INTRODUCTION

Lean Thinking versus *Muda*

Muda. It's the one word of Japanese you really must know. It sounds awful as it rolls off your tongue and it should, because *muda* means "waste," specifically any human activity which absorbs resources but creates no *value:* mistakes which require rectification, production of items no one wants so that inventories and remaindered goods pile up, processing steps which aren't actually needed, movement of employees and transport of goods from one place to another without any purpose, groups of people in a downstream activity standing around waiting because an upstream activity has not delivered on time, and goods and services which don't meet the needs of the customer.

Taiichi Ohno (1912–1990), the Toyota executive who was the most ferocious foe of waste human history has produced, identified the first seven types of *muda* described above and we've added the final one.[1] Perhaps there are even more. But however many varieties of *muda* there may be, it's hard to dispute—from even the most casual observation of what gets done in an average day in the average organization—that *muda* is everywhere. What's more, as you learn to see *muda* in the pages ahead, you will discover that there is even more around than you ever dreamed.

Fortunately, there is a powerful antidote to *muda: lean thinking*. It provides a way to specify value, line up value-creating actions in the best sequence, conduct these activities without interruption whenever someone requests them, and perform them more and more effectively. In short, lean thinking is *lean* because it provides a way to do more and more with less and less—less human effort, less equipment, less time, and less space—while coming closer and closer to providing customers with exactly what they want.

Lean thinking also provides a way to make work more satisfying by providing immediate feedback on efforts to convert *muda* into value. And, in striking contrast with the recent craze for process reengineering, it provides a way to create new work rather than simply destroying jobs in the name of efficiency.

Specify Value

The critical starting point for lean thinking is *value*. Value can only be defined by the ultimate customer. And it's only meaningful when expressed in terms of a specific product (a good or a service, and often both at once) which meets the customer's needs at a specific price at a specific time.

Value is created by the producer. From the customer's standpoint, this is why producers exist. Yet for a host of reasons value is very hard for producers to accurately define. Business school–trained senior executives of American firms routinely greet us when we visit with a slick presentation about their organization, their technology, their core competencies, and their strategic intentions. Then, over lunch, they tell us about their short-term competitive problems (specifically their need to garner adequate profits in the next quarter) and the consequent cost-cutting initiatives. These often involve clever ways to eliminate jobs, divert revenues from their downstream customers, and extract profits from their upstream suppliers. (Because we are associated with the concept of lean production, they are usually eager to label these programs "lean," although often they are only "mean.") By dessert, we may be hearing about their personal career issues in the current age of "downsizing."

What only comes up when we push it to the foreground is the specific products the firm expects specific customers to purchase at a specific price to keep the company in business and how the performance and delivered quality of these products can be improved while their fundamental costs are pushed steadily down. In raising this issue it's often revealing to ask these executives a simple question: Can you put yourself in the position of a design as it progresses from concept to launch, an order as information flows from initial request to delivered product, and the physical product as it progresses from raw material to the customer, and describe what will happen to you at each step along the way? Usually there is an awkward silence, and then, if we aren't persistent, these issues quickly slip out of sight to be replaced once more by aggregated financial considerations. In short, the immediate needs of the shareholder and the financial mind-set of the senior managers have taken precedence over the day-to-day realities of specifying and creating value for the customer.

When we've gone to Germany, until very recently, we've found a reverse distortion of value specification. For much of the post–World War II era, executives of private or bank-controlled companies could ignore the need for short-term financial performance and were eager to tell us all about their products and process technologies. Even the most senior executives could go into great detail about product features and new processing methods

But who specified their value? The engineers running the companies! Designs with more complexity produced with ever more complex machinery were asserted to be just what the customer wanted and just what the production process needed. But where was the evidence?

In pressing this point, it often became apparent that the strong technical functions and highly trained technical experts leading German firms obtained their sense of worth—their conviction that they were doing a first-rate job—by pushing ahead with refinements and complexities that were of little interest to anyone but the experts themselves. Our doubts about proposed products were often countered with claims that "the customer will want it once we explain it," while recent product failures were often explained away as instances where "the customers weren't sophisticated enough to grasp the merits of the product."

A central feature of the crisis of German industry in the period since the end of the cold war has been the dawning perception that the complex, customized designs and sophisticated processing technologies favored by German engineers are too expensive for customers to afford and often irrelevant to their real desires.

When we have traveled to Japan, also until very recently, we have encountered yet a third distortion. What's been really important for Japanese firms as they have defined value is *where* value is created. Most executives, even at firms like Toyota which pioneered lean thinking, have begun their value definition process by asking how they can design and make their product at home—to satisfy societal expectations about long-term employment and stable supplier relations. Yet most customers across the world like products designed with an eye to local needs, which is hard to do from a distant home office. And they like products made to their precise order to be delivered immediately, which ocean shipping from a Japanese production base makes impossible. They certainly do not define the value of a product primarily in terms of where it was designed or made.

