Loop Polarity, Loop Dominance, and the Concept of Dominant Polarity

George P. Richardson
The Nelson A. Rockefeller College of Public Affairs and Policy
University at Albany – State University of New York
Albany, NY 12222

Abstract
There is a conspicuous gap in the literature about feedback and circular causality between intuitive statements about shifts in loop dominance and precise statements about how to define and detect such important nonlinear phenomena. This paper provides a consistent, rigorous, and useful set of definitions of loop polarities, dominant polarity, and shift in loop dominance, and illustrates their application in a range of system dynamics models.

Consistent with the usual definitions, the polarity of a first-order feedback loop involving a level x and a single inflow $\dot{x}$ is defined to be the sign of $d\dot{x}/dx$. Loop polarity is shown to depend upon the sign of parameters not usually considered to be part of the loop itself. The definition of loop polarity is then extended to multi-loop first order systems. All positive loops with gain less than one, such as economic multipliers, are shown to be multi-loop systems with dominant negative polarity. The shifts in loop dominance that occur in nonlinear system arise naturally as changes in the sign of dominant polarity.

The concepts developed in the paper are then applied to simple higher-order nonlinear feedback systems. The final application to a bifurcating system suggests that all bifurcations in continuous systems can be understood as consequences of shifts in loop dominance at equilibrium points.

Loop Polarity, Loop Dominance, and the Concept of Dominant Polarity

George P. Richardson
The Rockefeller College of Public Affairs and Policy
University at Albany – State University of New York
Albany, NY 12222

Introduction

Underlying the formal, quantitative methods of system dynamics is the goal of understanding how the feedback structure of a system contributes to its dynamic behavior. Understandings are captured and communicated in terms of stocks and flows, the polarities of feedback loops interconnecting them, and shifts in the significance or dominance of various loops. However, there is a conspicuous gap in our literature between intuitive statements about shifts in loop dominance and precise statements about how we define and detect such important nonlinear phenomena.

This investigation is an attempt to bridge that gap. In the effort to construct formal definitions of shifts in loop dominance, it became clear that our common definitions of loop polarities were not sufficiently precise. There is an underlying unease in our own field and in the cybernetics literature that we do not really know what a positive loop is. Ashby, for example, was bothered by the convergent behavior of the discrete positive loop

$$x_{t+1} = (1/2)\, y_t,$$
$$y_{t+1} = (1/2)\, x_t.$$

He used its apparently contradictory goal-seeking behavior to support his claim of the "inadequacy" of feedback as a tool for understanding complex dynamic systems (Ashby 1956, p. 81). To avoid such anomalies, some define a loop to be positive if it gives "divergent behavior." Graham (1977) finds problems with that characterization and suggests instead that a loop be called positive if its open-loop steady state gain is greater than one. Richmond delightfully exposed our confusions by describing a well-meaning professor trying to explain to a concerned student: "Positive loops are … er, well, they give rise to exponential growth … or collapse … but only under certain conditions … under other conditions they behave like negative feedback loops…" He concluded that the nicest way out of the confusion is to define a positive loop to be a goal-seeking loop whose goal continually "runs off in the direction of the search" (Richmond 1980). Some, of course, ignore all the subtleties and obtain loop polarities simply by counting negative links (Richardson and Pugh 1981).

We begin then with a tighter, more formal definition of the polarity of a feedback loop. Our focus, however, is on the concept of loop dominance and the phenomenon of shifts in loop dominance in multi-loop nonlinear systems.

Rigorous Definition of Loop Polarity

We shall base our definition of loop polarity on the assumption that every dynamically significant feedback loop in a system contains at least one level (accumulation or integration).[1] The development will be in terms of continuous systems. A similar development holds for feedback processes couched in discrete terms, provided the principle of "an accumulation in every loop" is maintained.

Consider a single feedback loop involving a single level x and an inflow rate $\dot{x}$ = dx/dt.[2] Define the polarity of the feedback loop linking the inflow rate $\dot{x}$ and the level x to be

$$\text{sign}\left(\frac{d\dot{x}}{dx}\right) \qquad \left(= \text{sign}\left(\frac{dx/dt}{dx}\right)\right)$$

This formal definition is consistent with our more intuitive characterizations: "dx" can be thought of as "a small change in x" which is traced around the loop until it results in "a small change $d\dot{x}$" in the inflow rate $\dot{x}$ = dx/dt. If the change in the rate, $d\dot{x}$, is in the same direction as the change in the level, dx, then they have the same sign. Since $\dot{x}$ here is an inflow rate and thus is added to the level, the loop reinforces the initial change and is therefore a positive loop. In such a case, sign($d\dot{x}$/dx) is also positive, so the formal definition is consistent with the intuitive one. If the resulting change in the inflow rate is in the opposite direction to the change dx, then sign($d\dot{x}$/dx) is negative and the polarity of the loop is negative by both our intuitive and formal definitions. The formal definition is equivalent to defining the polarity of a first-order feedback loop to be the sign of the slope of its rate-versus-level curve.[3]

To extend the definition to feedback loops in which $\dot{x}$ is an outflow rate, we merely have to agree to attach a negative sign to the expression for $\dot{x}$ if it represents an outflow. Then the definition above holds for all loops involving a single level x and a single inflow, outflow, or net rate $\dot{x}$. The first few examples that follow are very familiar; they are intended to establish some confidence in this formal definition of loop polarity before we use it to derive some less familiar results.

Example (1): Exponential growth or decay.

Let $\dot{x}$ = bx, where b is a constant. Then the polarity of the feedback loop is

$$\text{sign}\left(\frac{d\dot{x}}{dx}\right) = \text{sign}\left(\frac{d(bx)}{dx}\right) = \text{sign}(b)$$

which is positive if b is positive and negative if b is negative.

The result makes intuitive sense, as may be seen by interpreting $\dot{x}$ as a net rate such as net population growth. If births exceed deaths, the coefficient b is positive and the loop produces exponential population growth. Similarly, if deaths exceed births, b is negative and the loop exhibits exponential decay behavior. The usual case is $b > 0$, and that prompts us to call all such first-order net-rate formulations positive loops. However, the polarity of such a loop in fact depends on a parameter whose sign is set by environmental conditions outside the loop. Without knowledge of the sign of b, the polarity of the loop represented by $\dot{x} = bx$ is undetermined.[4]

Example (2): Exponential adjustment to a goal.

Let $\dot{x} = \dfrac{(x^* - x)}{T}$, where x^* and T are constants.

Loop polarity = $\text{sign}\left(\dfrac{d\dot{x}}{dx}\right) = \text{sign}\left(\dfrac{(x^* - x)/T}{dx}\right) = \text{sign}\left(\dfrac{-1}{T}\right)$

which is negative if the time constant T is positive, and positive if T is negative.

