969, pp. 110-111) observes that:

> At a critical time in the life history of a given population, a physical factor such as light or a nutrient may be significant as a regulatory agent; at another time, parasitism, predation, or competition, or even some other physical factor may become the operative factor. As complex and as variable as the niche of any species is, it is unlikely that this regulation comes about by any single agency. However, there does appear to be considerable and mounting evidence, both empirical and theoretical, to suggest that populations are self-regulating through automatic feedback mechanisms. Various mechanisms and interactions appear to operate both in providing the information and in the manner of responding to it, and with the exceptional case of a catastrophe, the

stimulus to do so appears to depend directly on the density of the population. The end effect is one of avoiding destruction of a population's own environment and thereby avoiding its own extinction.

Crowding and Population Growth

In order to develop a simple model of population growth as an illustration of the general S-shape structure, attention will be given to only one of the single density dependent agencies Kormondy alludes to. The model will focus on a particular intraspecific interaction characteristic of mammal populations, mainly, crowding and infant mortality. The choice is arbitrary. Any single factor or combination of other factors affecting or limiting population growth could be used. The result, however, irrespective of the precise mechanism of the population check, will be the same as will be demonstrated.

Perhaps the most classic example of the effect of crowding on infant mortality is observed by B. F. Calhoun (Calhoun 1962, p. 139) in his experiments with Norway rats:

> Some years ago I attempted to submit [the question of population density on social behavior] to experimental inquiry. I confined a population of wild Norway rats in a quarter-acre enclosure. With an abundance of food and places to live and with predation and disease eliminated or minimized, only the animals' behavior with respect to one another remained as a factor that might affect the increase in their number. There could be no escape from the behavioral consequences of rising population density. By the end of 27 months the population had become stabilized at 150 adults. Yet adult mortality was so low that 5,000 adults might have been expected from the observed reproductive rate. The reason that this larger population did not materialize was that infant mortality was extremely high. Even with only 150 adults in the enclosure, stress from social interaction led to such disruption of maternal behavior that few young survived.

Calhoun devised an additional indoor experiment in order to examine in more detail the nature of the disruption. He found that mother rats failed to build nests or nurse their young adequately when exposed to overcrowding, causing infant mortality to rise (Calhoun 1962, p. 139).

Calhoun's description of the essentially sigmoid behavior of the rat population will form the qualitative basis of the model. From his description, the following assumptions are employed:

1. Confined space allowing no migration and no predation;
2. Ample and sufficient food supply;
3. Constant environment (no abnormal changes in weather, temperature, etc.).

For modeling simplicity, two further assumptions are made:

4. Disregard for the effects of age and of the onset of reproduction;
5. The sex ratio (males/females) of the population is 1.

The causal diagram for the rat population model is presented in Figure 40. The rat population is increased by births (Loop 1) and decreased by deaths (Loop 2). It was seen previously that in general Loops 1 and 2 alone can only produce the behavioral modes of Figure 33. Loop 3, an additional negative feedback loop based on Calhoun's observations, must then be responsible for transferring dominance from Loop 1 to Loop 2, necessary to produce equilibrium. As the population increases, the social stresses due to crowding are translated into a decrease in the aggregate birth rate of the population.

A flow diagram of Figure 41 is based on the causal loops described above. The rat birth rate RBR is defined as the number of infant rats per month that will survive infancy and attain adulthood. The normal rat fertility NRF is defined as the average number of infants per month produced by each adult female rat. It is equal to .4 (rats/female rat-month) for conditions of a relatively low or "normal" population density. Or equivalently, every female produces approximately 5 pups per year.

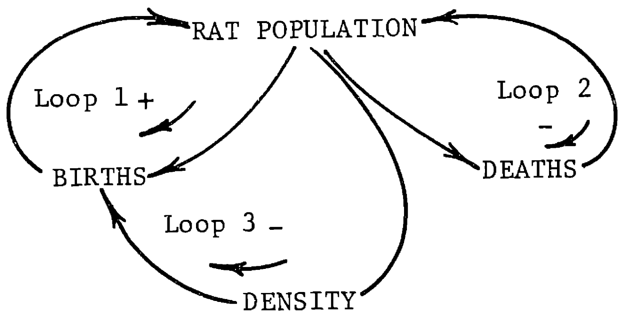


Figure 40
Causal Diagram--Rat Population

The adult rat death rate ARDR is computed as the number of adult rats divided by the average rat lifetime ARL. The average rat lifetime ARL is defined as the number of months an average rat survives during "normal" conditions and is taken as 22 months. This means that 4.5 percent of the population dies each month.

As Calhoun observed, the effect of high population density is an increase in infant mortality. Hence, the number of infant rats surviving to adulthood, the rat birth rate RBR, is reduced. This effect is captured in the model through alteration in the normal rat fertility NRF, by use of the infant survival multiplier ISM in the flow diagram.