What's more, the stay-at-home-at-all-costs thinking of Japanese senior managers, even as the yen steadily strengthened, depleted the financial resources these firms needed to do new things in the future. The immediate needs of employees and suppliers took precedence over the needs of the customer, which must sustain any firm in the long term.

Moving beyond these national distortions in the world's three most important industrial systems (and every country probably has its own unique set),[2] we are repeatedly struck how the definition of value is skewed everywhere by the power of preexisting organizations, technologies, and undepreciated assets, along with outdated thinking about economies of scale. Managers around the world tend to say, "This product is what we know how to produce using assets we've already bought, so if customers don't be doing instead is fundamentally rethinking value from the perspective of the customer.

One of the best (and most exasperating) illustrations of this backwards thought-process is the current-day airline industry. As frequent users of this service we have long been keeping detailed notes on our experiences and contrasting our own definition of value with that proposed by most companies in this industry. Our value equation is very simple: to get from where we are to where we want to be safely with the least hassle at a reasonable price. By contrast, the airline's definition seems to involve using their existing assets in the most "efficient" manner, even if we have to visit Timbuktu to get anywhere. They then throw in added features—like executive lounges in their hubs and elaborate entertainment systems in every seat—in hopes the inconvenience will be tolerable.

Just today, as this is written, one of us has traveled the 350 miles from his summer home in Jamestown in western New York State, across Lake Erie, to Holland, Michigan, in order to make a presentation on lean thinking to an industrial audience. What was needed was a way to fly from Jamestown directly to Holland (both of which have small airports) at an affordable cost. What was available was either an absurdly priced charter service from Jamestown to Holland (total door-to-door travel time of about two hours) or an eighty-mile drive to the Buffalo, New York, airport, a flight on a large jet to the Detroit sortation center of Northwest Airlines (where the self-sorting human cargo finds its way through a massive terminal from one plane to the next), another flight on a large jet to Grand Rapids, Michigan, and a forty-mile drive to the ultimate destination. (The lower-cost option required a total travel time of seven hours.)

Why aren't airlines like Northwest (and its global partner KLM) and airframe builders like Boeing and Airbus working on low-cost, point-to-point services using smaller jets instead of developing ever-larger aircraft? And why aren't they developing quick turnaround systems for small jets at small airports instead of constructing Taj Mahal terminals at the absurd "hubs" created in America after airline deregulation—and long present in Europe and East Asia due to the politically motivated practice of routing most flights of state-controlled airlines through national capitals? (One hour of the seven hours spent on the trip just cited was taxiing time in the Detroit hub and a second was occupied with self-sortation inside the terminal.)

Few firms are aggressively promoting this definition of value because the airlines and airframe builders start their thinking with extraordinarily costly assets in the form of large aircraft; the engineering knowledge, tooling, and production facilities to make more large aircraft; and massive airport complexes. Old-fashioned "efficiency" thinking suggests that the best way to make use of these assets and technologies is to get larger batches of people on larger planes and to do this by sending ever more passengers through the expensive sorting centers. This type of efficiency calculation, focused on the airplane and the hub—only two of the many elements in the total trip—loses sight of the whole. Much worse from the standpoint of value for the passenger, it simply misses the point.

The end result of fifteen years of this type of thinking in the United States is that passengers are miserable (this is not what they meant by value!), the aircraft producers make little money (because the airlines can't afford new planes), and the airlines (excepting Southwest and a few other start-ups pursuing the more sensible strategy of flying point-to-point, although still using large aircraft) have flown a decade-long holding pattern in the vicinity of bankruptcy. Europe and parts of East Asia are not far behind.

Lean thinking therefore must start with a conscious attempt to precisely define value in terms of specific products with specific capabilities offered at specific prices through a dialogue with specific customers. The way to do this is to ignore existing assets and technologies and to rethink firms on a product-line basis with strong, dedicated product teams. This also requires redefining the role for a firm's technical experts (like the inward-looking German engineers we just cited) and rethinking just where in the world to create value. Realistically, no manager can actually implement all of these changes instantly, but it's essential to form a clear view of what's really needed. Otherwise the definition of value is almost certain to be skewed.

In summary, specifying value accurately is the critical first step in lean thinking. Providing the wrong good or service the right way is *muda*.

Identify the Value Stream

The *value stream* is the set of all the specific actions required to bring a specific product (whether a good, a service, or, increasingly, a combination of the two) through the three critical management tasks of any business: the *problem-solving task* running from concept through detailed design and engineering to production launch, the *information management task* running from order-taking through detailed scheduling to delivery, and the *physical transformation task* proceeding from raw materials to a finished product in the hands of the customer.[3] Identifying the *entire* value stream for each product (or in some cases for each product family) is the next step in lean thinking, a step which firms have rarely attempted but which almost always exposes enormous, indeed staggering, amounts of *muda*.