In applications of this structure, as in exponential smoothing, the time constant T is always positive, so the loop is always negative. When $x^* = 0$, this formulation reduces to example (1) with $b = -1/T < 0$: again, a negative loop by both formal and intuitive definitions.

In each of these cases, the formal definition of loop polarity behaves appropriately but yields no new insights. Cases involving more than one loop provide more interesting testing ground.

Multi-Loop Structures: Loop Dominance

The formal definition of loop polarity leads to a precise concept of loop dominance in simple systems. Consider a first-order system containing several feedback loops and the level variable x.

Let $\dot{x}$ represent the net increase in x. Define the dominant polarity of the first-order system to be

$$\text{sign}\left(\frac{d\dot{x}}{dx}\right)$$

This simple extension of the formal definition of loop polarity to multi-loop first-order systems leads to new understandings of some familiar structures and a precise statement of what is meant by a shift in loop dominance. The examples below illustrate results for both linear and nonlinear systems.

Example (3): Logistic growth.

Let $\dot{x} = ax - bx^2$, $a >> b > 0$, $x_o > 0$.

This familiar structure can be thought of as a pair of feedback loops, one positive and one negative. One could rewrite the equation, for example, as

$$\dot{x} = (a - bx)\, x,$$

considering the factor $(a - bx)$ as a multiplier representing an endogenously changing fractional growth rate of x. If we take each factor as a separate first-order system, we have

$$\dot{x}_1 = x \quad \text{and} \quad \dot{x}_2 = a - bx.$$

The definition of loop polarity produces the expected results:

Polarity of loop 1 = $\text{sign}(d\dot{x}_1/dx) = \text{sign}(1)$ = positive

Polarity of loop 2 = $\text{sign}(d\dot{x}_2/dx) = \text{sign}(-b)$ = negative

Since $d\dot{x}/dx = a - 2bx$, the dominant polarity of this nonlinear system varies with the level x:

$$\text{Dominant polarity} = \text{sign}(a - 2bx) = \begin{cases} +, & \text{if } x < a/2b \\ -, & \text{if } x > a/2b \end{cases}$$

Thus the dominant polarity in this two-loop system shifts from positive to negative as the level variable x grows. The shift in dominant polarity suggests the following formal definition:

In a first-order system with level x and net rate of change $\dot{x}$, a shift in loop dominance is said to occur if and when $d\dot{x}/dx$ changes sign, that is, when the dominant polarity of the system changes.

In the logistic equation, a shift in loop dominance occurs when the level reaches half of its maximum value, the point of inflection in the logistic curve. The shift in loop dominance is a consequence of the nonlinearity of $\dot{x}$: in any first-order system containing any number of loops, if $\dot{x}$ is a linear function of x, $d\dot{x}/dx$ is constant and can not change sign. We conclude that first-order linear systems cannot show shifts in loop dominance.[5]

It should be noted that this definition does not capture all possible shifts in loop dominance -- only those that involve a change in dominant polarity. Presumably, it is entirely possible for a system to show a shift in dominance between two negative loops or two positive loops. Such a shift in dominance between loops of the same polarity would not show up as a change in dominant polarity and would have to be defined and detected by other means.[6]

Example (4): General nonlinear sigmoid growth structure.

Let $\dot{x} = x\, f(x)$, $f(x) > 0$, $x_o > 0$.

A suggestive example is the business construction formulation in several simple urban models (Alfeld & Graham 1976) in which

R BC. KL = BCN * BS. K * BLM. K,

where BC = business construction (structures/year),

BCN = business construction normal (fraction/year),

BS = business structures,

and BLM = business land multiplier (dimensionless),

which is a function of BS.

$$\text{Dominant polarity} = \operatorname{sign}\left(\frac{d}{dx}\left(x\, f(x)\right)\right)$$

$$= \operatorname{sign}\left(x\, f'(x) + f(x)\right) = \begin{cases} +, & \text{if } f'(x) > \dfrac{f(x)}{x} \\ -, & \text{if } f'(x) < \dfrac{f(x)}{x} \end{cases}$$

This result has a simple geometric interpretation. $f'(x)$ represents the slope of the tangent to the graph of $y = f(x)$ at the point $(x,f(x))$. On the same graph the term $f(x)/x$ represents the slope of the line from the origin to the point $(x,f(x))$.

Taken together, these considerations show:

> A nonlinear first-order feedback system of the form $\dot{x} = x\,f(x)$ shifts loop dominance at the point on the graph of $y = f(x)$ where the slope of the tangent is the negative of the slope of the line from the origin.

If such a point exists (that is, if loop dominance does indeed shift in the system), these two lines would form the diagonals of a rectangle with sides parallel to the x- and y-axes. Consequently, in a simple two-loop system the point of shifting loop dominance is relatively easy to pick out visually from a table function for $f(x)$. Figure 1 shows the determination of the point of shifting loop dominance for the business construction example cited above.

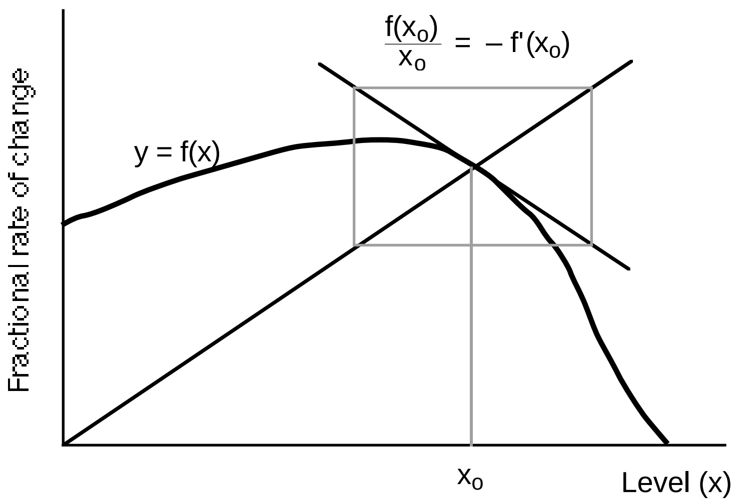

Figure 1: Locating on the graph of $y = f(x)$ the point x_o of shifting loop dominance in the first-order sigmoid growth system $\dot{x} = xf(x)$.

The criterion just derived applies neatly to the logistic equation as a special case. For $\dot{x} = ax - bx^2 = (a-bx)\,x$, the function $f(x)$ is $a - bx$, which is a straight line from $(0,a)$ to $(a/b,0)$. Therefore, the curve $y = f(x)$ itself becomes one of the diagonals of the rectangle that determine the point of shifting loop dominance, and the other diagonal is the line that runs from $(0,0)$ to $(a/b,a)$. Because the diagonals of a rectangle bisect each other, the point of shifting loop dominance is thus again found to be $x = a/2b$.

An analogous result, with an even simpler geometric interpretation, holds for nonlinear systems of the form $\dot{x} = (x^* - x)/f(x)$, $f(x) > 0$, so-called *nonlinear delays.* In such systems, x^* represents some goal state for the level variable x, and $f(x)$ represents a variable adjustment time dependent on the level. Examples of such formulations include pollution absorption in *World Dynamics* (Forrester 1971) and food regeneration in the KAIBAB model (Goodman 1974, Roberts et al. 1982). (In the former x^* would be zero since the absorption rate is simply the outflow from the pollution level.) In these cases, a computation[7] analogous to example (4) shows that loop dominance shifts when

$$f'(x) = \frac{f(x)}{x - x^*}$$

The geometric interpretation follows by noting that $f(x)/(x-x^*)$ can be viewed as the slope of the line joining $(x,f(x))$ and $(x^*,0)$. Loop dominance in such a system thus shifts when the slope of the tangent to the graph of $y = f(x)$ equals the slope of the line from the point of tangency to the point $(x^*,0)$.

As an example, Figure 2 shows the table function for pollution absorption time from Forrester (1971). The tangent line shown in the figure appeared in the original without explanation. Now we know its significance: since $x^* = 0$ here the line from $(0,0)$ tangent to the graph determines the location of the shift in loop dominance of this system. Because Forrester's table function formulation happens to lie *along* this line for $10 < \text{POLR} < 20$, the shift in loop dominance occurs not at a point but over an interval. For POLR < 10, the negative loop dominates and the system is capable of absorbing increases in pollution; for POLR > 20, the positive loop dominates and the system has the capability of exhibiting runaway pollution increases for constant or even declining rates of pollution generation. In the interval [10,20] neither loop dominates: when the pollution ratio falls in this range the system is essentially open-loop.

Figure 2: Table function for pollution absorption time from Forrester (1971), showing the line indicating the interval over which

loop dominance shifts from negative to positive as the pollution level grows.

Example (5): "Positive loops with gain less than one."

A classic example of this structure is the consumption multiplier (Samuelson 1939, Low 1980), shown in Figure 3. In the formulation of the loop used here, average income x is represented as an exponential smooth of GNP (Y), so

$$\dot{x} = \frac{Y - x}{T}$$

where T is a positive smoothing time constant.

Since Y = G + C = G + cx,

$$\dot{x} = \frac{(G+cx) - x}{T}$$

so

$$\frac{d\dot{x}}{dx} = \frac{c - 1}{T}$$

Therefore,

$$\text{Dominant polarity} = \text{sign}(d\dot{x}/dx) = \text{sign}\left(\frac{c - 1}{T}\right) = \begin{cases} +, & \text{if } c > 1 \\ -, & \text{if } c < 1 \end{cases}$$

Since the propensity to consume (c) must necessarily be a fraction between zero and one, we conclude that dominant polarity of the multiplier loop is always negative.

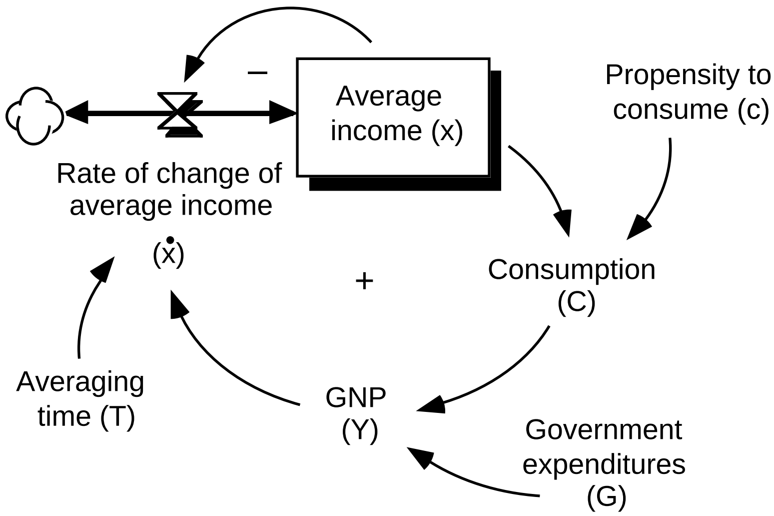

Figure 3: The consumption multiplier: for 0 < c < 1, a first-order system with negative dominant polarity.

The coefficient c in this system is commonly referred to as the open-loop steady-state gain or open-loop step gain of the positive loop connecting GNP (Y), consumption (C), and average income (x). The multiplier structure is thus usually characterized as a positive loop with gain less than one. From the point of view of loop dominance and dominant polarity, however, it is clearly seen to be a structure consisting of *two* loops, one positive and one negative, in which, for all sensible parameter values (0 < c < 1), the negative polarity always dominates. A similar but higher-order structure figures prominently in the market growth model in Forrester (1968a).

The goal-seeking behavior that such systems display is thus no surprise. It is intuitively reasonable that a system with dominant negative polarity should be goal-seeking. Furthermore, it is evident that one need not invoke an additional concept, such as "gain," to explain the apparent anomaly of "goal-seeking positive loops." The nonlinear notion of loop dominance, which is part of the system dynamicist's everyday stock-in-trade, suffices admirably in these special linear cases.

More Complex Systems

The goal of developing rigorous definitions of loop polarity, dominant polarity, and shift in loop dominance is to be able to say something significant about multi-loop nonlinear systems containing a number of different rates, levels, and auxiliaries. Taking auxiliaries first as the easiest to handle, let us make the obvious formal definition of the polarity of a link:

> Let variable A directly influence variable B. Define the polarity of the link from A to B to be $\text{sign}\left(\frac{\partial B}{\partial A}\right)$.

This definition is merely a formal statement patterned after our previous definitions, which expresses the intuitive notion of a change in A (∂A) resulting in a change in B (∂B) in the same or the opposite direction. (We've moved to partial derivatives because in higher-order systems a rate $\dot{x}$ can vary as a function of levels other than x).

Now suppose the rate $\dot{x}$ is linked to the level x through a sequence of auxiliaries, x ---> a_1 ---> a_2 ---> . . . ---> a_n ---> $\dot{x}$ ---> x. Repeated application of the chain rule for differentiation of composite functions yields

$$\frac{\partial \dot{x}}{\partial x} = \frac{\partial a_1}{\partial x} \frac{\partial a_2}{\partial a_1} \frac{\partial a_3}{\partial a_2} \cdots \frac{\partial a_n}{\partial a_{n-1}} \frac{\partial \dot{x}}{\partial a_n}$$

It follows that $\text{sign}(\partial\dot{x}/\partial x)$, the polarity of the feedback loop formed by this sequence of auxiliaries and $\dot{x}$ and x, is the product of the signs of the links in the loop, as we have in the past defined it.