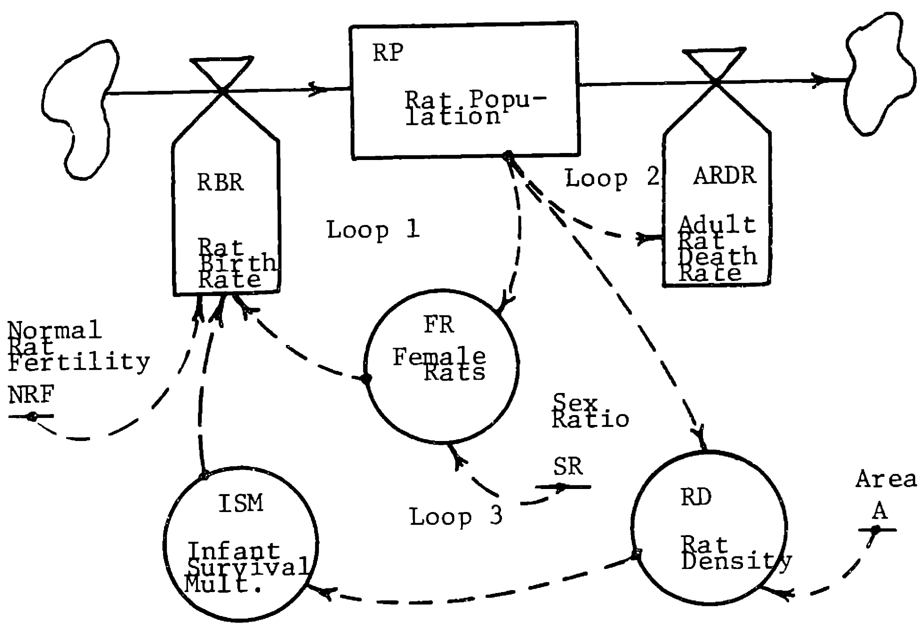


Figure 41

Flow Diagram--Rat Population

The infant survival multiplier table ISMT is produced in Figure 42. The curve is based on the assumption that at low population densities 100 percent of the pups born will survive. Under such circumstances, the normal rat fertility NRF is defined. As the population density PD increases, the percentage of rats surviving infanthood decreases. The normal rat fertility NRF is appropriately modified downward. In extremely crowded conditions, a very low percentage of the newly born pups will survive; the normal rat fertility NRF is 16 percent of its "normal" value. The equations for the system are given below.

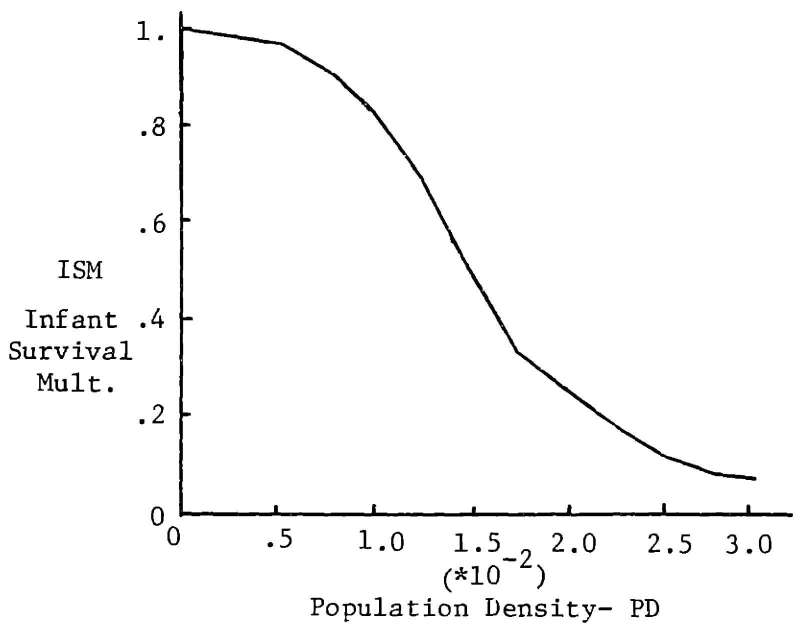


Figure 42

Infant Survival Multiplier Table

L	RP.K	=	RP.J + (DT)(RBR.JK-ARDR.JK)	(Rats)
N	RP	=	10	(Rats)
R	RBR.KL	=	NRF*FRP.K*ISM.K	(Rats/Month)
A	FRP.K	=	SR*RP.K	(Female Rats)
C	SR	=	.5	(Dimensionless)
C	NRF	=	.4	(Rats/Female-Month)

```
R  ARDR.KL  =  RP.K/ARL                               (Rats/Month)
C  ARL      =  22                                     (Months)
A  ISM.K    =  TABLE(ISMT,PD.K,0,.0225,.0025)         (Dimensionless)
A  PD.K     =  RP.K/A                                 (Rats/Sq. Ft.)
T  ISMT     =  1/1/.96/.92/.82/.7/.52/.34/.20/.16
C  A        =  11000                                  (Sq. Ft.)
```

By combining the rat birth rate RBR and adult rat death rate ARDR into a single net population growth rate NPGR and plotting the result as a function of the population, we obtain Figure 43.

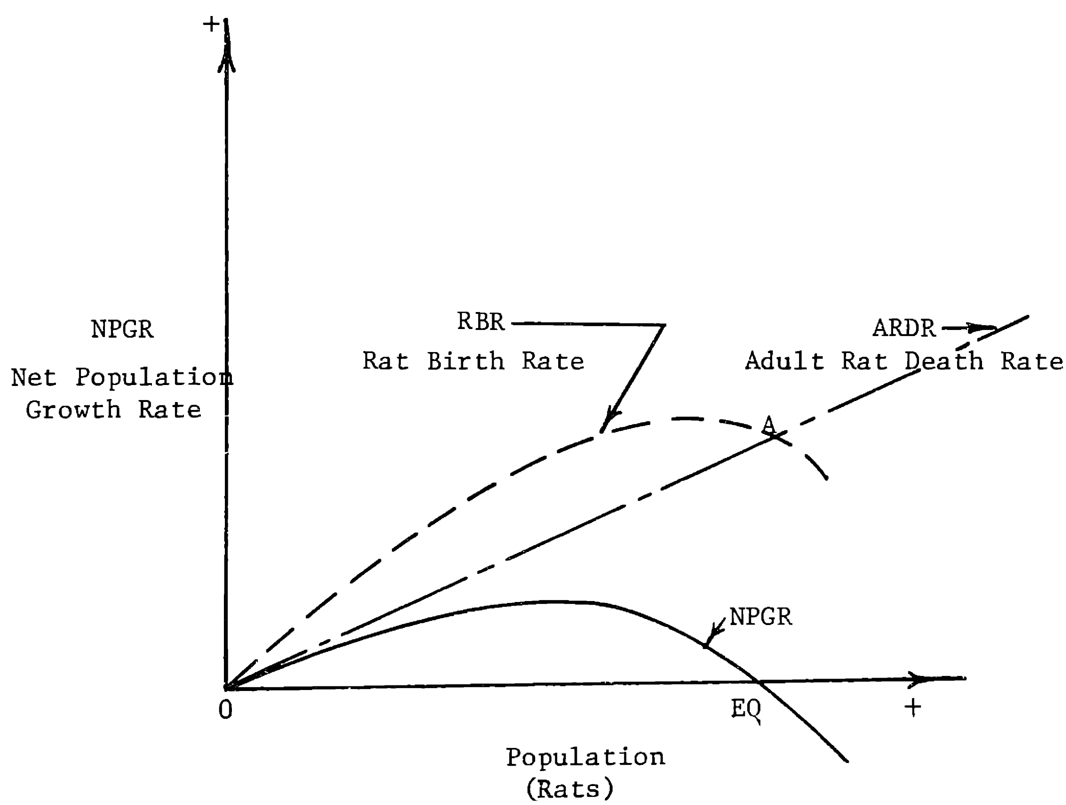


Figure 43

Net Population Growth Rate
with Linear Death Rate

The NPGR curve in Figure 43 is strikingly comparable to the curves in Figure 36. In the absence of "population pressure" due to crowding, the rat population grows exponentially. However, a reduction in the growth rate occurs as crowding sets in. Infant mortality begins to increase, reducing the rat birth rate RBR. Further population increase continues to reduce the net growth rate until growth stops. Births and deaths are in equilibrium (point A) and a steady state rat population is reached. It is expected that the net growth rate would have to cross the horizontal axis in order to achieve the stable population. Past point A the death rate is greater than the birth rate. The resultant sigmoid growth curve for the rat population is produced in the simulation run of Figure 44.

The model developed above accounts for only one density effect that alone could produce logistic growth. What about limiting factors acting on the adult rat death rate ARDR such as starvation, disease, or fighting? That is, assume that an ample food supply (Assumption 2), for example, did not exist. The effect of density might be included by an alteration of the linear death rate curve as seen in Figure 45. A decrease in the average rat lifetime ARL might be the mechanism involved. In Figure 45 a new nonlinear, density-dependent death rate curve intersects the birth rate curve at point B yielding a new and smaller "carrying capacity population" EQ'. The overall shape of the net growth rate curve, however, is not altered. Sigmoid growth would still result.

The conclusion to be drawn from this modeling exercise is graphically seen in Figure 45. A sigmoidal population growth trend results when either the birth rate or the death rate or both begin to decrease or increase respectively with density. This nonlinear effect of population growth on births and deaths produces a net growth rate relationship of the general type necessary for S-shape growth. Where the net population growth rate intersects the population axis, the "carrying capacity population" is determined.

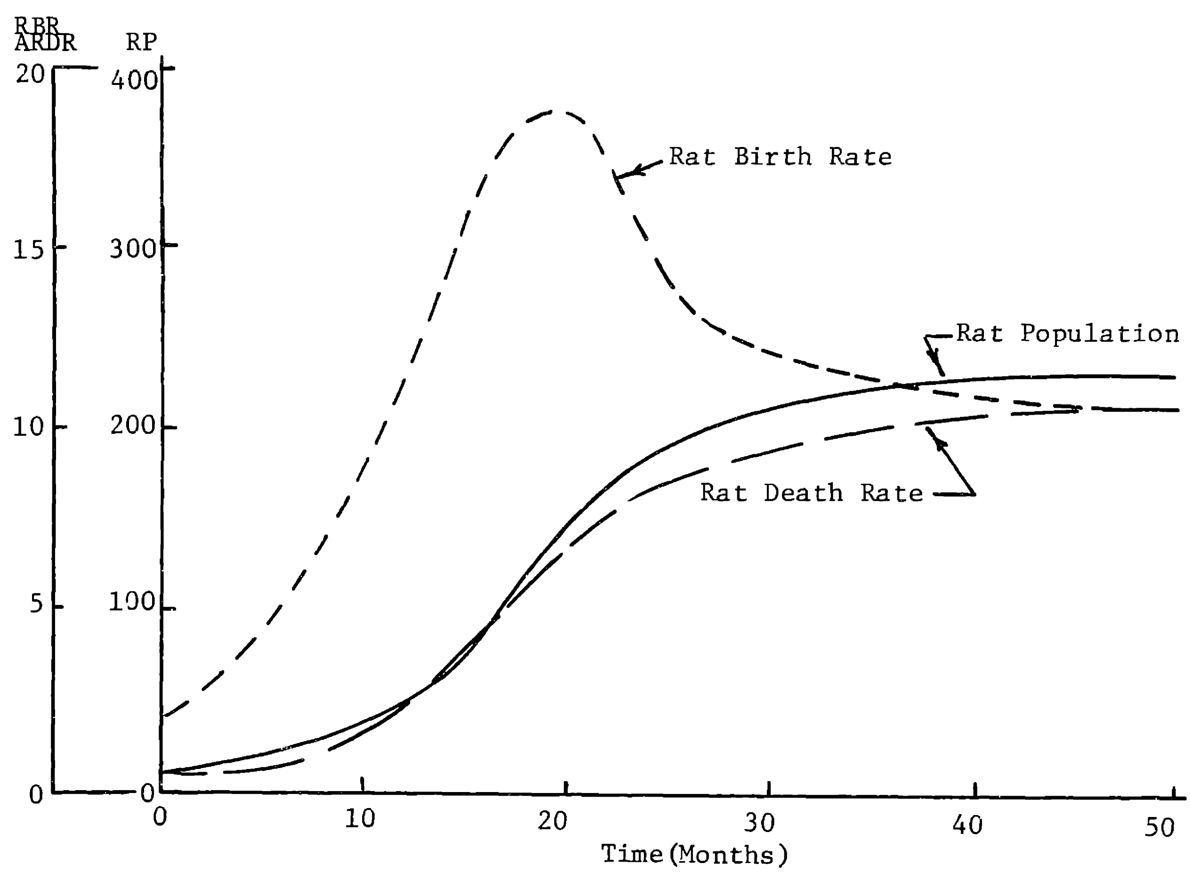


Figure 44
S-shape Rat Population Growth

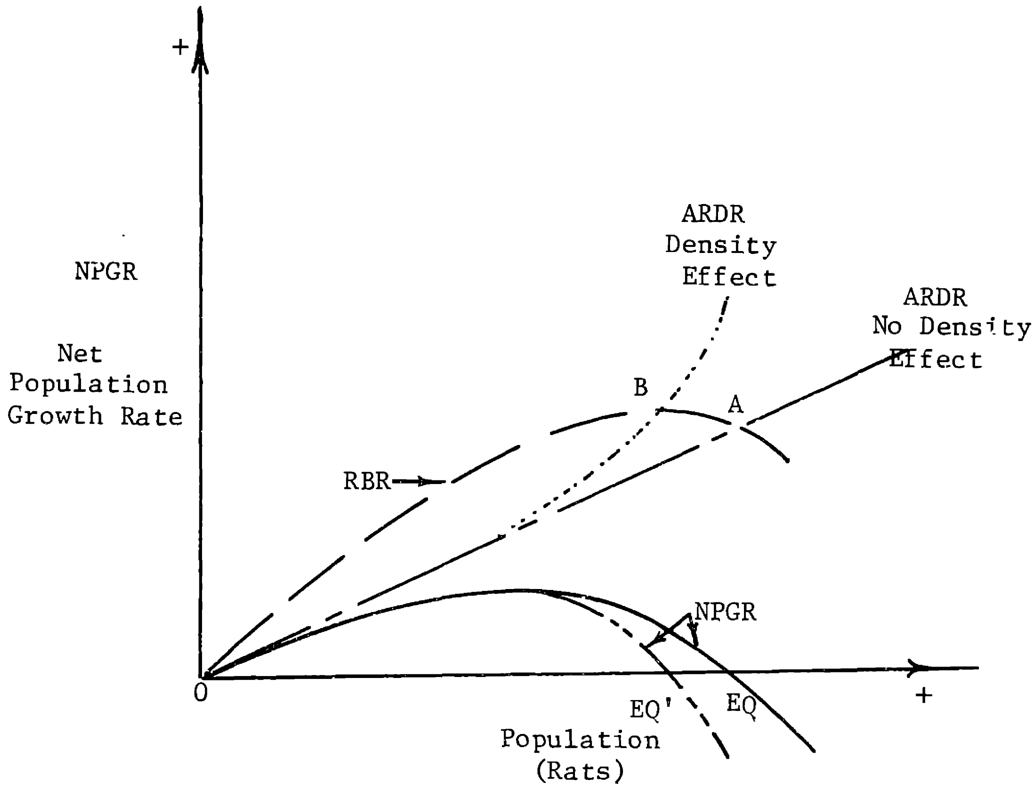


Figure 45

Net Population Growth Rate with Nonlinear Death Rate

Example 2--Epidemic Growth

The propagation of infectious diseases and the mouth-to-mouth diffusion of news have been observed under certain conditions to exhibit sigmoidal growth (Coleman 1961). A simple model of this epidemic behavior will be developed, employing the combined positive and negative feedback structure.