Specifically, value stream analysis will almost always show that three types of actions are occurring along the value stream: (1) Many steps will be found to unambiguously create value: welding the tubes of a bicycle frame together or flying a passenger from Dayton to Des Moines. (2) Many other steps will be found to create no value but to be unavoidable with current technologies and production assets: inspecting welds to ensure quality and the extra step of flying large planes through the Detroit hub en route from Dayton to Des Moines (we'll term these Type One *muda*). And (3) many additional steps will be found to create no value and to be immediately avoidable (Type Two *muda*).

For example, when Pratt & Whitney, the world's largest manufacturer of aircraft jet engines, recently started to map its value streams for its three families of jet engines, it discovered that activities undertaken by its raw materials suppliers to produce ultrapure metals were duplicated at great cost by the next firms downstream, the forgers who converted metal ingots into near-net shapes suitable for machining. At the same time, the initial ingot of material—for example, titanium or nickel—was ten times the weight of the machined parts eventually fashioned from it. Ninety percent of the very expensive metals were being scrapped because the initial ingot was poured in a massive size—the melters were certain that this was efficient—without much attention to the shape of the finished parts. And finally, the melters were preparing several different ingots—at great cost—in order to meet Pratt's precise technical requirements for each engine, which varied only marginally from those of other engine families and from the needs of competitors. Many of these activities could be eliminated almost immediately with dramatic cost savings.

How could so much waste go unnoticed for decades in the supposedly sophisticated aerospace industry? Very simply: None of the four firms involved in this tributary value stream for a jet engine—the melter, the forger, the machiner, and the final assembler—had ever fully explained its activities to the other three. Partly, this was a matter of confidentiality—each firm feared that those upstream and downstream would use any information revealed to drive a harder bargain. And partly, it was a matter of obliviousness. The four firms were accustomed to looking carefully at their own affairs but had simply never taken the time to look at the whole value stream, including the consequences of their internal activities for other firms along the stream. When they did, within the past year, they discovered massive waste.

So lean thinking must go beyond the firm, the standard unit of scorekeeping in businesses across the world, to look at the whole: the entire set of activities entailed in creating and producing a specific product, from concept through detailed design to actual availability, from the initial sale through order entry and production scheduling to delivery, and from raw materials produced far away and out of sight right into the hands of the customer. The organizational mechanism for doing this is what we call the *lean enterprise*, a continuing conference of all the concerned parties to create a channel for the entire value stream, dredging away all the *muda*.

Whenever we present this idea for the first time, audiences tend to assume that a new legal entity is needed, some formalized successor to the "virtual corporation" which in reality becomes a new form of vertical integration. In fact, what is needed is the exact opposite. In an age when individual firms are outsourcing more and themselves doing less, the actual need is for a voluntary alliance of all the interested parties to oversee the disintegrated value stream, an alliance which examines every value-creating step and lasts as long as the product lasts. For products like automobiles in a specific size class, which go through successive generations of development, this might be decades; for short-lived products like software for a specific application, it might be less than a year.

Creating lean enterprises *does* require a new way to think about firm-to-firm relations, some simple principles for regulating behavior between firms, and *transparency* regarding all the steps taken along the value stream so each participant can verify that the other firms are behaving in accord with the agreed principles. These issues are the subject of Part III of this book.

Flow

Once value has been precisely specified, the value stream for a specific product fully mapped by the lean enterprise, and obviously wasteful steps eliminated, it's time for the next step in lean thinking—a truly breathtaking one: Make the remaining, value-creating steps *flow*. However, please be warned that this step requires a complete rearrangement of your mental furniture.

We are all born into a mental world of "functions" and "departments," a commonsense conviction that activities ought to be grouped by type so they can be performed more efficiently and managed more easily. In addition, to get tasks done efficiently within departments, it seems like further common sense to perform like activities in batches: "In the Claims Department, process all of the Claim As, then the Claim Bs, and then the Claim Cs. In the Paint Department, paint all of the green parts, then shift over and paint all the red parts, then do the purple ones." Batches, as it turns out, always mean long waits as the product sits patiently awaiting the department's changeover to the type of activity the product needs next. But this approach keeps the members of the department busy, all the equipment running hard, and justifies dedicated, high-speed equipment. So, it must be "efficient," right? Actually, it's dead wrong, but hard or impossible for most of us to see.

Recently, one of us performed a simple experiment with his daughters, ages six and nine: They were asked the best way to fold, address, seal, stamp, and mail the monthly issue of their mother's newsletter. After a bit of thought their answer was emphatic: "Daddy, first, you should fold all of the newsletters. Then you should put on all the address labels. