SYSTEMS
ARCHETYPES III

Understanding Patterns of

Behavior and Delay

PEGASUS
COMMUNICATIONS

➢ BY DANIEL H. KIM

THE TOOLBOX REPRINT SERIES

Systems Archetypes I: Diagnosing Systemic Issues and Designing High-Leverage Interventions

Systems Archetypes II: Using Systems Archetypes to Take Effective Action

Systems Archetypes III: Understanding Patterns of Behavior and Delay

Systems Thinking Tools: A User's Reference Guide

The "Thinking" in Systems Thinking: Seven Essential Skills

TABLE OF CONTENTS

Dr. Joe Juran, a renowned pioneer and expert on Total Quality Management, often told top management teams that their businesses in effect consisted of two different factories of about equal size. One was designed to produce high-quality products every day of the year. The other was designed to continually churn out nothing but scrap and waste. Framing his observations about built-in waste in this way—especially the claim that the second factory was consciously designed to produce nothing but waste—certainly caught the attention of many top executives. We might find this an absurd concept, because no one would actually design such a system . . . or would they?

ADAPTIVE ACTIONS AT WORK

In his comments about the two different factories, Dr. Juran keenly summarized what often happens when a system is allowed to slowly adapt to its environment without conscious planning or analysis. He noted that when defect rates rise, a plant typically reacts at the event level by increasing product starts and ordering more raw materials. This action makes sense in the short run, because the factory still has to deliver a specified number of good products. Over a long time, however, the defect rate may continue to rise without setting off any alarm bells, because the increased number of production starts "solves the problem." During this period, the factory continually "recalibrates" the cost of producing the product to reflect the additional expenses. After a while, managers accept the higher costs as correct—and so it goes over the years. The company has adapted to higher scrap rates by permanently utilizing a higher volume of raw materials to yield the needed products. They have, in effect, disconnected the alarm bell that would have warned them about a fundamental problem with the system.

Dr. Juran points out that there are many behaviors we may never notice because they have faded into the background of what we have come to accept as "normal." Yet our deepest problems may stem precisely from those behaviors we consider to be acceptable—which is why a number of quality improvement tools (check sheets, pareto analysis, control charts) are specifically designed to help identify "invisible" patterns. The systems archetypes are additional tools that can help you detect invisible patterns in a much broader context.

SEEING THE SIGNATURE PATTERNS

In the first two systems archetypes volumes in the *Toolbox Reprint Series (Systems Archetypes I* and *Systems Archetypes II),* we focus on identifying and understanding the feedback loop structures that produce undesirable behaviors in our organizations. In this third volume, we take a closer look at what I refer to as the signature patterns of behavior associated with each archetype. By internalizing these signature patterns, we can enhance our capacity to detect situations where archetypal systemic forces are driving the (undesirable) results.

Identifying and charting patterns of behavior over time can help us detect adaptive responses that are often invisible in the relatively short time horizons of day-to-day life. Especially in situations where we have either too much of something (inventory, defects, complaints) or too little (time, people, products), constructing a behavior over time graph

can provide the longer time perspective needed for us to see larger patterns. (You might want to refer to the article "Four Steps to Graphing Behavior Over Time," in Volume 8, Number 3 of *THE SYSTEMS THINKER* Newsletter.) These signature patterns can help us identify the archetypal forces at work in a given situation and guide us to more effective, longer term actions.

HOW THIS VOLUME IS ORGANIZED

Because we focus on behavior over time in this volume, the archetypes are organized according to similarities in their dynamic patterns of behavior—specifically, whether they are driven more by reinforcing "engines" or by balancing "corrections."

Reinforcing Engines

The first four archetypes—"Limits to Success," "Tragedy of the Commons," "Growth and Underinvestment," and "Success to the Successful"—are initially driven by the growth engines of reinforcing loops. Each of these archetypes shares a similar overall behavior of initial growth in something, which in turn drives other behaviors in other parts of the system. In the first three archetypes, the growth eventually pushes on limits in the system and produces a balancing response in the form of slowing growth. An important point to understand about these archetypes is that their long-term behaviors depend not only on the specific actions taken but the timing of those actions. The fourth archetype in this group, "Success to the Successful," differs from the other three in that the reinforcing engine does not reach any limits. Indeed, the very downside of this structure derives from the way the "failure" in one of its loops has been "hard-wired" into the structure from the very beginning.

Balancing Corrections

The second four archetypes—"Fixes That Fail," "Shifting the Burden," "Escalation," and "Drifting Goals"—are all driven by balancing forces that are intended to "fix" something in the short term but often produce undesirable results in the long term. All four of these archetypes attempt to stabilize a situation by taking some form of corrective actions. However, they end up worsening things over time, producing an undesirable, reinforcing growth pattern of behavior in the long term. Owing to the delays in the system, people involved in the system tend to continue taking those corrective actions, even though these actions keep reinforcing the problem. This is because the connection is not obvious to them, and/or it is not "politically" wise to cease taking those actions.

HOW TO USE THIS VOLUME

This volume builds on and extends your understanding of systems archetypes gained through the first two volumes. By focusing on the behavior over time aspect of the archetypes, the volume provides you with another entry point for identifying situations in which your organization may be stuck in one or more of these archetypes. Charting behavior over time is often an easier and more natural place for some people to start their systemic inquiry into an issue. We also encourage you to engage others in constructing

behavior over time graphs, again by following the guidelines outlined in "Four Steps to Graphing Behavior Over Time." The conversations and insights a group can gain by drawing these patterns are often valuable in themselves.
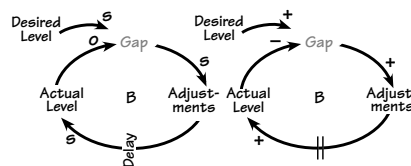
ACKNOWLEDGMENTS

THE LANGUAGE OF LINKS AND LOOPS

A causal link between two variables, where a change in X causes a change in Y in the same direction, or where X adds to Y.

A causal link between two variables, where a change in X causes a change in Y in the opposite direction, or where X subtracts from Y.

R A "reinforcing" feedback loop that amplifies change.

B A "balancing" feedback loop that seeks equilibrium.

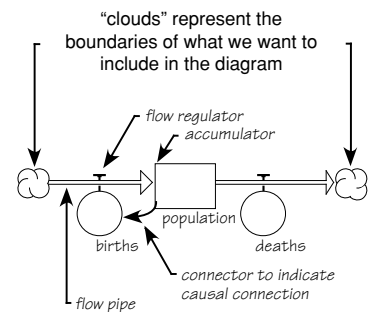### BALANCING LOOP EXAMPLE

If there is a gap between the desired level and the actual level, adjustments are made until the actual equals the desired level. The starting variable is grey.

THE LANGUAGE OF ACCUMULATORS

"clouds" represent the boundaries of what we want to include in the diagram

flow regulator
accumulator
population
births
deaths
connector to indicate causal connection
flow pipe

FROM GROWTH TO UNCERTAINTY: THE DYNAMICS OF "LIMITS TO SUCCESS"

As you may recall, in the "Limits to Success" archetype, certain actions initially lead to success, encouraging the organization to engage in even more of those same efforts. Over time, however, the system encounters limits that slow improvements in results. As performance declines, the organization tends to focus even more on implementing the actions that initially led to growth.

For instance, marketing is a common engine of growth in most companies. Businesses allocate a certain percentage of revenues to marketing and engage in various marketing activities that generate sales (see "Limits to Computer Sales"). More sales lead to increased revenues and a corresponding rise in the marketing budget (R1). However, in our example, as computer sales grow, the installed customer base naturally expands, along with calls for technical support (B2). If the company's technical-support capacity does not increase fast enough to meet the growing demand, the adequacy of that capacity begins to drop. This trend leads to lower customer satisfaction and a downward pressure on computer sales. (We say "downward pressure" rather than a "decline" because sales are likely to continue to grow, although at a reduced or slower rate.)

THREE "REGIONS" OF BEHAVIOR OVER TIME

Understanding the dynamic behavior over time of key variables and grasping how they interrelate accomplishes at least two things. First, it reveals patterns that might help us see similar dynamics occurring in

our own organizations. Second, it shows us how delays in the system can affect our ability to respond to various scenarios, and provides us with guidelines for when and where to implement corrective actions. In the case of "Limits to Success," we can view its behavior over time graph as consisting of three distinct "regions" (see "Behavior over Time Graphs" on p. 8). In this figure, the top graph tracks what happens on the reinforcing side of the archetype, and the bottom one tracks what happens on the balancing side. The dynamic begins with a period of seemingly unhampered growth. The next region is marked by slowing growth and diminishing returns for the efforts expended. The final region involves uncertain outcomes, because the shape of the graphs depends on what happened in the previous regions and on the range of options available at this point. Below we look at the dynamics of each of these regions.

REGION I: UNHAMPERED GROWTH

The early phases of any growth effort can be exhilarating. When a product or service takes off and becomes hugely successful, we feel like we can't do anything wrong. Growth begets more growth, as the reinforcing engine kicks into high gear.

In our example, as marketing activities

increase, sales growth follows. Because we began with ample tech-support capacity, we initially have no problems handling the trickle of calls that come in from our first customers. We may even take pride in the fact that we are steadily increasing our tech-support capacity even when we seem to have more than we need.

REGION II: DIMINISHING RETURNS

The continued growth in sales leads to an even faster increase in the installed customer base. If sales remain constant, the installed base grows linearly. If sales increase linearly (with a positive slope), the installed base grows exponentially. If sales increase exponentially, the installed base grows superexponentially. Because of this relationship between sales and the installed customer base, an organization can easily be caught off guard by how quickly the impact of growth hits it. If tech-support calls are driven more by the number of
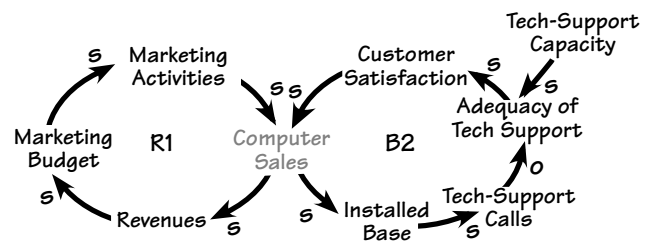
More sales lead to increased revenues and a rise in the marketing budget (R1). However, as sales grow, the installed customer base expands, along with calls for technical support (B2). If the technical-support capacity does not increase fast enough to meet the growing demand, the adequacy of that capacity drops, leading to lower customer satisfaction and sales.

installed customers rather than by current sales, the call volume is likely to increase at more or less the same rate as the installed base.