To get a sense of the applicability of these ideas to higher-order systems, let us consider a familiar nonlinear system containing two system states:

Example (6): The Lotka-Volterra predator-prey equations.

$$\dot{x} = ax - bxy$$

$$\dot{y} = -cy + dxy,$$

where x represents the prey population and y the predators.

Applying the definition of dominant polarity to each of these equations independently, we find[8]

$$\text{sign}\left(\frac{\partial \dot{x}}{\partial x}\right) = \text{sign}\,(a - by) = \begin{cases} +, & \text{if } y < a/b \\ -, & \text{if } y > a/b \end{cases}$$

and

$$\text{sign}\left(\frac{\partial y}{\partial y}\right) = \text{sign}(-c + dx) = \begin{cases} +, & \text{if } x < c/d \\ -, & \text{if } x > c/d \end{cases}$$

In this situation, these expressions tell us the conditions under which each population's behavior is dominated by its own positive loop or negative loop processes, that is, births or deaths. For the prey, the positive loop dominates and the prey flourish as long as the predator population is small (< a/b). For the predators, the positive loop comes to dominate only when the prey population exceeds a certain critical size (> c/d).

Population	Dominant polarity					
x (prey)	+	+	-	-	+	...
y (pred)	-	+	+	-	-	...
Conditions	x < c/d y < a/b	x > c/d y < a/b	x > c/d y > a/b	x < c/d y > a/b	x < c/d y < a/b	... ...

Figure 4: Patterns of loop dominance about the individual levels in the Lotka-Volterra equations.

While these expressions for dominant polarity about each of the levels independently do not tell the whole story of the behavior of the system, they do strongly suggest that the system ought to oscillate. One could reason as follows. Say the system starts with both populations small: x < c/d, y < a/b. Then according to the above calculations, the x population is dominated by its positive (births) loop, and the y population is dominated by its negative (deaths) loop. Thus x should grow and y should decline. If x grows to exceed c/d, the system experiences a change in loop dominance: the y population comes to be dominated by its positive (births) loop, so y ought to cease declining and start to rise. Eventually, if y grows to exceed a/b, another shift in loop dominance takes place: the x population comes to be dominated by its negative (deaths) loop, so it ought to peak and begin declining. If x falls far enough (< c/d), loop dominance for the y population again shifts to the negative and the y population must start to decline. Eventually, y ought to fall far enough to shift the dominant loop of the x population, causing the prey to start to rise, and bringing us back to the start of this analysis to repeat the cycle. Figure 4 shows the recurring pattern of dominant polarities in the behavior of this predator-prey system.

It should be emphasized that these computations of loop dominance tell only part of the story about the behavior of the system. The major negative loop in the system is only implicitly being taken into account. It is the mechanism that brings about the shifts described above, and as a negative loop with more than one accumulation it is the actual source of the oscillatory tendencies of the system, but we did not explicitly make use of its structure or polarity. To see what we might be missing it is instructive to consider what an eigenvalue analysis of a linearized version of this system would look like:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} f(x,y) \\ g(x,y) \end{bmatrix} \approx \begin{bmatrix} f(x_o,y_o) \\ g(x_o,y_o) \end{bmatrix} + \begin{bmatrix} f_x(x_o,y_o) & f_y(x_o,y_o) \\ g_x(x_o,y_o) & g_y(x_o,y_o) \end{bmatrix} \begin{bmatrix} x - x_o \\ y - y_o \end{bmatrix}$$

In the Lotka-Volterra system, the essential matrix is

$$\begin{bmatrix} f_x(x_o,y_o) & f_y(x_o,y_o) \\ g_x(x_o,y_o) & g_y(x_o,y_o) \end{bmatrix} = \begin{bmatrix} a - by_o & - bx_o \\ dy_o & dx_o - c \end{bmatrix}$$

Thus in such a linearization our partial derivatives $\partial\dot{x}/\partial x$ and $\partial\dot{y}/\partial y$ would appear (as f_x and g_y), but so would $\partial\dot{x}/\partial y$ and $\partial\dot{y}/\partial x$ (as f_y and g_x, respectively). By investigating $\partial\dot{x}/\partial x$ and $\partial\dot{y}/\partial y$ alone we are ignoring terms off the main diagonal in the linearized state-space matrix. It looks as if the potential for this development of the notion of dominant polarity is limited to systems in which the off-diagonal terms are few and far between, or are for some other reason not particularly significant.[9]

In spite of that apparent limitation, we can apply these ideas to higher-order feedback systems and learn something. To do so we first need a rigorous definition of the polarity of a major loop. Consider a loop composed of levels $x_1, x_2, \ldots, x_n$ connected in order: $\dot{x}_1$ ---> x_1 ---> $\dot{x}_2$ ---> x_2 ---> … ---> x_n ---> $\dot{x}_1$. Following the pattern of our previous definitions, define the polarity of the major loop to be

$$\text{sign}\left(\frac{\partial\dot{x}_1}{\partial x_n}\ \frac{\partial\dot{x}_2}{\partial x_1}\ \frac{\partial\dot{x}_3}{\partial x_2}\ \cdots\ \frac{\partial\dot{x}_n}{\partial x_{n-1}}\right)$$

Applied to the major loop in the Lotka-Volterra system, for example, this definition states that its polarity is

$$\text{sign}\left(\frac{\partial \dot{x}}{\partial y}\ \frac{\partial \dot{y}}{\partial x}\right) = \text{sign}\ ((-bx)(dy))$$

which is indeed negative, as it should be, since b, d, x, and y are all greater than zero.

This definition of the polarity of a major loop is consistent with, and in fact depends upon, the principle of feedback systems that asserts that rates and levels alternate around a loop (Forrester 1968b). It is also consistent with our intuitive characterizations, for it amounts to tracing around the major loop the implications of a small change in one of the levels. Note that, when integrated with the definition of the polarity of a string of auxiliaries, this definition asserts that the polarity of a feedback loop containing any number of rates, levels, and auxiliaries is the product of the signs of the links in the loop.

In the development that follows, we shall need one other fact, a theorem about the steady-state behavior of the smooth of an exponentially growing or declining variable. Say $y(t) = y_0 e^{gt}$, and let $z(t)$ be an exponential smooth of $y(t)$. That is, $\dot{z} = (y - z)/T$, for some time constant T. Then in the steady-state,[10]

$$z(t) = \frac{1}{1+Tg}\, y(t)$$

That is to say, the smooth of y(t) is also growing (or declining) at the same exponential rate g, but it "lags behind" by a factor of 1/(1+Tg). The result holds for first-order exponential material delays as well as for information smoothing. This property can be used to simplify the computation of dominant polarity in some higher-order systems, as the following extended example demonstrates.