Simple Epidemic

Under the following circumstances, it is possible to develop a single level model which will replicate the growth of an epidemic:

1. Constant population allowing no migration;
2. Infectious people are never sufficiently ill to withdraw from circulation and are not cured of the infection during the course of the epidemic (thus reinfection is minimized);
3. Fairly homogenous mixing of the susceptible population and the infectious population.

Such circumstances seem to correspond to epidemics involving mild infections of the upper respiratory tracts such as the common cold, flu, and mild virus (Bailey 1955).

A causal diagram embodying the three assumptions is produced in Figure 46. The contagion rate depends on both the infected population and the susceptible population. In Loop 1, all else being equal, an increase in the infected population will result in an increase in the contagion rate. An increase in the contagion rate adds additional people to the infected population and so forth in a positive feedback manner. Since an infinite supply of population does not exist, however, all else is not the same. Loop 2, the negative feedback loop, accounts for the finite supply of susceptible population. As the infected population increases, the susceptible population, the difference between the total population and the infected population, decreases. This action leads to a suppression of the contagion rate. Eventually, the contagion rate must reach zero when the entire population has contracted the disease.

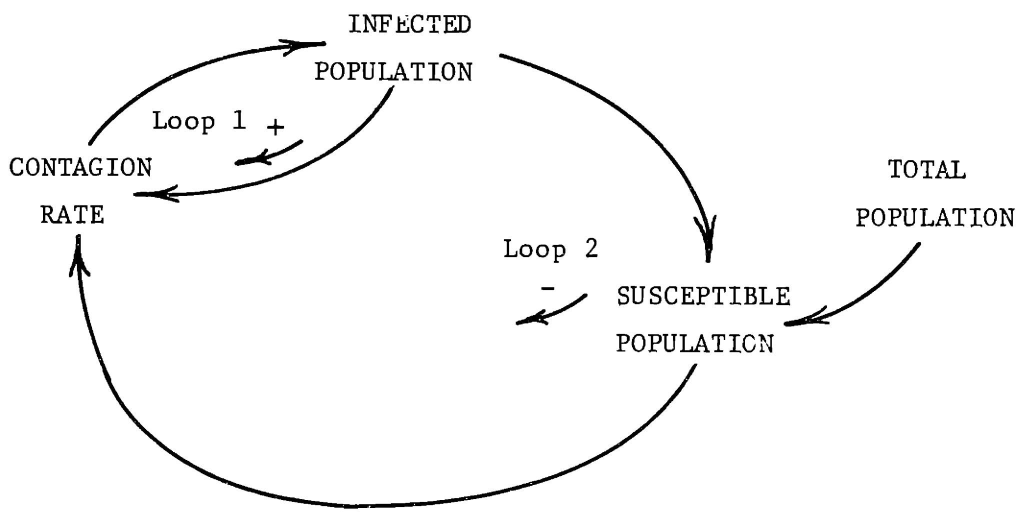
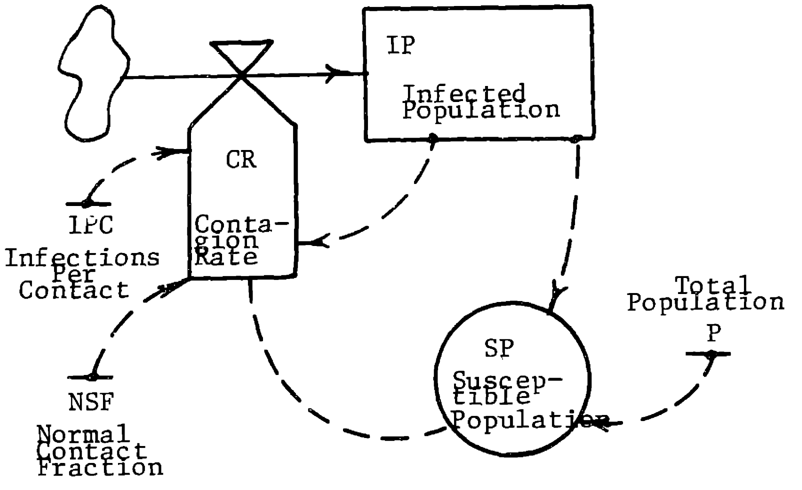


Figure 46
Causal Diagram of
Epidemic Model

The flow diagram and equations of Figure 47 represent the single level epidemic model. The contagion rate CR is developed from probabilistic considerations of the likelihood of infection between individuals making contact in a closed environment with uniform mixing (Bailey 1955). The term IP.K*SP.K is the total number of possible contacts that can occur. NCF is the percentage of these total contacts that actually do occur per day. The product of the three terms then yields the total number of contacts occurring per day. When this product is multiplied by the fraction of contacts producing an infection IPC, the number of infections (i.e., infected people) per day, or the contagion rate CR is generated.