So, in this region, a company's tech-support capacity is outstripped by the volume of calls. The effect of this pattern shows up in the top graph, where the rate of sales growth begins to slow (sales are still increasing, but at a decreasing rate). When we notice the decrease in sales, we typically respond by stepping up marketing efforts, depicted in the upper graph

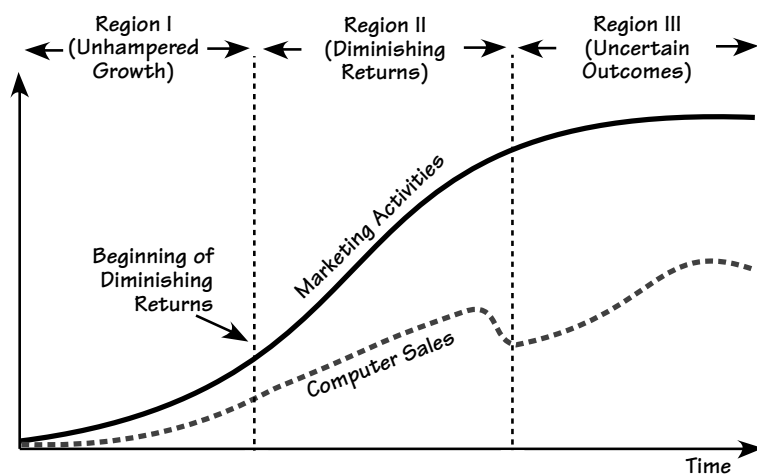with a steeper slope. But notice that the gap between marketing and sales widens over time. Although we do not measure the two variables in the same units, we can still compare the degree of change between the two over time. In Region II, a marked increase in marketing does not result in the same rate of growth in sales that it did in Region I. The growing gap between the two reflects the diminishing returns on marketing dollars spent: It takes more marketing dollars to attract the same number of new customers.
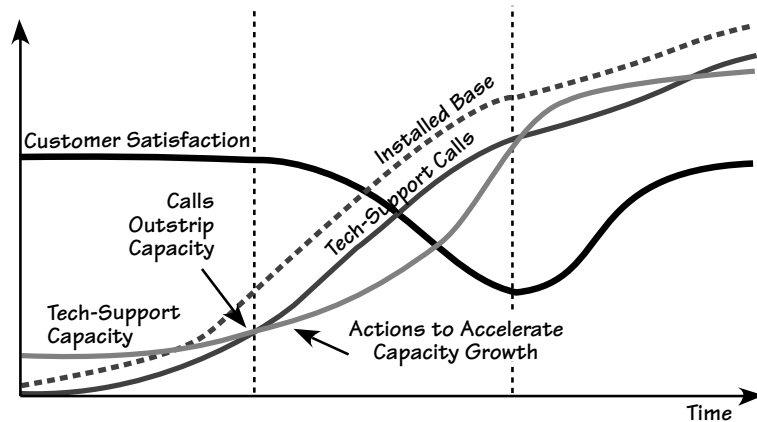
Meanwhile, we can detect one cause for the declining sales in the behavior over time of the variables in the balancing loop. As call volume outstrips tech-support capacity, customer satisfaction starts to decline. We may have noticed the capacity shortfall, and we may have started to expand capacity faster than we had originally planned. But, in the short term, the gap keeps getting wider because it takes time to add such capacity. During this period, customer satisfaction continues to decline, further hampering sales. All the while, we are increasing marketing efforts as we try to chase down more sales. Because of our investments, tech support finally reaches a level where it can adequately meet the call volume. Of course, part of this change is the result of slowing sales. The sad truth may be that sales have declined from what they could have been if we had had enough capacity, indicated by the flattening of the installed base curve.

REGION III: UNCERTAIN OUTCOMES

By the time we have addressed the limit that we have hit (in this case, tech-support capacity), we have entered Region III, where we face many possible future outcomes. How the system behaves from this point forward has a lot to do with the particular business. Regions I and II are relatively predictable because the behaviors are driven by the generic archetypal structures. In Region III, however, the dynamics are also determined by the specifics of the product and the market.

For example, for some products and markets, if we resolve the technical-support problems, customers will resume their buying patterns, albeit at a slower rate than before. This scenario is depicted in our graph, which shows sales recovering as customer satisfaction increases. America Online's situation in the late 1990s

---

BEHAVIOR OVER TIME GRAPHS

REINFORCING LOOP BEHAVIOR OVER TIME

Region I (Unhampered Growth) | Region II (Diminishing Returns) | Region III (Uncertain Outcomes)

Marketing Activities

Beginning of Diminishing Returns

Computer Sales

Time

BALANCING LOOP BEHAVIOR OVER TIME

Customer Satisfaction

Installed Base

Tech-Support Calls

Calls Outstrip Capacity

Tech-Support Capacity

Actions to Accelerate Capacity Growth

Time

We can view the "Limits to Success" archetype as consisting of three distinct "regions" of a behavior over time graph. The top graph tracks what happens on the reinforcing side, and the bottom the balancing side. The dynamic begins with a period of seemingly unhampered growth. The next region is marked by slowing growth and diminishing returns. The final region involves uncertain outcomes.

---

appears to mirror this experience. However, other situations may not turn out as well. Once customers sour on a product or service, the satisfaction level may never rebound, even after a company remedies whatever shortfall they had initially experienced.

Another possibility is that a business may go through several cycles of growth and decline as it either hits the same limit again later or hits a new one. This situation triggers another round of ups and downs, which may set the company on a permanent downward slide as customers become fed up with its inconsistency.

WHAT'S A MANAGER TO DO?

So, how can we make use of this information, whether we are just starting out or are already in Regions I, II, or III? A key lesson of this archetype is to understand the importance of knowing where in the system the impact of growth will be felt the *fastest* and where the delays will be the *longest*.

To begin, we want to find out if growth in one area has a multiplying effect on another part of the system. Otherwise, we risk addressing one rate of growth when, in actuality, we should be concerned with a much higher rate of growth that is occurring elsewhere. In our example, we see that the rise in sales has a strong impact on the installed base, and hence on the number of service calls that hit tech support. If we link growth in tech support to the increase in sales, we will consistently suffer from undercapacity. But by identifying this connection in advance, we can better prepare for the true impact of the growth on all parts of the system.

Nevertheless, even after we have identified the variable that is likely to create a bottleneck, we don't know when to start investing in additional capacity and at what rate. Having too much idle capacity for long periods of time is almost as bad as

having too little. We thus need to understand where the longest delay in increasing capacity exists. In addition, we must design the appropriate feedback channels to provide sufficient lead time for the capacity to come online when it is needed. In our example, if we can quickly boost tech-support capacity in response to an increase in the number of service calls, we can initiate capacity expansions based on the tech-support call volume (see R3 in "Reengaging the Growth Engines"). On the other hand, if it takes a relatively long time for us to upgrade tech-support capacity, we may need to expand capacity based on either installed base (R4) or marketing activities and projections (R5).

Unfortunately, many companies end up relying on negative customer feedback to correct the problem they now face (B6). Or worse still, they wait until sales actually fall (well into Region III) before they get the signal that something needs to be done.

PLANNING FOR SUCCESS

As you plan for growth, automatically assume that something will eventually limit your expansion, and then look for those limits. Study other companies or groups that have embarked on similar ventures to see what limits they may have encountered. Walk yourself through your processes or services, looking at them from the point of view of all members of the organization as well as from that of customers, suppliers, and other outsiders. Ask again

how growth could be limited by factors in a number of other areas: customer service, training, delivery. And don't overlook intangible elements—attitudes, values, beliefs, feelings, and relationships.

Remember, the real leverage in a "Limits to Success" situation does not lie in pushing harder on the engine of growth, but in finding and managing the factor or factors that are limiting success while you still have the time and money to do so. This strategy may involve taking politically difficult steps, such as investing in new capacity before you actually need it or developing new management systems when everyone else is screaming that you need more R&D instead. Mapping the anticipated behavior over time in detail, as we have done here, may help you build a common understanding of these dynamics within your organization and make difficult actions a little bit easier to implement. ■

REENGAGING THE GROWTH ENGINES
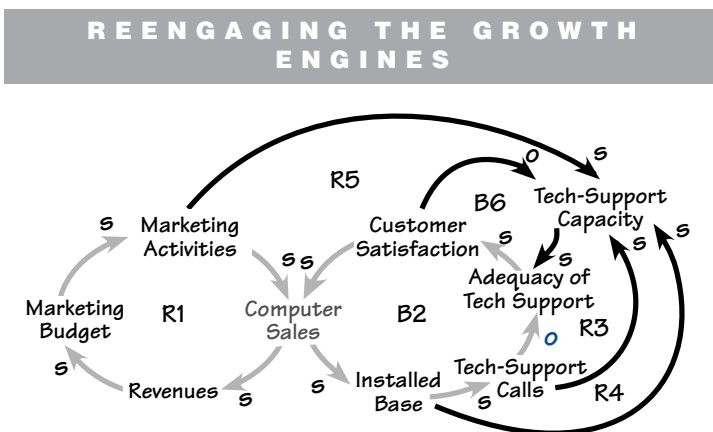
If we can quickly boost tech-support capacity in response to rising service calls, we can initiate capacity expansions based on the tech-support call volume (R3). If instead it takes a long time to upgrade tech support, we may need to expand capacity based on either installed base (R4) or marketing projections (R5). Unfortunately, many companies rely on negative customer feedback to drive the increases (B6).

MACRO OUTCOMES FROM MICRO BEHAVIORS: THE DYNAMICS OF "TRAGEDY OF THE COMMONS"

The "Tragedy of the Commons" structure is a slightly more complex, multiplayer variation on the "Limits to Success" story (see "The 'Tragedy of the Commons' Structure.") All the players in the system discover that they can gain from utilizing a common resource without having to pay much, if anything, in terms of direct costs. The more they utilize the resource, the more they gain from their activity (loops R1 and R2). So to maximize their individual benefit, they continue to take advantage of the commons as much as they want. As long as the total activity or total draw on the resource from all players stays within the carrying capacity or limits of the commons, everything is fine. The "Tragedy" never gets triggered, and the players keep doing what they're doing.

When consumption begins to exceed the resource's ability to replenish itself, however, gains per individual effort will start to decline—fewer fish in the nets, longer time for prototypes to be completed, more mistakes in the documents. When this happens, participants often respond by redoubling their efforts (R3 and R4). They may try to get to the commons faster, initiate more requests for the commons, or just outright grab more of the commons before others get to it. Of course, these kinds of tactics are quickly copied by everyone else, which further accelerates the depletion of the commons. Left unmanaged, these kinds of actions will eventually bring on the collapse of the commons. Then there will be no more gains for the participants, so their activities will finally cease or be greatly curtailed (B5 and B6).

OF LIMITS AND TRAGEDIES

Both the "Limits to Success" and "Tragedy of the Commons" archetypes are affected by limits, but the two archetypes differ in an important way. In "Limits to Success," the limits encountered are ones that could be expanded through judicious planning and timely investments in the resource. The primary lesson of "Limits" is therefore about balancing capacity investments and growing demand such that future growth is not hindered by inadequate capacity. In "Tragedy of the Commons," the limits are considered "fixed" during the relevant time frame. This archetype's primary lesson is therefore about managing consumption of the resource in a way that never allows the system to enter the undesirable region of rapid decline. Thus, behavior over time graphs for the two archetypes are quite distinct.