Example (7): Corporate growth from product development.

The structure shown in Figure 5 is the essence of the self-regenerating process behind the growth of a product-driven company. The major loop (the revenue loop) is positive, and the minor loops are all negative.

To find the dominant polarity of the system, we compute:

$$\frac{\partial \dot{PD}}{\partial PD} = \frac{\partial}{\partial PD}(IR - CR) = \frac{\partial}{\partial PD}(CR + CPD - CR) = \frac{\partial}{\partial PD}(CPD)$$

$$= \frac{\partial}{\partial PD}\left(\frac{PS - PD}{PAT}\right) = \frac{1}{PAT}\left(\frac{\partial PS}{\partial PD} - 1\right)$$

Tracing the chain of variables around with the chain rule,

$$\frac{\partial PS}{\partial PD} = \frac{\partial PS}{\partial RDB}\frac{\partial RDB}{\partial AR}\frac{\partial AR}{\partial REV}\frac{\partial REV}{\partial PP}\frac{\partial PP}{\partial PD}$$

which will be much more convenient for us in the form

$$\frac{\partial PS}{\partial PD} = \frac{\partial PS}{\partial RDB}\frac{\partial RDB}{\partial AR}\frac{\partial AR}{\partial REV}\frac{\partial REV}{\partial PP}\frac{\partial PP}{\partial OR}\frac{\partial OR}{\partial CR}\frac{\partial CR}{\partial PD}$$

(Note that the last three terms equal $\partial PP/\partial PD$.)

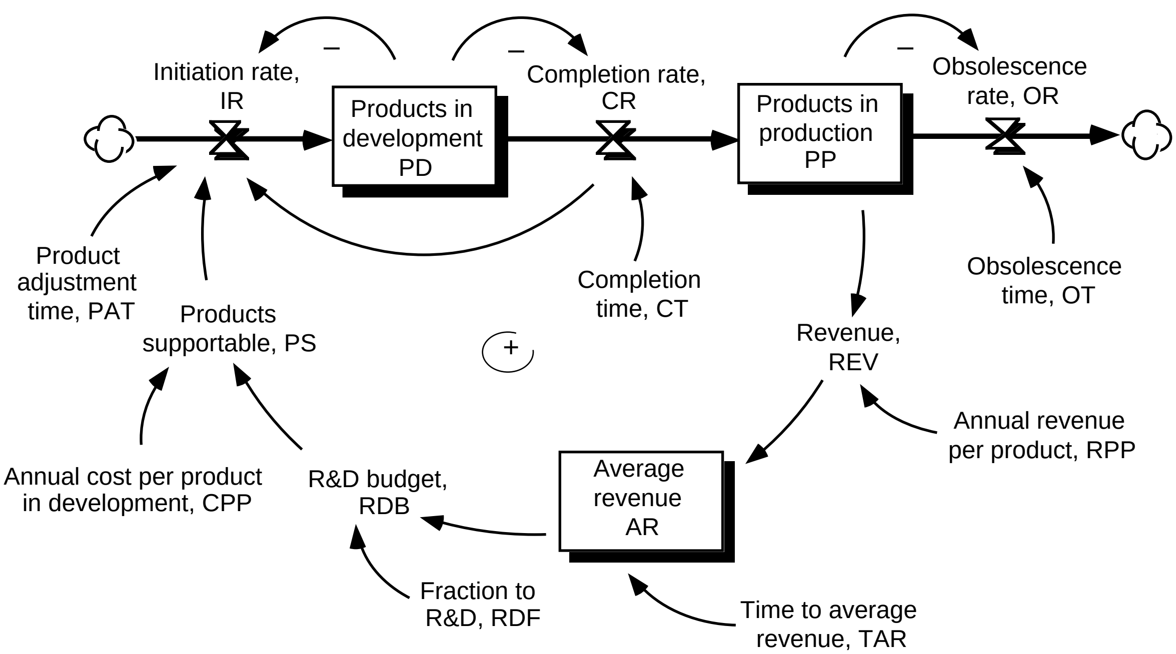

Figure 5: Corporate product-development structure.

At this point we see that we can compute each of these terms from their defining equations except $\partial AR/\partial REV$ and $\partial OR/\partial CR$. These are fractional rates of change of an exponential smooth and a delay. As noted above, their magnitudes depend upon the exponential rate of growth of the system, which is in fact what we are looking for in $\partial \dot{PD}$

$/\partial PD$. As feedback thinkers we are accustomed to circularities, but this is too much; we have to make an additional assumption to proceed.

Let us ask what this expression for dominant polarity becomes if and when the system has reached some *steady state* of exponential growth or decline. Let the system be growing or declining exponentially at the fractional rate g. Then from the equations in the system and the property of exponential smooths and delays cited above,

$$\frac{\partial PS}{\partial PD} = \frac{1}{CPP}\ RDF\ \frac{1}{1 + TARg}\ RPP\ OT\ \frac{1}{1 + OTg}\ \frac{1}{CT}$$

Substituting into the equation for $\partial \dot{PD}/\partial PD$ and rearranging slightly, we find

$$\frac{\partial \dot{PD}}{\partial PD} = \frac{1}{PAT}\left(\frac{RDF\ RPP\ OT}{CPP\ CT}\ \frac{1}{1 + TARg}\ \frac{1}{1 + OTg} - 1\right)$$

Under the simplifying assumption of steady-state exponential growth, this expression equals the fractional growth rate g. Setting it equal to g, multiplying through by various denominators, and rearranging, we obtain the following pleasing polynomial in g:

$$\frac{RDF\ RPP\ OT}{CPP\ CT} = (1 + OTg)\ (1 + TARg)\ (1 + PATg)$$

Because of the steady state assumption, the dominant polarity we seek is $\text{sign}(\partial \dot{PD}/\partial PD) = \text{sign}(g)$, where g is a solution of this polynomial. If g is small relative to the time constants OT, TAR, and PAT, as presumably it always would be, then this equation may be written

$$\frac{RDF\ RPP\ OT}{CPP\ CT} = 1 + (OT + TAR + PAT)\ g + O(g^2) \approx 1 + (OT + TAR + PAT)\ g$$

where the higher order terms in g are dropped because they are insignificant.[11] Thus

$$\frac{\partial \dot{PD}}{\partial PD} = g \approx \left(\frac{RDF\ RPP\ OT}{CPP\ CT} - 1\right)\left(\frac{1}{OT + TAR + PAT}\right)$$

so dominant polarity $= \text{sign}\left(\frac{\partial \dot{PD}}{\partial PD}\right) = \text{sign}(g)$

$$= \text{sign}\left(\frac{RDF\ RPP\ OT}{CPP\ CT} - 1\right)$$

$$= \begin{cases} + \text{ if } \dfrac{RDF\ RPP\ OT}{CPP\ CT} > 1 \\ - \text{ if } \dfrac{RDF\ RPP\ OT}{CPP\ CT} < 1 \end{cases}$$

Thus the dominant polarity of this product development system is determined by the quantity

$$\frac{RDF\ RPP\ OT}{CPP\ CT}$$

If it is greater than one the dominant polarity of the structure is positive and corporate growth ensues; if it is less than one the dominant polarity is negative and the company declines. If as the system evolves over time the product completion time CT were to rise sufficiently to pull this ratio below one, then the dominance in this system would shift from the positive growth loop to the negative loops and constrain corporate growth.