L	IP.K	=	IP.J + (DT)(CR.JK)	(People)
R	CR.KL	=	IPC*NCF*IP.K*SP.K	(People/Day)
N	IP	=	10	(People)
C	IPC	=	.1	(Dimensionless)
C	NCF	=	.02	(Fraction/Day)
A	SP.K	=	P - IP.K	(People)
C	P	=	100	(People

Figure 47
Epidemic Model

The susceptible population SP is simply the total population P less the infected population IP. Eliminating the auxilary equation for the susceptible population SP by incorporating it into the rate equation yields:

CR.KL = IPC*NCF*IP.K(P-IP.K)

Plotting the contagion rate CR versus the infected population IP (Figure 48) yields the rate-level relationship necessary to produce sigmoidal growth. As the number of infected people increases, the infection rate increases. This causes further growth in the number of infected people. However, as the infected population IP increases, the uninfected population pool is rapidly depleted. The likelihood of contact between an infected person and an uninfected person is diminished, even though the infected population IP is large. Finally, when the entire population contracts the disease, the contagion rate CR must cease. Since the contagion rate CR is by definition a one-way flow, the curve cannot extend below the horizontal axis.

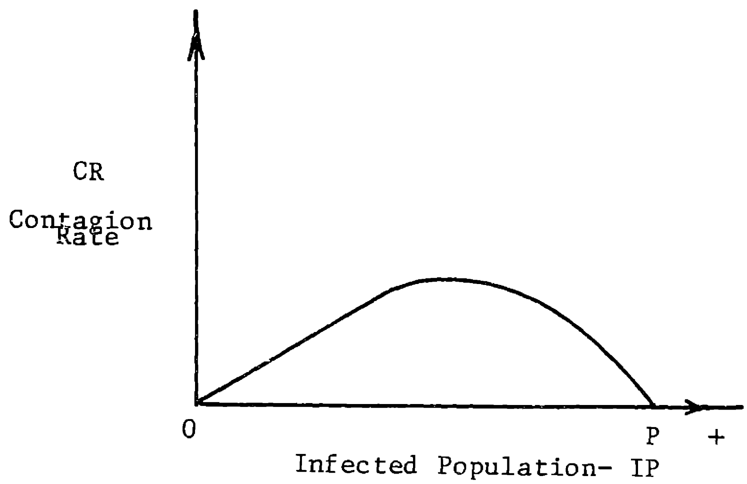


Figure 48

Contagion Rate vs Infected Population

Figure 49 is a simulation run of the model. The infected population IP time shape displays sigmoidal growth. The inverted sigmoid curve of the susceptible population SP also is noteworthy. Its shape is predictable since the sum of infected population IP and the susceptible population SP must all times equal the total population P.

While the model in this example was developed for a mild disease, with definitional changes, it would as well apply to a riot or a rumor. In general, epidemic processes that begin exponentially and eventually "consume" all available population must follow the sigmoid time form produced in Figure 49, at least in part.