THREE REGIONS OF BEHAVIOR OVER TIME

In "Tragedy of the Commons," we want to pay attention to how three key variables change over time (see "Behavior Over

THE "TRAGEDY OF THE COMMONS" STRUCTURE
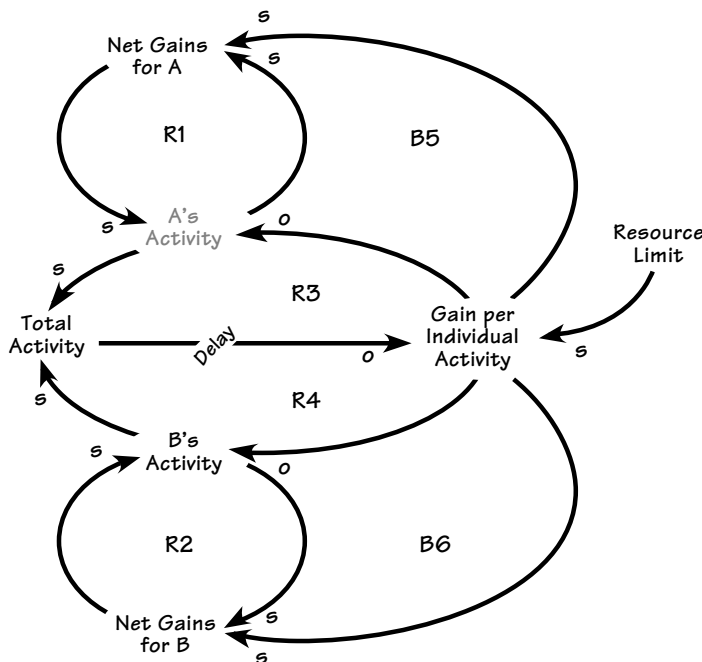
In previous versions of this archetype, the accelerating R loops (R3 and R4) were left as implicit parts of the structure. We are making them more explicit here because we want to link these loops more precisely to the patterns of behavior over time that they generate.

Time"). Total Activity represents the sum of all the individual parties' efforts. Paying attention to changes in Common Resource provides an indicator of what is happening to the "commons" itself. Gain per Activity measures the delayed impact of everyone's use of the commons. There are three distinct phases to the timeline that are significant in a "Tragedy of the Commons" archetype.

Phase I: Stability. In all cases, we start out in the stable phase, where the total activity is small relative to the resource available. In this phase, increases in total activity do not decrease the gains per activity, and people are not aware that limits may even exist. This stage can go on indef-

> **Not only are we consuming the resource faster than it can replace itself, but our consumption is actually causing the replacement rate to decrease.**

initely, as long as our consumption rate remains slower than the resource's regeneration rate (e.g., we fell fewer trees than are planted, or demand less overtime hours than the employees' rejuvenation rate).

Phase II: Gradual Decline. When our consumption of the resource becomes greater than its replacement rate, we enter the region of gradual decline. Here, the resource level begins to drop imperceptibly at first, but with greater velocity as total activity accelerates. In this region, the growth in total activity is fueled largely by the increasing popularity of the resource, both among the current users as well as new ones.

Phase III: Rapid Decline. At some point, the consumption reaches a level such that it affects the regeneration rate itself. In other words, not only are we consuming

the resource faster than it can replace itself, but our consumption is actually causing the replacement rate to decrease—which means the resource will get depleted at an even faster rate. When this happens, we have entered the third and final region of rapid decline, which is the precursor to a total collapse if corrective actions are not taken. In this region, the total activity is heavily influenced by the growing scarcity of the resource (which shows up in the precipitous decline in Gain per Activity). This scarcity can lead to a panicked consumption characterized by an "I've got to get my share before it's all gone" herd mentality. Total activity grows superexponentially and then drops sharply as the commons collapses.

## MACRO OUTCOMES FROM MICRO BEHAVIORS

The "Tragedy of the Commons" structure deserves special attention because it represents a macro view of a dynamic behavior that is being produced by a lot of individual actors at a micro level. In "The 'Tragedy of the Commons' Structure," for example, A and B represent dozens or thousands or more of individual actors, each of whom is enjoying the benefits of using the common resource. As each player gains from his or her activity, there is a tendency to increase the activity level, because gains increase without a proportionate increase in costs (R1 and R2). In addition, the number of participants is likely to increase, as others begin to hear about

the gains to be had. Both of these tendencies accelerate the increase in the Total Activity level, which will eventually lead to a decrease in Gain per Individual Activity.

Now, if this were a "self-correcting" system (see "Self-Correcting Versus Accelerating Forces" on p. 12), the outside balancing loops (for example, B5) would kick in. The decrease in Gain per Individual Activity would translate into a decrease in individual net gains, leading to a decrease in individual activity and Total Activity—which would eventually lead to an increase in Gain per Individual Activity.

Unfortunately, the setup of this archetype encourages people to do exactly the opposite and increase their activities in the belief that they can compensate for the fall in gains with greater efforts (R3, for example). This strategy does appear to pay off in the short run as long as we stay in Phase II. Unfortunately, the dynamic often continues into Phase III.
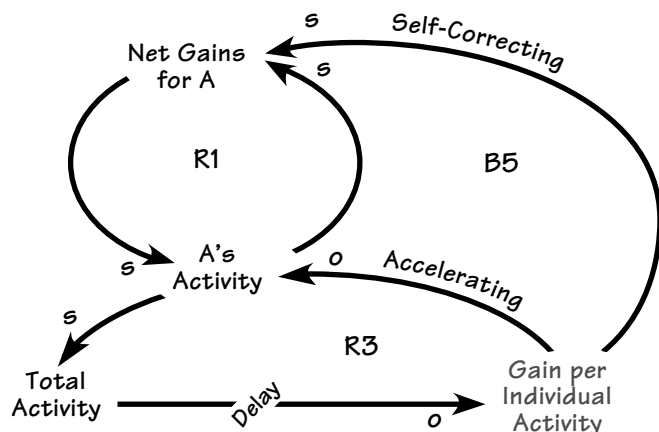
Even without the more usual "acceleration" dynamic, the "Tragedy of the Commons" archetype has a propensity for overshoot and collapse because of the delay


BEHAVIOR OVER TIME
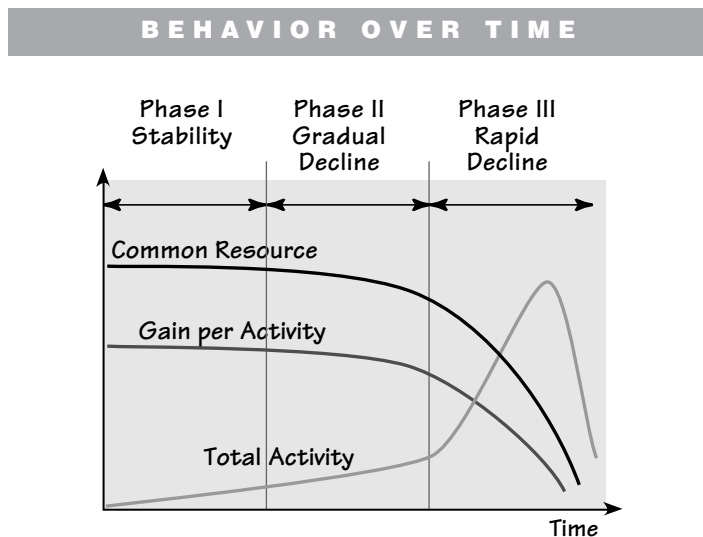
A "Tragedy of the Commons" situation progresses through three stages: stability, gradual decline, and rapid decline. If corrective actions are not taken, the common resource will eventually collapse.
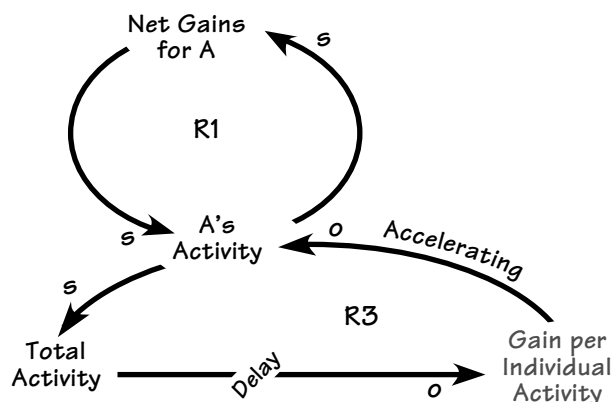
TWO COMPETING FORCES: When gain per individual activity decreases, there are two possible paths—self-correcting and accelerating—for the energy to move through the system.
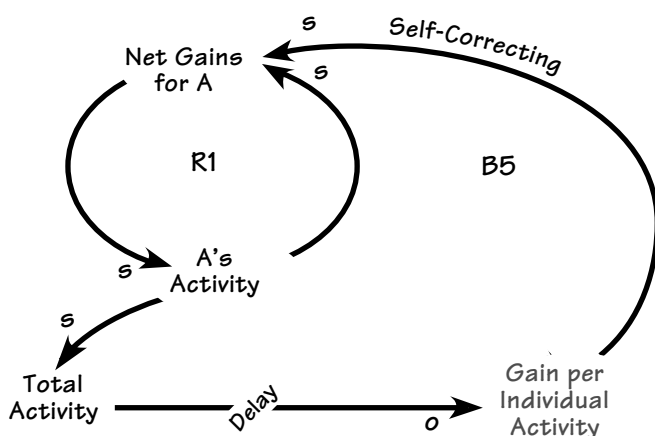
SELF-CORRECTING DYNAMICS: When gain per individual activity decreases, this path sends the self-correcting signal to decrease individual and then total activity. R1 goes in a decreasing reinforcing spiral until the system recovers and begins to yield higher gains.

ACCELERATED DEPLETION DYNAMICS: When gain per individual activity decreases, this path encourages individuals to compensate for the loss by increasing individual activity. R1 goes in an increasing reinforcing spiral until the resource is completely depleted.

between when the total activity level has risen beyond a sustainable level and when that feedback shows up in the way of diminishing returns on the activities. By the time those indicators do show up, there is so much momentum in the consumption activity that it is extremely difficult to get any participants to voluntarily reduce their activity.

MANAGING THE COMMONS

To detect a "Tragedy of the Commons" situation at work, look for two key factors: a common resource that two or more players have relatively free and equal access to; and the absence of any larger oversight or management responsibility to a single governing authority. The primary challenge in "Tragedy of the Commons" is coming to collective agreement on exactly what common resource is being overburdened and on what to do about the overuse. If people do not see how their individual actions will eventually reduce everyone's benefits, the level of debate is likely to revolve around why player B thinks that player A should stop doing what she is doing and why it's okay for player B to keep doing what he is doing. Chances are that there may be endless debates about one course of action or another, but little actual change.

This structure provides a critical insight: It shows us that the solution for averting a "Tragedy" does not lie at the individual level. As long as the system is designed to provide immediate individual gain without a way to make the long-term collective pain more evident, the players will keep using the resource. The critical and challenging steps are (1) to get the whole community of users to identify and agree on the commons, its limits, and the real potential for its depletion; (2) to assess the impact of the collective use of the commons; and (3) to design the mechanisms required to measure and manage the usage. ◻

"GROWTH AND UNDERINVESTMENT": THE ROLE OF RELATIVE DELAY

The "Growth and Underinvestment" structure is a direct extension of the "Limits to Success" archetype (see "The 'Growth and Underinvestment' Structure"). In "Growth and Underinvestment," we focus explicitly on the dynamics of investment decisions as the system reaches certain limits. As in the "Limits to Success" archetype, a business's financial performance slows down as the system approaches a limit, such as customer service capacity. Performance could continue to grow if the organization were to invest to expand the capacity of whatever is limiting further growth, for instance, by adding additional phone lines and service representatives. Unfortunately, the organization makes these investments only after a significant delay. The delay leads to a further decline in growth, causing additional lags in investment or even reductions in capacity—and even worse performance.