The expression that determines the dominant polarity of this system involves some time constants as well as proportionality factors. The significance of each of these parameters for a healthy company is clear. The greater the ratio of revenue-to-development cost per product (RPP/CPP), and the greater fraction of revenues the company sends to R&D (RDF), the greater the development effort the company can afford, leading to the prospect of a high continuing flow of new products into production. The expression shows that long product-lifetimes in the marketplace (OT) also contribute to the growth potential of the company, while long product development completion times (CT) threaten that potential. It is worthwhile observing that the importance of these two time constants in the growth potential of the company is derived here without reference to corporate reputation or feedback effects of delivery delays. In addition to these well-known reputation effects, completion times and obsolescence times figure directly in the potential of the positive, revenue-generating loop to dominate in this structure.

It is interesting to observe that in this example the time constants TAR and PAT do *not* influence the dominant polarity of the system. Products in development PD and Average Revenue AR are SMOOTHs (first order exponential averages) of Products

Supportable PS and Revenue REV, respectively. The time constants for these SMOOTHs, PAT and TAR, appear in the expression for $\partial \dot{PD}/\partial PD$ in a way that affects not *whether* the system grows or declines, but rather *how rapidly* it moves in the direction that other parameters in the system dictate. The result is generalizable: exponential averaging of a variable affects the rate of growth or goal-seeking adjustment in a system but does not have any role in determining loop dominance.[12]

Stepping back from the details of this example, we have found that the dominant polarity of the product development loop is positive if a particular combination of parameters affecting the loop is greater than 1, and negative if the combination of parameters is less than 1. The particular mix of parameters,

$$\frac{RDF \ RPP \ OT}{CPP \ CT},$$

is the *open-loop steady state gain*, or the *open-loop step gain*, of the system. We have thus concluded in this example that:

> The dominant polarity of the product development system is positive if its open-loop step gain is greater than one, and is negative if its open-loop step gain is less than one.

This is Graham's suggestion for the definition of a positive loop (Graham 1977). It should be noted that it is a statement not about a single loop but about loop dominance in a *multi-loop system* comprising a major positive loop and a number of negative loops.

Bifurcations and Loop Dominance

A bifurcation is a sudden shift in the goal state of a continuous, nonlinear system.[13] It is natural to ask how shifts in loop dominance relate to bifurcations. The following analysis of a well-known simple example suggests that bifurcations occur at equilibrium points that are also points of shifting loop dominance.

Example (9): Bifurcation and shifts in dominant polarity in a first-order system.

Let $\dot{x} = xf(x) - bx$.

For specifics, interpret x as a population (such as right whales or passenger pigeons) and f(x) as its net birth rate factor. Let b represent the fraction of the population harvested per year.

The dominant polarity of the system is

$$\text{sign}\left(\frac{d\dot{x}}{dx}\right) = \text{sign}\left(f(x) + x\, f'(x) - b\right)$$

$$= \begin{cases} +, & \text{if } f(x) + x\, f'(x) > b \\ -, & \text{if } f(x) + x\, f'(x) < b \end{cases}$$

This system bifurcates if the net birth factor $f(x)$ rises to a peak before it declines to zero. Figure 6A shows such a graph for $y = f(x)$; Figure 6B shows the corresponding rate-versus-level graph for the net birth rate $xf(x)$.

If the harvesting rate equals the value labeled b_1 in Figure 6B, the system seeks and maintains a stable equilibrium population x_1: if x rises above x_1, the net rate $xf(x) - bx$ is negative and x falls back to x_1; and if x falls below x_1, the net rate $xf(x) - bx$ is positive and x rises back to x_1. Note that the slope of the net birth rate curve $y = xf(x)$ is negative at $x = x_1$. Since that slope equals $f(x) + xf'(x)$, that guarantees that the dominant polarity of the system in a neighborhood of $x = x_1$ is always negative. The system should be nicely stable and goal-seeking around x_1.

However, if the harvesting rate equals the value labeled b_2 in Figure 6B, the equilibrium population x_2 is stable only if approached from above. If the harvesting rate b were to rise at all above b_2, or if the population x were to fall a bit below x_2, the goal state of the system switches suddenly to zero. Thus x_2 and b_2 determine a bifurcation point of the system.

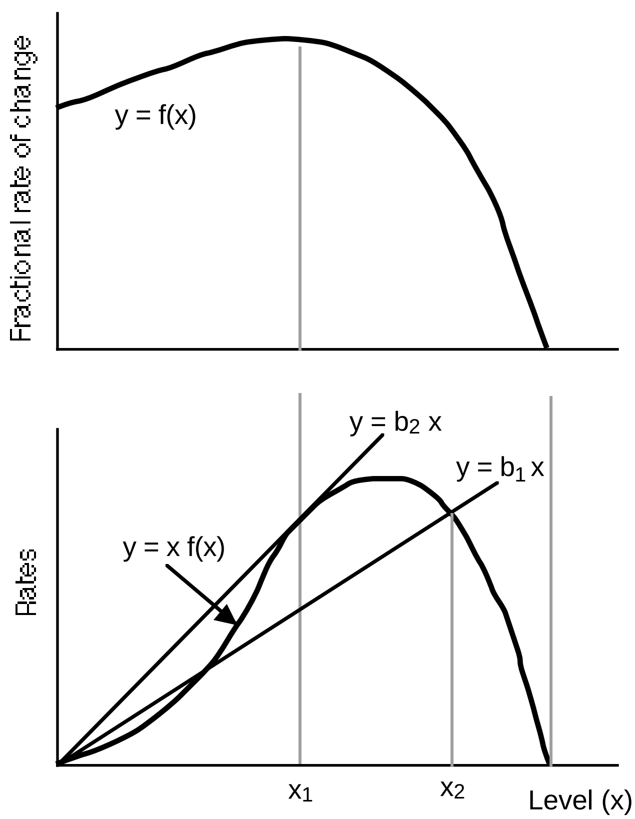

Figure 6: Graphs associated with the bifurcating system $\dot{x} = xf(x) - bx$.