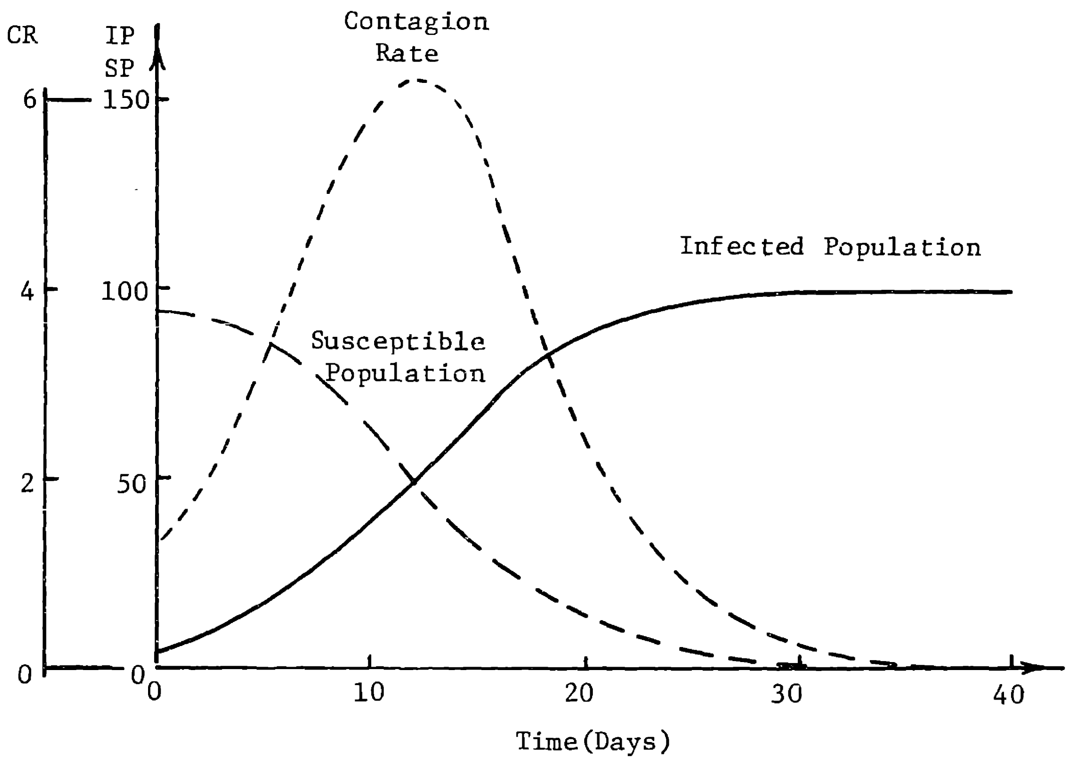


Figure 49

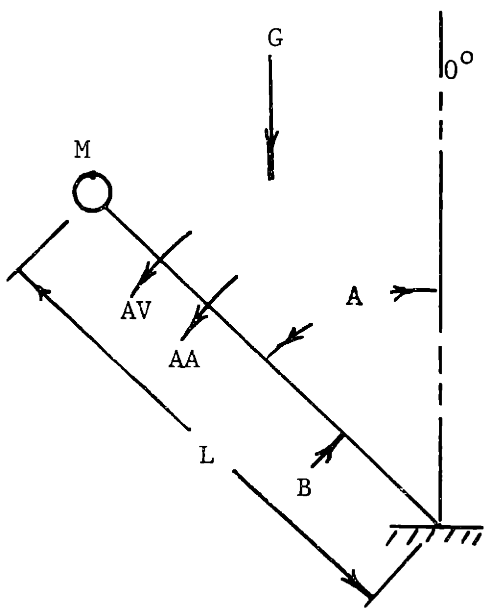
Epidemic Growth

Example 3--Damped Pendulum

This final example is taken from the physical sciences. Under certain conditions, it is possible to view the positional behavior of a damped pendulum as being a single level feedback system displaying sigmoidal growth.

The Damped Pendulum

Consider a pendulum in the vertical plane having a mass M, a damping coefficient B, and a length L as pictured in the free body diagram of Figure 50. When the pendulum is displaced from its zero degree (vertical) position, it will have an angular velocity AV and an angular acceleration AA.



A	=	Angle
AV	=	Angular Velocity
AA	=	Angular Acceleration
B	=	Damping Coefficient
L	=	Length
M	=	Mass
G	=	Gravitational Constant

Figure 50

Damped Pendulum

By considering the torques about the pivot point of the pendulum, we can form the following equation:

$$T_g + T_d = T_T \qquad (1)$$

where

T_g = Torque due to gravity

T_d = Torque due to damping

T_T = Total torque

The torques are in turn equal to the following:

$$T_g = F_g*L*\sin(A) \qquad (2)$$

$$T_d = F_d*L \qquad (3)$$

and

F_g = Force of gravity = $M \cdot G$

F_d = Force due to damping = $-B \cdot AV$

Substituting equations (2) and (3) into equation (1) yields equation (4):

$$M*G*L*\sin(A) - B*AV*L = T_T \qquad (4)$$

If it is assumed that the net torque T_T is very small as in the case of high damping, the following approximation can be made:

$$T_g + T_d \approx 0$$

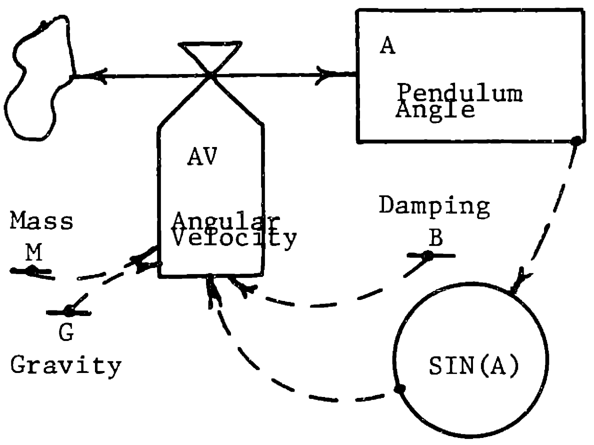
and

$$B*L*AV = M*G*L*\sin(A)$$

or

$$AV = (MG/B)*\sin(A) \quad (5)$$

Equation (5) is a rate equation where the angle A is the level. The model of the pendulum is given in Figure 51.



L	A.K	=	A.J + (DT)(AV.JK)	(Angle)
N	A	=	1	(Angle)
C	M	=	1	(Slugs)
R	AV.KL	=	(M*G/B)*SIN(A.K*6.28/360)	(Degrees/Second)
C	G	=	32	(Feet/Second Square)
C	B	=	2	(Pounds Force-Second/Degree)

Figure 51

Pendulum Model

A plot of the angular velocity AV for various values of the angle A according to equation (5) is seen in Figure 52. The rate curve is the general nonlinear form needed to produce the S-curve time shape seen in Figure 53. As the pendulum is displaced from its unstable equilibrium at A = 0°, the velocity increases until reaching A = 90°. At A = 90°, it has achieved its maximum velocity. The pendulum continues to lose speed until at A = 180° its stable equilibrium value of zero velocity is reached. It is not possible for the pendulum to overshoot 180° under the assumption of zero net torque employed in the model.