INVESTING IN THE FACE OF FALLING DEMAND

The reinforcing loop (R1) in the "Growth and Underinvestment" structure is the growth engine: It includes the growth effort itself, such as marketing, and the demand, such as sales, which is the variable the organization wants to grow. The growth generated by the reinforcing loop in turn strains current limits on capacity. Additional growth is thus slowed by the impact of the limiting factors represented in loop B2; for example, sales may be constrained by limits to service quality.

What distinguishes "Growth and Underinvestment" from "Limits to Success" is the additional balancing loop

(B3)—the investment loop. This part of the structure provides a mechanism for reducing the impact of the limiting factor through the allocation of additional resources. So, when the limiting factor—for example, service quality—starts to have a noticeable effect on the organization's performance, the perceived need to invest in capacity to overcome this limit rises. The company then invests in additional capacity—they hire more customer service representatives. Delays play a crucial role here, because it takes time for people's perceptions to change and for them to then translate their new understanding into actual investment decisions. And, once decisions have been made, it takes additional time for new capacity to come online. As capacity finally expands, the constraints imposed by the limiting factor are reduced, which can reverse the direc-

tion of B2 and reinvigorate growth.

One of the primary reasons why this structure is so difficult to manage successfully is that it often requires you to *add* capacity during times of *falling* demand. Most organizations are reluctant to invest in capacity even when times are good; when customers seem to be abandoning you, it can be nearly impossible to feel confident about committing more resources.

THE DYNAMICS OF GROWTH AND UNDERINVESTMENT

The interplay among the three loops in this structure can create a wide range of behaviors. A graph of the distinctive behavior over time of "Growth and Underinvestment" includes at least four variables: demand, capacity, perceived need to invest, and capacity investments/reductions (see

THE "GROWTH AND UNDERINVESTMENT" STRUCTURE
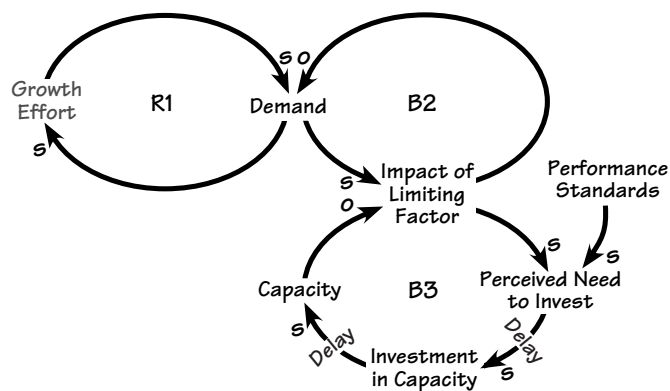
The reinforcing loop (R1) is the growth engine: It includes the growth effort itself and the demand. The growth in turn strains current limits on capacity and is thus slowed (B2). Loop (B3)—the investment loop—provides a mechanism for reducing the impact of the limiting factor through the allocation of additional resources.

"Behavior over Time for 'Growth and Underinvestment'"). As we will see, the dynamic behavior of the variables involved in this structure are more tightly linked than in most of the archetypes. In particular, the perceived need to invest drives capacity investment, which in turn affects capacity and, ultimately, demand. The gap between demand and capacity then influences perceived need to invest, thus closing the loop.

An organization may implement one of several strategies for pursuing growth. For example, it may peg growth at a steady level through policy regardless of the level of current demand or capacity, or it may seek to increase growth whenever demand is flat. Alternatively, a company may have demand drive growth efforts, increasing those efforts when demand goes up and decreasing them when it goes down.

In "Behavior over Time for 'Growth and Underinvestment,'" Time "A" marks the point when demand begins to exceed

capacity, which in turn causes the perceived need to invest to rise. After some delay (depending on the organization), capacity investments are made and capacity starts to increase. When capacity begins exceeding demand at Time "B" (because demand is falling due to shortages, and capacity is rising due to investments), perceived need to invest drops to zero. But because of delays, capacity continues to grow, as investments in the pipeline come online. With excess capacity, the company is better able to serve its customers, so demand picks up again until it once more exceeds capacity at Time "C"—and the cycle repeats.

FROM GROWTH TO "DRIFTING GOALS"

These ups and downs can be dangerous, because in today's competitive environment, customers may not return after experiencing the delays inherent in this archetypal structure—even after the company has expanded capacity. Thus, the peaks of demand may become smaller and smaller each time, as customers defect to other suppliers. The company could then find itself in a downward spiral of shrinking demand and falling capacity. Specifically, when demand fails to recover after Time "D," the company cuts capacity below the demand level (Time "E"), causing demand to fall even lower. With even

fewer customers than before, the company feels compelled to institute yet another round of capacity cuts.

For those familiar with all the archetypes, the dynamics described above may seem similar to "Drifting Goals." In a way, the two balancing loops in "Growth and Underinvestment" (B2 and B3) function much like the "Drifting Goals" structure, in that the deteriorating performance —or declining demand (B2)—causes us, over time, to lower our performance standards, thereby shortcircuiting the force that normally drives our perceived need to invest in capacity (B3). The difference is that, in a "Growth and Underinvestment" scenario, the leverage for improving the system lies primarily in maintaining the performance standards loop (B3). Why? Because this archetype presumes that nothing more can be done in the other two loops (R1 and B2) to improve the situation until the capacity limitation is addressed.

In extreme situations, companies can actually destroy themselves by slashing capacity investment or not investing early enough. We can see this in "The 'Growth and Underinvestment' Structure" when the impact of the limiting factor causes a decrease in the performance standard, leading to a vicious downward spiral.

Here's how this scenario plays out: When the limiting factor causes performance to decline, there is pressure to lower the performance standard rather than to invest in capacity. The two balancing loops then begin to function as a single reinforcing loop that has an insidious effect: The organization responds to a decline in growth by continuing to withhold or withdraw investment. This response leads to even lower growth (or even a decline), which, in turn, causes a further drop in performance standards. This dynamic can choke the growth engine completely and ultimately prove fatal for the enterprise (as shown in the latter part of the Behavior
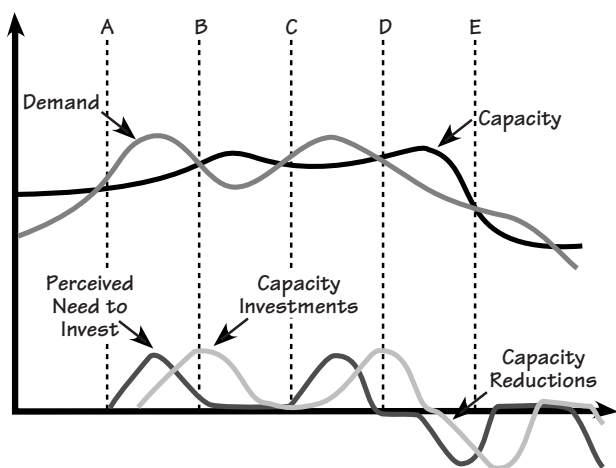
The ups and downs inherent in this archetypal structure can be dangerous, because customers may not return after experiencing the delays— even after the company has expanded capacity. The company could then find itself in a downward spiral of shrinking demand and falling capacity.

over Time diagram, beyond Time "E"). This eroding-performance-standards dynamic has been exemplified by several high-tech companies that sold divisions and laid off employees until the organization could no longer function.

MANAGING DIFFERENT CYCLE TIMES

A key challenge in managing this archetypal situation effectively is understanding the relative time delays in the two balancing loops and appreciating the importance of managing the interrelationship between the two. For example, "Managing Cycle Times" shows an organizational dynamic in which increases in order backlog lead to customer defections.
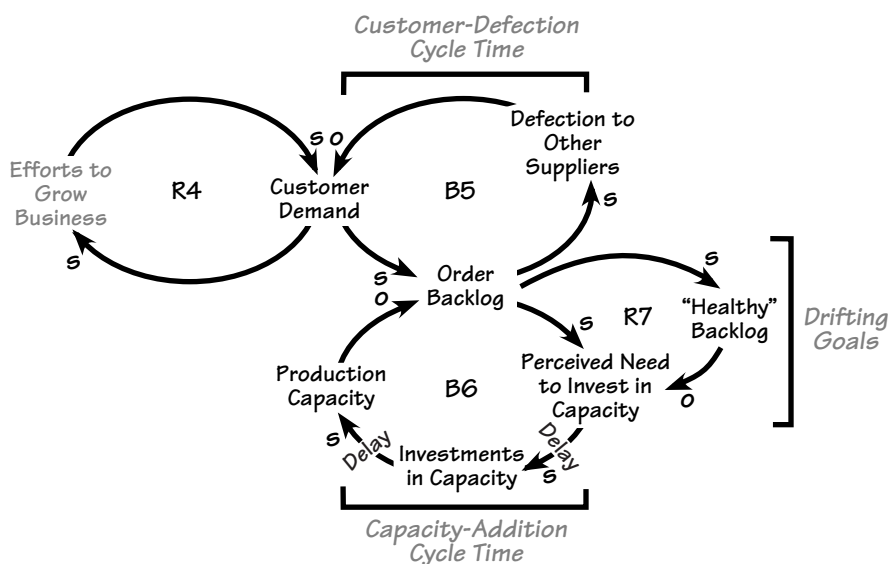
If the company can make capacity adjustments in less time than it takes customers to defect, then they should be able to respond to the backlog problem without suffering many defections. This is because their "healthy" backlog standard would send the signal that additional investments are needed. They would make their investments—and the added capacity would kick in—before the backlog translated into defections. In other words, as long as the capacity-addition cycle time is less than the customer-defection cycle time, this system works—with one caveat. The caveat has to do with the potential of this structure to degrade into a "Drifting Goals" dynamic. In this scenario, the definition of "healthy" drifts over time, allowing the actual backlog to grow higher. This increase in what is deemed an acceptable backlog slows the overall capacity-addition cycle time. Over the long run, the capacity-addition cycle time may end up being longer than the customer-defection cycle time—but no one notices this critical change; instead, they assume that the policy is still robust. Therefore, organizations must pay vigilant attention to the changing relationship between the two cycle times.

If, on the other hand, the company is already in a situation where the capacity-addition cycle time is greater than the customer-defection cycle time, then they are likely to suffer the negative consequences of this archetype. By the time they perceive the need to invest and launch those investments, customers are already defecting. The defection then causes the backlog to decrease, which lowers the perception that further investments are needed. This dynamic traces out a figure-eight reinforcing spiral of lower demand and lower investments. In such a case, two high-leverage actions are possible: (1) finding a leading indicator that provides an earlier signal of impending undercapacity, thereby reducing the perception delay, or (2) finding ways to shorten the capacity-acquisition delay itself. Either way, the main goal is to make the capacity-addition cycle time shorter than the customer-defection cycle time.

THE IMPORTANCE OF RELATIVE DELAYS

We all understand the negative consequences of excessive delays in a system; numerous books and articles have been written about finding ways to squeeze out many sources of delay. The key lesson of the "Growth and Underinvestment" archetype is that what may be more important than the absolute measure of delays is the assessment of delays *relative to other delays*. In other words, the way we define excessive delay is always based on some reference point. The reference point we use may or may not be appropriate. Assessing what is appropriate requires a deeper understanding of the relationship among various delays. This archetype provides a structural view of this challenge, which can then facilitate conversations about making those important assessments. ◘

MANAGING CYCLE TIMES

In this example, increases in order backlog lead to customer defections. If the company can make capacity adjustments in less time than it takes customers to defect, then they should be able to respond to the backlog problem without suffering many defections. Otherwise, by the time they perceive the need to invest and launch those investments, customers are already defecting.