A: Graph of $y = f(x)$

B: Graphs of the rates $xf(x)$ and bx versus the level x.

But x_2 and b_2 also determine a point of shifting loop dominance in the system. At that critical point, the slope of the net birth rate curve momentarily equals the slope of the harvesting curve. More precisely, in a neighborhood of x_2,

$$f(x) + x\,f'(x) \begin{cases} < b_2 & \text{if } x > x_2 \\ = b_2 & \text{if } x = x_2 \\ > b_2 & \text{if } x < x_2 \end{cases}$$

We see immediately that if $b = b_2$ the dominant polarity of the system shifts from negative to positive as x drops through the value x_2. The positive polarity for $x < x_2$ means a self-reinforcing decline in x: since $f(x_2)$ is the maximum value of $f(x)$, $\dot{x} = xf(x) - bx < 0$ for $x < x_2$. Thus the system shows a sudden shift in goal state because it experiences a shift in dominant polarity from goal-seeking negative to goal-divergent positive.

It is common to assert that as b increases through the bifurcation point in this system, the goal state of the system suddenly shifts to zero. However, that is not quite what happens. For b slightly bigger than b_2, or for $b = b_2$ and x slightly less than x_2, the system has positive dominant polarity, and its net rate $\dot{x}$ is negative. The system's goal is, for a time at least, *negative infinity*. As x drops more and more precipitously, the system experiences another shift in dominant polarity, from positive back to to negative. But here the goal of the negative polarity system is no longer $x = x_1$ or x_2, but rather $x = 0$. It is at this second shift in dominant polarity as x declines that it becomes appropriate to say that the goal of the system shifts to zero.

To see that there are two points of shifting dominant polarity in the system given by $\dot{x} = xf(x) - bx$, consider Figure 7. Dominant polarity shifts when $f(x) + xf'(x) - b$ changes sign, which in a continuous system implies $f(x) + xf'(x) = b$. Geometrically, that means that in this system dominant polarity changes when the slope of the tangent to the net birth rate curve equals the slope of the line representing the harvesting rate. A visual check of the slopes in Figure 7 shows that

$-\infty < x < x_1$ ===> slope of $x f(x) < b$ ===> polarity negative;

$x_1 < x < x_3$ ===> slope of $x f(x) > b$ ===> polarity positive;

$x_3 < x < \infty$ ===> slope of $x f(x) < b$ ===> polarity negative.

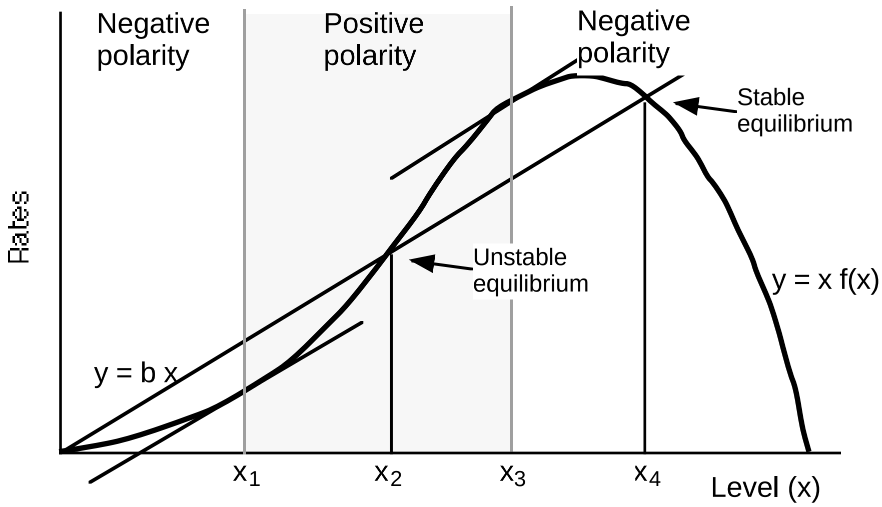

Figure 7: Illustration of the two points of shifting loop dominance in the bifurcating system $\dot{x} = x f(x) - bx$.
x_1 & x_3: points of changing dominant polarity;
x_2 & x_4: equilibrium points.

The points indicated by x_2 and x_4 in Figure 7 are, along with $x = 0$, the possible equilibrium points of the system. At $x = x_2$ the equilibrium is unstable, since it is in the interval of positive dominant polarity. Any deviation in either direction from $x = x_2$ is reinforced, moving x away from x_2 at an increasing rate. At $x = 0$ and $x = x_4$ the equilibria are stable, since both occur in intervals of dominant negative polarity in which deviations from equilibrium are counteracted.

Thus in Figure 6, when the system experiences a bifurcation at $b = b_2$ and $x = x_2$, the goal state of the system actually shifts to negative infinity. Then as x drops more and more rapidly the system eventually reaches the point where the slope of the tangent to $y = x f(x)$ again equals the slope of the line $y = b_2 x$. At that point the dominant polarity shifts back to negative, and, since $\dot{x}$ equals a negative number times x in this range, the goal for x becomes zero. Without its second shift in dominant polarity, this bifurcating system would not end up at a finite goal.

Mindful of the dangers of generalizing from one example, I conjecture that all instances of bifurcation in continuous systems result from shifts in loop dominance. More precisely, it seems reasonable that all such bifurcations occur at equilibrium points which are also points at which the dominant polarity can shift from negative to positive. The apparent sudden shift in goal state is a consequence of a shift in dominant polarity.

Summary and Conclusions

Shifts in loop dominance lie at the heart of significant feedback system dynamics. The concepts and definitions made more rigorous in this paper move in the direction of clarifying what we mean by such shifts. The necessary first step in that direction is to develop rigorous and reliable definitions of link and loop polarities and shifts in loop polarities. This paper has suggested such a set of definitions. They have the desirable property that they are formal definitions with clear and immediate connections to the intuitive characterizations in common use. Furthermore, they result in several simple algebraic and geometric tests for determining dominant polarities and shifts in dominance in simple systems.

The concept of dominant polarity developed here suggests the possibility of linking three areas in the study of dynamic feedback systems. First, dominant polarity bears a clear connection to the notion of open-loop steady-state gain. That connection takes all mystery away from "goal-seeking positive loops." From the point of view of dominant polarity, "positive loops with gain less than one" are no more mysterious than the structure and behavior of the logistic equation. Both are multi-loop systems in which the negative polarity can dominate. In positive loops with gain less than one, the negative polarity always dominates. Second, the notion of dominant polarity aims in the direction of identifying dominant loops and shifts in loop dominance in nonlinear systems. In that sense it is in the spirit, if not yet the significance, of efforts to use eigenvalue elasticities and participation factors to identify dominant loops (N. Forrester 1982). Third, there is the distinct possibility of a rigorous understanding the phenomena of bifurcation and perhaps even mathematical chaos in terms of shifts in loop dominance.