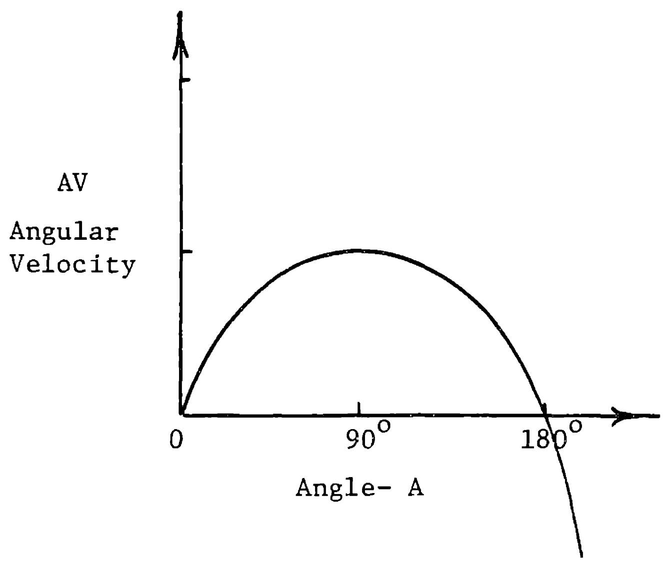


Figure 52
Angular Velocity vs Angle

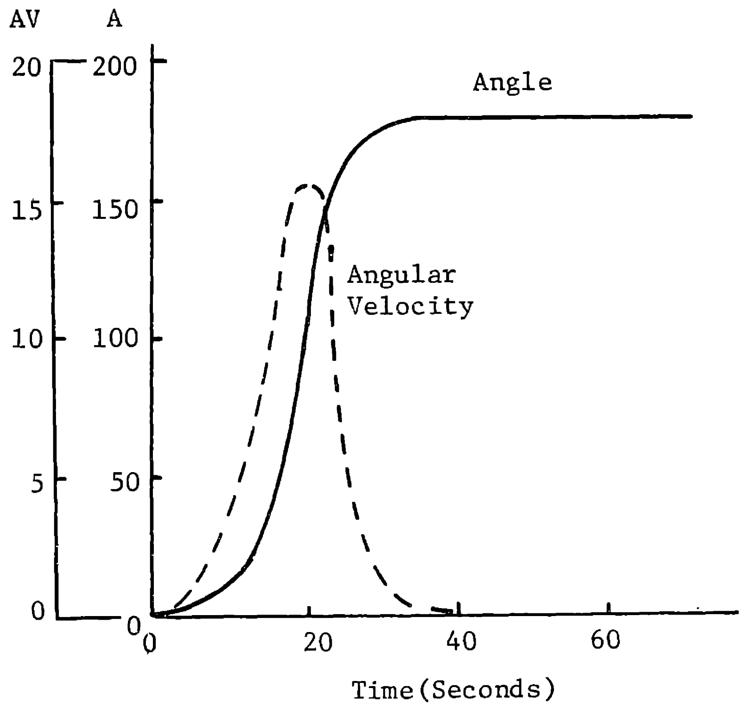


Figure 53
Damped Pendulum Behavior

The damped pendulum besides serving as an illustration of sigmoid growth behavior also demonstrates how the notion of feedback can be applied to purely physical phenomenon. It is intuitive that the pendulum velocity alters the angular position. This unilateral view of cause and effect, however, obscures the other less observable process at work. Also present is the cause and effect relationship between the angular position and the velocity. From Figure 52 it can be seen that the effect (angular position) becomes the cause and the cause (velocity) becomes the effect. Looking at the pendulum system in this circular fashion enhances our understanding about the nature of many physical phenomena.

Chapter 4
Conclusions

From the study of simple models comes the tools for constructing and understanding larger models. This thesis has explored two elementary structures.

A simple negative feedback structure as in Chapter 2 produces behavior that is goal-seeking. The structure attempts to maintain a goal when perturbed from outside the system. When the outside disturbance is in the form of constant input, the resulting equilibrium state of the system is different than the desired state. The first order linear feedback loop provides insight into the dynamics of a common inventory maintenance system and a coffee cup cooling to room temperature. The simple nonlinear negative feedback structure involving a rate which saturates (i.e., remains constant) with respect to the level shares many of the same behavioral characteristics as the purely linear structure. However, when the input to the system is beyond the tolerance of the structure as illustrated in the pollution absorption model, goal diverging behavior results.

Exponential growth followed by exponential decay (sigmoid growth) is the characteristic time shape of the structure in Chapter 3. Such behavior can only be generated by a particular type of rate-level relationship unachievable in a linear structure: a shift in dominance from positive feedback to negative feedback. In the negative feedback region of the sigmoidal growth structure, the behavior is characteristic of linear negative feedback, demonstrating how knowledge of the dynamics of one elementary structure can be exploited to explain another. Population growth, epidemic growth, and pendulum behavior phenomena seem to involve the structural relationships capable of producing sigmoidal growth.

REFERENCES

Bailey, N. T. J.

1955 "Some Problems in the Statistical Analysis of Epidemic Data." Royal Statistical Society 118, Series A, 35.

Calhoun, J. B.

1962 "Population Denisty and Social Pathology." Scientific American 206, 139.

Coleman, J. S.

1964 Introduction to Mathematical Sociology, Free Press, New York, New York.

Forrester, J. W.

1968 Principles of Systems, Wright-Allen Press, Cambridge, Massachusetts.

1971 World Dynamics, Wright-Allen Press, Cambridge, Massachusetts.

Kormondy, E. J.

1969 Concepts of Ecology, Prentice Hall, Englewood Cliffs, New Jersey.

Lotka, A. J.

1956 Elements of Mathematical Biology, Dover Publications, New York, New York.

Meadows, D. H., et. al.

1972 Limits to Growth, Universe Books, New York, New York.

Odum, H. T.

1971 Fundamentals of Ecology, W. D. Saunders, Philadelphia, Pennsylvania.

Taagepera, Rein

1968 "Growth Curves of Empires." General Systems 22, 171.