A TALE OF TWO LOOPS: THE BEHAVIOR OF "SUCCESS TO THE SUCCESSFUL"

"It was the best of times. It was the worst of times. . . ." So begins Charles Dickens's classic novel *A Tale of Two Cities.* Unfortunately, in the "Success to the Successful" archetype, the best and the worst of times are often hard-wired into the structure, so that it is always the best of times for one alternative and the worst of times for the other. To understand this "tale of two loops," let's consider a common—and timely—example.

NOT-SO-NEW NEW YEAR'S RESOLUTIONS

As we start a new year, many of us take the time to jot down some New Year's resolutions. If you are like most people, you may find that a few of the items on your list were also there last year—and the year before and the year before that. Why can't we make the changes that we have "resolved" to make and that, in most cases, we have the power to accomplish?

One easy response may be that we don't really want to do some of the things that we commit to doing. We come up with "politically correct" items like eating less red meat and more organic vegetables so that we can dutifully produce our list when someone asks us, "Are you making any New Year's resolutions?" Let's remove those gratuitous pronouncements from consideration and look at the changes in behavior that we really *do* want to accomplish, such as losing weight. Are we just too lazy or weak-willed to fulfill our commitments? Before we berate ourselves yet again, we may want to examine our situation from a systemic perspective. The

"Success to the Successful" archetype can help us understand the structural forces that are preventing us from carrying out our well-intentioned resolutions.

ORGANIZATIONAL LAW OF INERTIA

The "Success to the Successful" structure is largely driven by inertia. In physics, the principle of inertia means that, barring outside influences, an object that is in motion will tend to stay in motion; an object that is at rest will generally stay at rest. In the case of "Success to the Successful," the person or project that initially succeeds will continue to succeed. On the other hand, the person or project that, for whatever reason, gets a late start will tend to fail.

> **Why can't we make the changes that we have "resolved" to make and that, in most cases, we have the power to accomplish?**

So, it's easier to maintain existing habits (such as eating more and exercising less than we should) than to establish new ones (like sticking to a diet and walking at lunch time). The structure of our organizational systems—and our own mindsets—contributes to the forces that produce these predictable results. To better understand how this dynamic works, let's take a detailed look at the behavior over time of this archetype (see "Initial Dynamics of 'Success to the Successful'").

The center of the diagram illustrates the overall results of this archetype: As the resources dedicated to A and A's success both increase, the resources invested in B and, in turn, B's success decline. The insets above and below this graph provide a more detailed look at the initial dynamics that play a critical role in this long-term outcome. We devote resources to A (which represents the original or the more favored person, product, or activity) for some time with no visible success; therefore, in the beginning, the net returns for A are low or even negative. Sustained investment, however, eventually leads to A's success. The key here is that, if we sustain our investment in A beyond a critical point, A begins to generate positive returns. Beyond this critical threshold of positive returns, A's success is likely to be self-sustaining, because continued investment brings ongoing positive net returns.

In the case of B, we start by making the same initial investments as for A but, for whatever reasons (poor timing, external forces, the effects of learning curves, etc.), B takes longer than A to become successful. In many cases, the reason for A's comparative success is that it had a head start in and is already beyond the critical threshold of positive returns. Thus, B's net returns stay low or negative longer than A's, and B begins to look less attractive as an alternative. As a result, we decide to invest less and less in B, which delays B's achievement of success even further. At a certain point, we may even begin to take away resources, such as people and equipment, because we don't want to waste

them on a "lost cause." In turn, B's performance only declines further. We eventually conclude that B is a failed experiment and abandon it.

In the case of our New Year's resolutions, the success of our old habits in giving us satisfaction makes it difficult for our new efforts to produce equally compelling benefits in the first few months. So, we may pat ourselves on the back for trying to drop a few pounds, mutter something to the effect of "It just wasn't meant to be," and comfort ourselves with another hot fudge sundae.

OVERCOMING THE "SURVIVAL OF THE FITTEST" MENTALITY

In a way, the "Success to the Successful" archetype helps show why something that looks like a fair and equal setup is often rigged to favor one party over another. The imbalance can stem from some random external event, a personal bias, or simply the momentum of the first party's current success. Unfortunately, many management decisions are based on a "survival of the fittest" mentality that ignores the effects of this initial imbalance. As a result, we may not end up with a particular person, product, or activity because it is the "fittest," but rather because it was either the first or the most widely available option. In this way, we may ultimately accept an inferior outcome over what could have been—and possibly ruin a career or two along the way.

In order to achieve the best possible outcome, we need to be sure that we gave the second alternative a fair shake instead of dooming it to failure from the start. This is particularly important when A is already well established, because any comparisons of B to A tend to make B look less appealing. In this case, comparing A and B would be like judging the performance of a five-year-old child against that of a ten-

year-old and concluding that the younger child is inferior and not worth further investments. But, in actuality, the five-year-old may be much better at accomplishing the task than the older child ever was at that same age. Without separating our evaluation of B's performance from A's, we may end up sticking with current levels of competency at the expense of developing competency for the future.

People and organizations often suffer from this "competency trap" because, in the short run, it seems to make more sense to invest in something that is already successful than in something new and untried. The downside of this tendency is that we may unwittingly continue to use adequate but inferior tools or methods simply because we are familiar with them. This inclination can have dire consequences when we fail to invest in newly emerging competencies (e.g., when IBM was slow to recognize the importance of personal computers).

To break out of a competency trap, we must clarify our goals for the new product or initiative and identify the resources needed to achieve those objectives. We then must examine how the success of the current effort may systematically undermine support for the new initiative, and find a way to decouple the two.

"Success to the Successful" raises questions about what drives success in certain situations. It also shows how, if we are not clear about the overall result that we are trying to achieve, the differences in initial conditions alone can have powerful long-term effects on the outcome. Finally, this archetype illustrates how we can persuade ourselves to stay in old lines of business or outmoded ways of doing things simply because we are already good at them. To escape from this trap, we need to look beyond what works and clarify what we actually want in the longer term. We may then be in a better position to keep some of our resolutions this year—and next. ◘

INITIAL DYNAMICS OF "SUCCESS TO THE SUCCESSFUL"
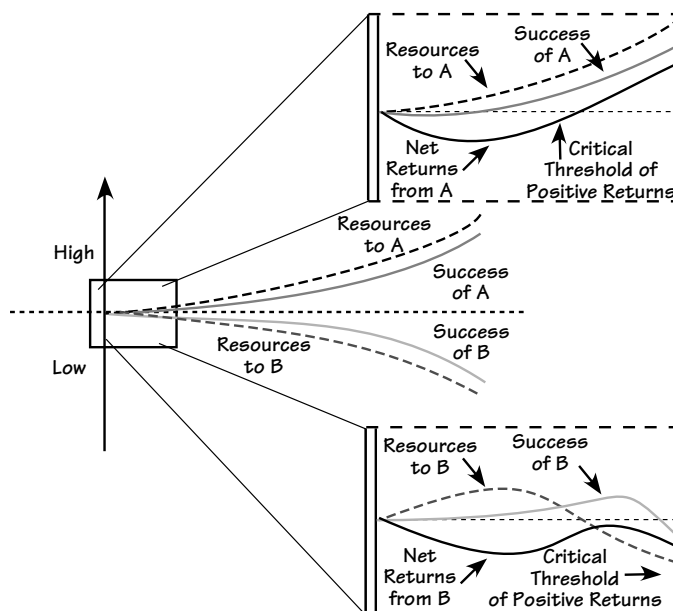
The center of the diagram illustrates the overall results of this archetype: As the resources dedicated to A and A's success both increase, the resources invested in B and, in turn, B's success decline. The insets above and below this graph provide a more detailed look at the initial dynamics that play a critical role in this long-term outcome.

"FIXES THAT FAIL": WHY FASTER IS SLOWER

Most of us are familiar with the paradox that asks, "Why is it that we don't have the time to do things right in the first place, but we have time to fix them over and over again?" That is, why do we keep solving the same problems time after time? The "Fixes That Fail" archetype highlights how we can get caught in a dynamic that reinforces the need to continually implement quick fixes.

THE STORYLINE

In this structure, a problem symptom gets bad enough that it captures our attention; for example, a slump in sales. We implement a quick fix (a slick marketing promotion) that makes the symptom go away (sales improve). However, that action triggers unintended consequences (diverts attention from aging product line) that make the original symptom reappear after some delay—often worse than before.

Some people know this dynamic from mismanaging their finances. Whenever they run short of cash, they use their credit cards to "solve" this shortfall. Unfortunately, the additional debt increases their monthly credit-card payments, causing them to run short of cash the next month. They again "fix" the problem by using their credit card to cover an even greater shortfall (because more dollars are going to pay the finance charges on the debt). Many juggle their debt among several credit cards by paying one card off with checks written on another. But with each round, the debt burden grows heavier and heavier, which may be why we currently have the highest consumer debt levels in history and record personal bankruptcies—all in a booming economy! This is the basic storyline of the "Fixes That Fail" archetype. Let's take a closer look at how and why this systemic structure behaves the way it does.

HEROES AND SCAPEGOATS

Many managers report that their organizations experience certain problems over and over again. Most seem to accept these challenges as a fact of life. Only a few see the cause as "hard-wired" into their businesses. However, from a systemic perspective, whenever patterns of behavior recur over time, they must be driven by structures that are designed into the way the system operates—intentionally or not. To better understand why we would create such structures, we need to take a closer look at the behavior of this archetype (see "The Hero-Scapegoat Cycle").

Organizations usually have target levels—for example, inventory—against which they monitor performance. If a problem symptom exceeds its desired level, such as excess inventory, we may notice but not act on it right away, because we're focused on other crises. When the symptom eventually reaches crisis proportions, we then shift our attention to that problem. At this point, because the situation has become so dire, we often look for someone who can "save the day" (e.g., slash inventory levels). Sure enough, we find a person who can drive the symptom down to the desired level in a hurry, and then reward her with a promotion.

In the meantime, the delayed consequences of the hero's actions (lack of product availability due to low inventories) begin to have an impact, and the problem symptom returns (higher inventory levels). When it again reaches the crisis level, we blame the person who is currently oversee-

THE HERO-SCAPEGOAT CYCLE
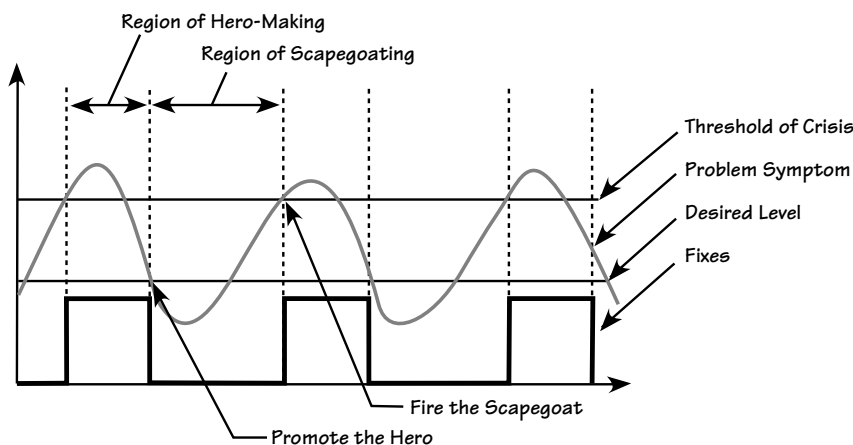
When a problem symptom becomes a crisis, we look for a hero to drive it back to acceptable levels using quick fixes. By the time the unintended consequences of those fixes cause the problem symptom to reach crisis level again, we've promoted the hero. We therefore scapegoat the new manager for failing to keep the problem under control.

ing that function for failing to do his job, fire him, and look for our next hero. However, in this archetype, it may well be that the first "hero" is the person who put the current crisis in motion and that the "scapegoat" is the person who set the stage for a more lasting solution to take hold. But, because of delays in the system, these realities are often obscured.