These connections to other ideas about dynamic systems suggest there is reason to develop these nascent notions further. Yet there is one more reason for a serious pursuit of the ideas of dominant polarity and shifts in loop dominance: in applied system dynamics work the concept of shifting loop dominance is an easily communicated, intuitive idea. Shifts in loop dominance and their implications for policy can be described and explained in terms of nonquantitative causal-loop diagrams. The concept of loop dominance becomes an important bridge between complex interactions in a simulation model and the intuitions and understandings of people the modeler hopes to influence. It may help us to move our more significant quantitative advances, such as eigenvalue analyses, bifurcation theory and chaotic systems, from the forefront of research to the arena of applicable policy analysis.

Notes

1. The statement is a principle of feedback system dynamics. It is also something of a tautology, however, because it can be viewed as an implicit definition of what is meant by the phrase "dynamically significant."

2. Define the rate of change $\dot{x}$ to be an inflow to x if $\dot{x}$, whatever its sign, is added x. That is, $\dot{x}$ is an inflow if

$$x(T) = x_o + \int_o^T \dot{x}\, dt$$

whether or not $\dot{x}$ itself is positive or negative. Similarly, $\dot{x}$ is an outflow from x if $\dot{x}$, whatever its sign, is subtracted from x.

3. Nonetheless, in nonquantitative causal-loop diagrams, we often know unambiguously the polarity of feedback loops. The signs of the parameters are usually given by the words used to name or describe the variables.

4. See Goodman (1974) or Alfeld & Graham (1976) for discussion and examples of rate-versus-level curves.

5. Higher-order linear systems also can not change loop dominance, but this simple development of dominant polarity is not sufficient to prove that fact. One way to justify it is to appeal to eigenvalue analysis and participation matrices; see N. Forrester (1982) for developments in the use of these ideas.

6. One simple technique that ought to work with first-order systems containing two loops of the same polarity is to change one of the loops arbitrarily to the opposite polarity in the expression for $\dot{x}$ and compute the shift in dominant polarity as before. Then interpret the result as the point of shifting loop dominance between the two loops of the same polarity. Presumably, a shift in loop dominance between two negative loops means a change in the goal state of the system.

7. The computation is:

dominant polarity = sign($d\dot{x}/dx$)

$$= \text{sign}\left(\frac{f(x)(-1) - (x^* - x)\, f'(x)}{f(x)^2}\right)$$

$$= \text{sign}\left(- f(x) + (x - x^*)\, f'(x)\right)$$

$$= \begin{cases} + \text{ if } f'(x) < \dfrac{f(x)}{x - x^*} \\ - \text{ if } f'(x) > \dfrac{f(x)}{x - x^*} \end{cases}$$

8. Since $\dot{x} = ax - bxy$ and we are only interested in $\partial\dot{x}/\partial x$, this example illustrates the need for partial derivatives when applying these ideas in higher-order systems.

9. The notion of dominant polarity reveals shifts in loop dominance, however, which are not observable in analyses of linearized versions of a nonlinear system. Efforts to blend the strengths of eigenvalue analyses and dominant polarity concepts may bear more fruit than either approach by itself.

10. If $y(t) = y_0 e^{gt}$, then the solution of the differential equation for $z(t)$ is of the form

$$z(t) = \frac{1}{1 + Tg}\, y(t) + ke^{gt}$$

where k is a constant dependent on initial conditions.

11. Root-locus analysis shows that the polynomial

$$k = (1 + T_1 g)(1 + T_2 g) \ldots (1 + T_n g)$$

with positive coefficients T_i always has a real root g that passes from negative to positive as k increases through one.

12. For a further example the reader is invited to apply the technique to the frequently

analyzed salesman loop in Forrester (1968a) and show that loop dominance shifts when the ratio (RS*SE)/SS moves through 1. See Rahn (1982) for further discussion of the salesman loop and loop gain.

13. For recent system dynamics references to bifurcations, see Andersen (1982) and From the Physical Sciences to the Social Sciences: Proceedings of the 7th International Conference on System Dynamics (1982).

References

Alfeld, Louis E. and Allen Graham (1976), *Introduction to Urban Dynamics* (Cambridge, MA: The MIT Press, Wright-Allen Series).

Andersen, David F. (1982), Conference Report — Developing New Concepts in System Dynamics: A review of the 7th International System Dynamics Conference at Brussels, *Dynamica* 8,2 (Winter 1982) 113-116.

Aracil, Javier

Ashby, W. Ross (1956), *Introduction to Cybernetics* (New York: John Wiley and Son).

Forrester, Jay W. (1971,1973), *World Dynamics* (Cambridge, MA: The MIT Press, Wright-Allen Series).

------- (1968a), Market Growth as Influenced by Capital Investment, *Industrial Management Review*, 9,2 (Winter, 1968) 83-115.

------- (1968b), *Principles of Systems* (Cambridge, MA: The MIT Press, Wright-Allen Series)

Forrester, Nathan (1982), A Dynamic Synthesis of Basic Macroeconomic Theory: Implications for Stabilization Policy Analysis, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA. 02139.

From the Physical Sciences to the Social Sciences: Proceedings of the 7th International Conference on System Dynamics, June 16-18, 1982, Universite Libre de Bruxelles, Brussels, Belgium.

Goodman, Michael (1974), *Study Notes in System Dynamics* (Cambridge, MA: The MIT Press, Wright-Allen Series).

Graham, Alan K. (1977), Principles of the Relationship between Structure and Behavior of Dynamic Systems, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA 02139.

Low, Gilbert W. (1980), The Multiplier-Accelerator Model of Business Cycles Interpreted from a System Dynamics Point of View, in Jørgen Randers, ed., *Elements of the System Dynamics Method* (Cambridge, MA: The MIT Press).

Mosekilde, Erik

Rahn, R. Joel (1982), Unity-Gain Positive Feedback Systems, *Dynamica* 8,2 (winter 1982): 92-104.

Richardson, George P. and Alexander L. Pugh III (1981), *Introduction to System Dynamics Modeling with DYNAMO* (Cambridge, MA: The MIT Press).

Richmond, Barry (1980), A New Look at an Old Friend, *Plexus* (Resource Policy Center, Thayer School of Engineering, Dartmouth College, Hanover, NH).

Roberts, Nancy, et al. (1983), *Introduction to Computer Simulation* (Reading, MA: Addison-Wesley).

Samuelson, P.A. (1939), Interactions between the Multiplier Analysis and the Principle of Acceleration, *Review of Economics and Statistics*, 21 (May 1939) 75-79.