WIN TODAY, LOSE TOMORROW

So, why do so many organizations fall into the "Fixes That Fail" trap? Why can't people recognize the vicious cycle that keeps repeating the same patterns of events? One of the reasons is that the delays in the system mask the true nature of the cause-and-effect relationship. The narrow time frames that often drive decision-making in organizations also compound the problem.

For instance, our results are likely to deteriorate over time if the delay for the unintended consequences to affect the system is long. This is because, without the feedback supplied by the unintended consequences, the "improvements" actually appear to make things better in the short term (see "Fixes That Fail over Time"). And yet, when we view the situation from a longer time horizon, we find that today's "desired" levels are far higher than yesterday's "crisis" levels. From a longer perspective, we see that those short-term successes are part of a series of steps toward long-term failure. This pattern shows how companies can go bankrupt even as individuals are continually rewarded for doing a great job.

WHAT ALARM BELLS?

Another problem associated with this archetype can occur even when we do make changes so that quick fixes are no longer needed. Now, this may sound like a good thing, but it all depends on how we do it. Unfortunately, many organizations solve the problem by adapting to the poorer performance level, which then becomes the new norm (or desired level).

For example, we may have had a desired first-run capability of 95 percent or better from our production line (that is, 95 percent of our motorcycles run the first time off the assembly line), but we often find ourselves operating at a crisis level of only 90 percent. Because our plans are based on the higher level, our ability to provide predictable performance drops.

In order to improve predictability, we lower our desired level to one we know we can achieve (90 percent), with plans to eventually bring our capability back up to 95 percent. The danger of such a move is that once we have factored the poorer performance into operating plans, it becomes less visible as an issue that needs attention. In other words, what once caused alarm bells to ring no longer rings any bells, because we have in effect disconnected them. Although we no longer reach the crisis level or require frequent fixes, we have embedded the poorer performance in our system, and we no longer notice it.

In this situation, we have fixed our problem by getting caught in a different archetypal structure called "Drifting Goals."

We end up "fixing" things by changing our criteria of what constitutes a crisis.

FINDING FIXES THAT LAST

Of course, the answer is not that we should never apply quick fixes. There are many circumstances for which we absolutely have to implement short-term solutions. The danger lies in failing to recognize that all quick fixes are merely stopgap measures that buy us time to get to the root causes of those problem symptoms.

One of the most important points to address about this archetype is the relationship between the delay for the unintended consequences to show up and the timing of organizational performance assessments. If you suspect that you may be caught in a "Fixes That Fail" dynamic, look for a repeating pattern of quick fixes, determine how often these fixes occur, and compare that to the frequency with which you typically review performance. If the review time horizon is about the same as or shorter than the time between fixes, then try lengthening the time frame so that it's at least three or four times the delay period. This will help provide a more accurate picture of the actual "progress" being made. ◘

FIXES THAT FAIL OVER TIME
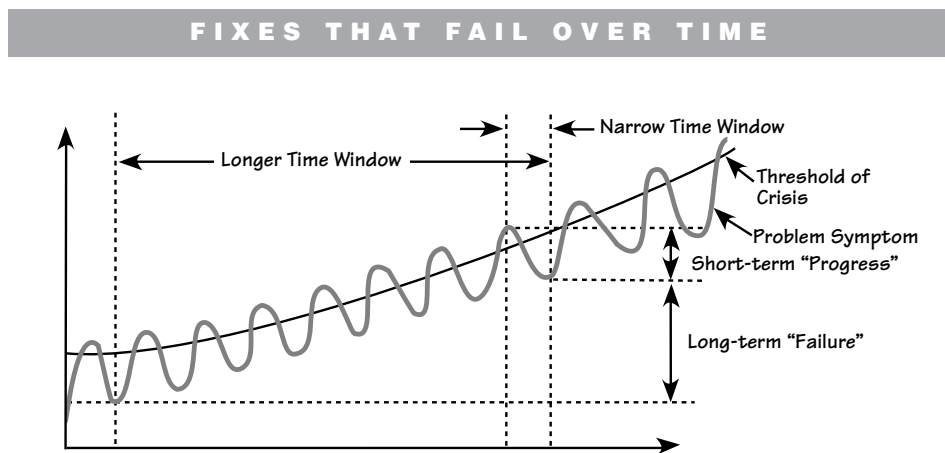
Over the short term, we applaud the progress we are making. And yet, when we view the situation from a longer time horizon, we find that the current "desired" levels are far higher than yesterday's "crisis" levels used to be. Thus, those short-term successes are actually part of a series of steps toward long-term failure.

"SHIFTING THE BURDEN": MOVING BEYOND A REACTIVE ORIENTATION

Although the parable of the boiled frog has become a familiar story in organizational learning circles, it does not yet seem to prevent organizations from suffering the same fate. The story goes that if you toss a frog into a pot of boiling water, it will jump out to save itself. However, if you put it in a pot of luke-warm water and slowly turn up the heat, the frog will happily swim around until it boils to death. Not a pretty picture, especially if you don't enjoy frog's legs. So, why doesn't the frog jump out in the second instance? The reason is that it is designed to detect sudden, large shifts in temperature, not small, gradual changes. So, it never senses the danger in the second scenario until it's too late to respond.

As Peter Senge points out in *The Fifth Discipline,* many organizations suffer from the same learning disability as the frog. Their internal detection mechanisms are geared for responding only to quick, dramatic changes in their environment, not to slow, more incremental ones. Hence, the same businesses that would sound all kinds of alarm bells if they experienced a sudden 5-percent drop in market share will quietly adapt to an annual 0.5-percent erosion over 10 years without recognizing this slow downward spiral as a crisis.

Although the boiled-frog syndrome has often been associated with the "Drifting Goals" archetype, there are many ways in which an organization can "get boiled." In the "Shifting the Burden" structure, things seem to improve in the short term, even as the water gets hotter

and hotter over time. Therefore, this archetype warns us about the long-term consequences of relying on a symptomatic approach to addressing problems.

THE URGENCY OF *NOW*

The "Shifting the Burden" systems archetype produces behavior quite similar to that generated by the "Fixes That Fail" structure (see "Fixes That Fail: Why Faster Is Slower" on p. 18). Both archetypes tend to cause people to take actions in response to acute problems, and both tend to reinforce the use of quick fixes. In this way, the two archetypes are driven by the urgency of the here and now, which leads to unintended consequences that end up making the original situation worse in the future.

The difference with the "Shifting the Burden" archetype is that it requires a deeper understanding of what's needed to keep the system healthy than does the "Fixes That Fail" structure. This is because addressing a "Shifting the Burden" scenario often necessitates identifying not so much a solution to a problem but rather the fundamental capability that the organization needs to develop over the longer term.

PRUDENT OUTSOURCING OR SHIFTING THE BURDEN?

Let's look at an example. In "Boiled Through Outsourcing," we see a situation in which a refrigerator manufacturer faces a shortage of engineers to work on a new product design (labeled "Problem-Symptom Peaks" in the diagram). Management knows that they need to add more engineers if the company is going to be able to handle these kinds of projects internally. But because they must begin

BOILED THROUGH OUTSOURCING
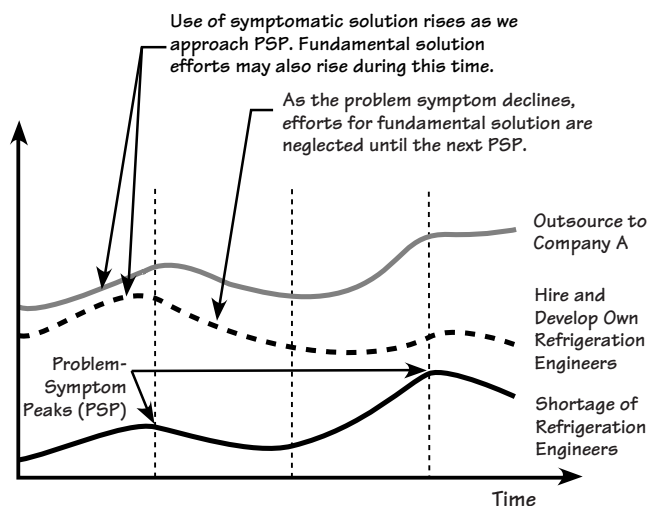
The manufacturer faces a shortage of engineers to work on a new product design (problem symptom). Management knows that they need to add more engineers (fundamental solution). But because they must begin work on the new product right away, they choose to outsource the engineering (symptomatic solution). The company repeats the same dynamics time and again, lessening its own internal capability.

work on the new product right away, they choose to outsource the engineering to Company A.

At the same time, because company leaders recognize the need to build up their own staff, they initiate actions to hire and develop internal capacity. However, those efforts quickly wane when the problem symptom—the need for experienced refrigeration engineers—declines because Company A is doing such a good job. Other, more pressing issues occupy the company's attention, and the capacity-building effort gets put on the back burner until the next staffing shortfall occurs. At that point, the company repeats the same dynamics.

This approach leads to a pattern of behavior in which the problem symptom continually resurfaces. Each time, the company makes efforts to address both the symptomatic and fundamental problems. However, when the quick fix proves successful in handling the problem in the short term, the organization continues to rely on that tactic over the longer run. As a result, efforts to seek a lasting, more fundamental solution decline. If left unchecked, the company will eventually "boil" like the poor frog—that is, face serious financial and performance difficulties.

BREAKING OUT OF A REACTIVE ORIENTATION

Although choosing to invest in the more fundamental solution is better than pursuing a symptomatic solution, both actions are inherently reactionary. This is because the two approaches are driven by the need to solve what is currently wrong rather than by the desire to create the future you want. Hence, even opting for the fundamental solution can produce problem symptoms that come and go. This is because no matter how the symptom gets reduced, the amount of effort devoted to its solution varies with the severity of the

symptom—it rises when the problem is acute and falls when it is "under control."

Breaking out of this reactive orientation requires a shift from problem-solving to developing a vision of what you want to create—a generative orientation (see "From a Reactive to a Generative Orientation"). In our example of the refrigerator manufacturer, this approach would mean having a clarity of vision about the kind of engineering capability the company wants to maintain and then developing that skill base—regardless of whether the organization is experiencing shortfalls at the moment or not. The company may still experience problems with staffing shortages during this time. However, when it encounters them, the organization will be able to use symptomatic solutions as temporary stopgap measures, while it continues to steadily build its underlying capacity.

Does our refrigerator manufacturing example mean that all outsourcing is a case of "Shifting the Burden"? The answer to that question depends on your organization's vision of what it wants to keep as its core competencies. If you inadvertently ended up outsourcing what you considered a core competency, such as refrigeration design, then you would be caught in a "Shifting the Burden" dynamic. On the other hand, if you decided that competence in payroll systems and health-benefit programs

was not key to your core business, outsourcing those functions might be a prudent decision.

OUT OF THE BOILING POT AND . . .

The "Shifting the Burden" structure shows that, in addition to refining our organizations' mechanisms for detecting slow, gradual changes, we need to develop better direction-setting systems. Otherwise, we may improve at making course corrections but never clarify what course we really want to take. That approach would be analogous to our poor frog jumping from one pot to another whenever it feels the water heating up, but never pursuing a more fundamental solution by seeking a nice lily pond instead. Even with improved temperature-sensing mechanisms, if the frog keeps hopping from one pot to the next, the odds are that, sooner or later, it will end up on someone's dinner plate. ◘
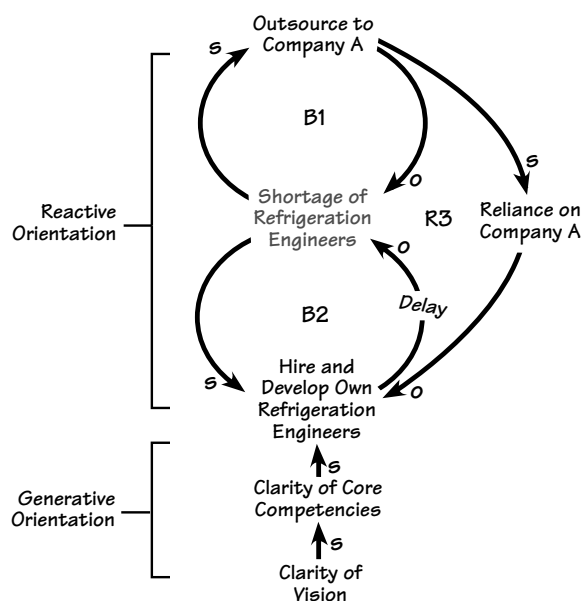
Breaking out of a reactive orientation requires shifting from problem-solving to developing a vision of what you want to create—a generative orientation.

"ESCALATION": THE UNDERLYING STRUCTURE OF WAR

The lessons we learn from the "Fixes That Fail" and "Shifting the Burden" archetypes center on the kinds of actions that we choose to take and the long-term consequences of those actions. In "Escalation," the situation becomes more complex, because our actions directly affect the actions that *others* take. But unlike what we learned in physics—where every action produces an equal and opposite reaction—our actions are amplified with each round, leading to a phenomenon known as escalation. If left unchecked, the escalation dynamic can spiral out of control, going far beyond what either side may have intended.

EYE OF THE BEHOLDER

In the U.S., the expression "keeping up with the Joneses" describes the rivalry that some homeowners fall into with their neighbors. So, if the Joneses buy a new car, the Smiths feel compelled to replace their old vehicle with the latest model. When the Joneses have their yard landscaped, the Smiths do the same. And on it goes.

In this case, escalation occurs when we equate acquiring material things with success. Once we become involved in a competition—whether it's over which neighbor has a neater lawn or which airline is offering the lowest fares—we unconsciously raise the ante with each additional action that we take. For example, even though the Smiths believe they are merely "keeping up" when they buy their new car, they may choose one with bells and whistles that the Joneses' car doesn't have, triggering another round of escalating conspicuous consumption.

Because parity is in the eye of the beholder, escalation dynamics can erupt in any relationship that involves even the slightest hint of rivalry. On the playground, we have all seen how an accidental bump quickly escalates into a shoving match and then into an all-out fight. On a larger scale, we have lived through perhaps the largest escalation dynamic in human history—the nuclear arms race between the Soviet Union and the United States.

INNOCUOUS BEGINNINGS, DESTRUCTIVE ENDINGS

Why does escalation often spin out of control? One reason is that delays contribute to distortions in the information flowing between the two sides. One delay occurs between the actions that each party takes and the results of those actions. The other is between the relative position of each participant vis-à-vis the other and the perceived threat that this positioning causes the parties to feel. Information gets distorted along every link in the system; however, the second delay may have the greatest effect, in that it leads each side to overestimate the impact of its rivals' activities on their relative position.

Thus, when BigCo increases its level of promotions, the results of these activities do not show up immediately in higher sales (B1 in "The Structure of Escalation Dynamics"). Thus, BigCo may engage in more promotions than it originally intended; for instance, by prolonging a special offer. This delay contributes to the escalation dynamics, because LilCo then perceives BigCo as aggressively promoting its products. In the short-term, LilCo may respond by engaging in "Tit-for-Tat" behavior (see "Three Regions of Escalation").

Eventually, the results of BigCo's actions do become visible. But, because of the delay between relative results and feelings of being threatened, LilCo remains complacent about its level of activity relative to BigCo. When LilCo finally realizes that it has fallen behind, the gap between BigCo's

THE STRUCTURE OF ESCALATION DYNAMICS
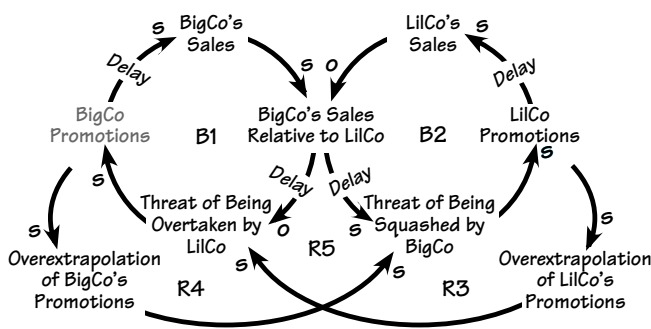
When BigCo increases its promotions (B1), it takes time for the effects to show up. Thus, LilCo does not feel threatened initially. When LilCo finally realizes that it has fallen behind, it tries to catch up (B2). BigCo feels threatened by these aggressive actions and overextrapolates the impact of LilCo's promotions (R3). In turn, LilCo sees BigCo's efforts as an even worse threat and again increases its own promotional activity (R4).

sales and LilCo's sales is wider than it might have been if LilCo had seen the relative impact of BigCo's actions sooner. When LilCo takes action, it does so from a heightened state of threat and tries to catch up to BigCo as fast as it can (B2). BigCo then interprets this increased level of activity as an attempt by LilCo to raise the stakes. So, BigCo now overextrapolates LilCo's catch-up activity as a threat to its own position and, in turn, increases its activities (R3). LilCo sees BigCo's increased marketing efforts as an even worse threat and again increases its own promotional activity (R4). Both sides are fast approaching the turbo-charged region of All-Out Escalation.

In All-Out Escalation, time delays become compressed. Because the parties have previously been caught off-guard as a result of delays, neither side wants to wait for additional results to materialize before taking action. The problem is that those subsequent actions are based on each party's extrapolations—usually inflated—of the other's activities (R5). When escalation reaches this level, activity by one party begets more activity by the other with ever-increasing speed and volatility until something devastating happens.

In the case of BigCo and LilCo, it may appear that there is nothing wrong because sales continue to rise for both companies. However, promotional costs are rising faster than sales, so margins are shrinking even while sales are growing. Companies have engaged in these kinds of dynamics to the point where they sell their products at a loss because they are so focused on not being "outsold" by their competition!

EARLY WARNING SYSTEMS

Escalation dynamics can occur in numerous business settings, such as price-cutting wars, promotional competitions, and product-feature battles. So, how can you keep from getting lured into these dynamics in the first place? As with most conflicts, the

best time to deal with escalation is early in the process, before the dynamics take on a life of their own. For the "Escalation" archetype, this means paying attention to the interplay between you and your rival while you are still in the relatively harmless Tit-for-Tat stage. Take the first rumblings of an escalation dynamic as your early warning to proceed with caution. Immediately assess the value proposition that you are offering your customers. For instance, when a competitor begins to target your customers by emphasizing a lower price, it is easy to respond by lowering your price as well. But perhaps your competitor picked price as the variable because that is the only thing that they can compete on.

The problem with responding in kind to this gambit is that you let your competitor set the ground rules. This proved a costly mistake for Texas Instruments, when it allowed Commodore to choose price as the competing variable for the home computer. Instead of emphasizing the superiority of its product, Texas Instruments lowered its prices. Price cuts followed price cuts until TI finally had to write off the TI 99/4A computer, which cost the company hundreds of millions of dollars.

Instead of letting the competition dictate your strategy, refocus your business strategy. When competition in the overnight delivery business stiffened and some companies began to lower their prices, FedEx could have joined in. However, it chose to emphasize a value proposition that was

even more important to customers than price: reliability. The company thus reestablished its leadership role in the overnight delivery business and was able to maintain higher pricing than its competitors.

ENDING THE WAR GAMES

In the hit movie *War Games,* a Defense Department computer assumed control of all U.S. nuclear warheads. As the computer was in the process of cracking the security code that would allow it to launch the entire U.S. nuclear arsenal at the U.S.S.R., the programmer-hero engaged it in playing tic-tac-toe over and over again, hoping it would learn that trying to win the game was futile. In the end, the computer did learn that lesson and concluded that all thermonuclear war scenarios would lead to a no-win situation. Even though *War Games* was fictional, it accurately captured the potentially destructive quality of escalation. More individuals, companies, and countries embroiled in escalating struggles could learn a valuable lesson from understanding the pitfalls of this structure. ◼

THREE REGIONS OF ESCALATION
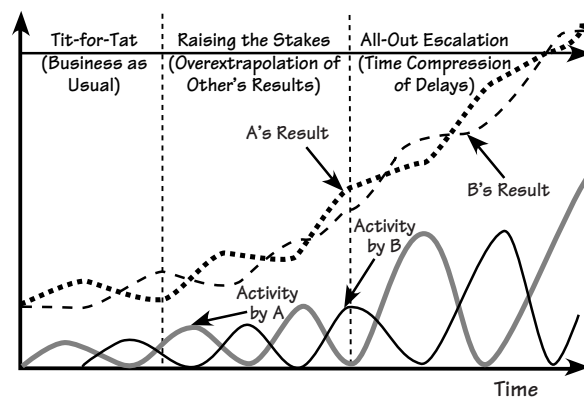
In the early stage of "Escalation," the parties engage in "Tit-for-Tat" behavior. As each company overextrapolates the other's activities, they "Raise the Stakes" with each action. When the rivals reach "All-Out Escalation," they act with ever-increasing speed and volatility until something devastating happens.

"DRIFTING GOALS": THE CHALLENGE OF CONFLICTING PRIORITIES

It's 7:30 A.M., and you are hurriedly getting your children ready for the day. You finally buckle the kids into the car, rush across town, and drop them off at school—only to find yourself stuck in bumper-to-bumper traffic on the way to the office. You glance at your watch. It is 8:03. You want to be early for your first meeting at your new job, but everything seems to be conspiring against you. Finally, the traffic clears as you pass the site of the accident that caused the logjam. You glance at your watch again as you pull into the parking lot at work—it's now 8:52. "So much for getting a cup of coffee before the meeting," you mutter. You walk into the conference room, a little breathless but on time at 8:58, only to find that you are the first one there. You check your calendar to make sure that you have the right date, time, and place. Yep, you do. Around 9:05, some of your coworkers show up, and by 9:10, everyone has arrived for the "9 o'clock" meeting. So, what do you learn from this experience? Probably the same thing the others have already learned—that the "real" starting time for meetings is never the stated time. This is a common example of the "Drifting Goals" archetype.

THE SLIPPERY SLOPE

Many of us have had experiences similar to the one described above. As a group, we commit to a certain meeting time or project deadline with every intention of fulfilling that promise. Nevertheless, "life" intervenes—in the form of traffic jams, more pressing deadlines, and urgent phone calls—so we relax our standards for keeping the commitment. We think to ourselves, "The rest of the group is bound to be late, so I'll spend one more minute polishing this presentation" or "Waiting an extra day for the new release won't kill our customers!" We say 9:00 A.M., but, through our own tardiness or lack of reaction when others arrive late, we tacitly accept that it's O.K. to begin the meeting no later than 9:05. Well, maybe 9:10, but we absolutely should start by 9:15. So, why don't we just schedule the meeting for 9:15? Because then it's likely to start at 9:30! This dynamic reminds us of the old adage "Give him an inch, and he'll take a mile." It seems that once we compromise a little, we are headed down a slippery slope with no bottom in sight.

One obvious solution to drifting meeting times would be to establish a company-wide norm that meetings must begin as scheduled no matter what. Many groups have experimented with different incentives (or more accurately, disincentives) to encourage people to arrive on time—ranging from monetary penalties to singing a song for being late—with mixed results. For numerous organizations, though, delayed meetings are just a surface manifestation of a larger—and potentially more serious—pattern of drifting goals.

THE DANGER IN DEADLINES

Perhaps with things like meetings, it's not such a big deal if everyone translates 9:00 to mean 9:15. The problem with such habits is that they have a way of spreading to other areas, such as quality standards, new product launches, and marketing campaigns. The danger lies in the tendency for all goals to drift, depending on the forces that are operating at the moment. In other words, we want a quality of 10, but when time is tight, we will settle for 9.5. If we are even more pressed, 9.3 will do. And on it goes.

CLASHING GOALS
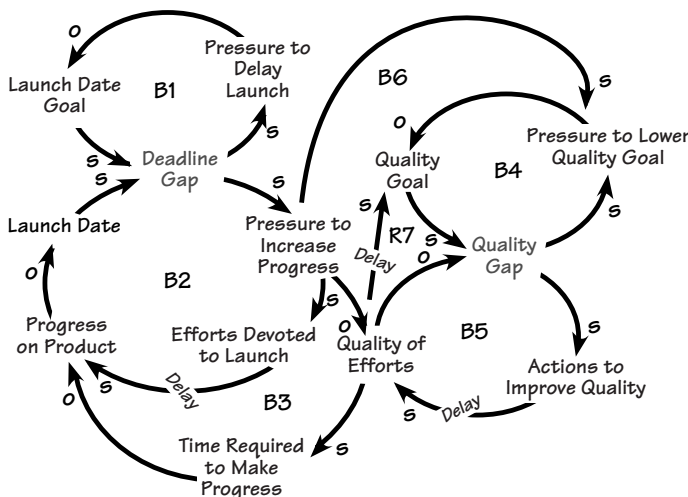
Quality and schedule sometimes conflict (B1 and B2 conflict with B4 and B5). The pressure to expedite a project increases the pressure to cut corners (B3) and lower the quality goal (B6). Over time, as the quality of our efforts declines, the quality goal declines, which reduces the gap. This leads to a further decrease in the quality of our efforts and a subsequent lowering of product quality (R7).

Some standards are more important to maintain than others. For example, new product launches generally need to stay on schedule so the company can fulfill advanced sales. But more often than not, deadlines begin to slip, often because people are juggling multiple demands. When this happens, the project manager has at least two choices about how to address the gap between the desired and actual deadline (see "Clashing Goals"). One way is to simply delay the launch date (B1), which is not an acceptable alternative in most cases. Another way is to increase the amount of effort or resources devoted to the project so that progress can be made faster and the launch date can be met (B2). If management makes it clear that the deadline must be maintained at all costs, then this second scenario will likely occur. But if the organization doesn't allocate the resources needed to expedite the project, people in the system must find other ways to reach the goal. One solution is to reduce the quality of efforts on the project; that is, to cut corners, which will lower the time required to produce the end product (B3).

In some cases, taking such a shortcut makes sense in order to get a critical product out on time, even though the quality may not be up to our usual standards. The problem with this approach is that it rarely remains an isolated event, but rather becomes a part of the way we do things. The next time we get into a time bind, we may "cheat" a little on quality again because it worked the last time. So by setting rigid deadlines in isolation of other factors, we can actually create undesirable long-term outcomes, such as lower-quality products.

COMPETING GOALS

The "Drifting Goals" phenomenon occurs more often when we are juggling competing objectives than when we are trying to meet a single target. Ideally, we would like to produce a high-quality product on schedule every time, but what happens when these two requirements seem to conflict (when B1 and B2 conflict with B4 and B5)? In "Clashing Goals," we see that the pressure to expedite a project does two things. It increases the pressure to lower the quality goal (B6) and it lessens the quality of the efforts that we can put forth.

Over time, this decline in quality of efforts also erodes the quality goal itself, which creates a dangerous reinforcing dynamic. Specifically, as the quality of our efforts declines, the quality goal declines, which reduces the gap. This leads to a further decrease in the quality of our efforts and a subsequent lowering of product quality (R7).

The figure "Drifting Goals over Time" shows the long-term dynamics of this structure at work. The quality goal appears to stay stable for periods of time and drops slowly relative to the wider swings of the actual quality of efforts. This dynamic serves to mask the long-term downward trend, which is why this archetype is often referred to as the "Boiled Frog Syndrome." The changes in the goal are slow enough that nobody detects the dangerous trend until the company is in serious "hot water."

IDENTIFYING INTER-DEPENDENT GOALS

An important lesson in managing the "Drifting Goals" structure is to look beyond the individual goals and identify interdependent goals. By mapping the interrelationships, you can more intentionally decide which goal you are going to emphasize this time,
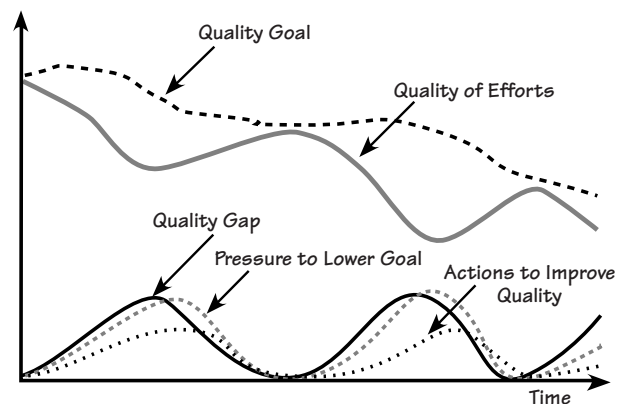
and you can put mechanisms in place to prevent you from plummeting down the slippery slope of drifting goals. This action alone won't necessarily stop each goal from drifting, but it will help you to become more aware of the consequences of your actions.

Returning to our original example, people in organizations constantly juggle the competing goals of getting to meetings on time and attending to a whole slew of tasks they need to accomplish. One leverage point would be to emphasize the importance of actually starting as scheduled and to ask what it would take for everyone to keep that commitment. We may discover that 9 A.M. is not the best time to accomplish this goal because there are too many other competing variables—traffic, urgent messages to return, and problems to troubleshoot. It may be that gathering at lunchtime will make the goal more achievable—especially if lunch is provided! Lunch or no lunch, the principle is to establish the importance of meeting a specific goal in the context of multiple goals, and then to set up structures to minimize the conflicts between competing demands and priorities. ◘

The quality goal appears to stay stable for periods of time and drops slowly relative to the wider swings of the actual quality of efforts. This dynamic serves to mask the long-term downward trend.

INDEX TO *THE SYSTEMS THINKER*

ABOUT THE TOOLBOX REPRINT SERIES

Systems Archetypes III: Understanding Patterns of Behavior and Delay is the newest volume in the Toolbox Reprint Series. Other volumes include *Systems Archetypes I: Diagnosing Systemic Issues and Designing High-Leverage Interventions, Systems Archetypes II: Using Systems Archetypes to Take Effective Action, Systems Thinking Tools: A User's Reference Guide,* and *The "Thinking" in Systems Thinking: Seven Essential Skills.* All volumes are available for $16.95 each. As these booklets are often used in training and introductory courses, volume discounts are available. Call 1-800-272-0945 for details.

The Toolbox Reprint Series has been compiled from *The Systems Thinker*® Newsletter, which presents a systems perspective on current issues and provides systems tools for framing problems in new and insightful ways. *The Systems Thinker* includes articles by leading systems thinkers, case studies of systems thinking implementation, software and book reviews, a calendar of workshops and events, and numerous other columns geared to different levels of systems thinking ability. To learn more about *The Systems Thinker* or to subscribe, go to http://www.thesystemsthinker.com.

PEGASUS PUBLICATIONS

The Toolbox Reprint Series
Systems Archetypes I: Diagnosing Systemic Issues and Designing High-Leverage Interventions
Systems Archetypes II: Using Systems Archetypes to Take Effective Action
Systems Archetypes III: Understanding Patterns of Behavior and Delay
Systems Thinking Tools: A User's Reference Guide
The "Thinking" in Systems Thinking: Seven Essential Skills

The Pegasus Workbook Series
Systems Archetype Basics: From Story to Structure   *Systems Thinking Basics: From Concepts to Causal Loops*
The Learner's Path: Practices for Recovering Knowers

Learning Fables (available as soft-cover books or as e-books)
Outlearning the Wolves: Surviving and Thriving in a Learning Organization
Shadows of the Neanderthal: Illuminating the Beliefs That Limit Our Organizations
The Lemming Dilemma: Living with Purpose, Leading with Vision
The Tip of the Iceberg: Managing the Hidden Forces That Can Make or Break Your Organization
Listening to the Volcano: Conversations That Open Our Minds to New Possibilities

Newsletters (available electronically)
THE SYSTEMS THINKER®   *LEVERAGE POINTS*® *for a New Workplace, New World*

The Innovations in Management Series
Concise, practical volumes on systems thinking and organizational learning tools, principles, and applications.

The "Billibonk" Series
Billibonk & the Thorn Patch   *Frankl's "Thorn Patch" Fieldbook*
Billibonk & the Big Itch   *Frankl's "Big Itch" Fieldbook*

Other Titles
When a Butterfly Sneezes: A Guide for Helping Kids Explore Interconnections in Our World Through Favorite Stories

Anthologies
Managing the Rapids: Stories from the Forefront of the Learning Organization
Reflections on Creating Learning Organizations
The New Workplace: Transforming the Character and Culture of Our Organizations
Organizational Learning at Work: Embracing the Challenges of the New Workplace
Making It Happen: Stories from Inside the New Workplace
Essential Readings for the Innovative Organization (four titles, available individually in print as well as in PDF format)

Pegasus Communications, Inc. is dedicated to providing resources that help people explore, understand, articulate, and address the challenges they face in managing the complexities of a changing world. Since 1989, Pegasus has worked to build a community of practitioners through newsletters, books, audio and video tapes, and its annual *Systems Thinking in Action*® Conference and other events.

For more information, contact:
Pegasus Communications, Inc. • One Moody Street • Waltham, MA 02453-5339 USA • www.pegasuscom.com
Phone: 800-272-0945 / 781-398-9700 • Fax: 781-894-7175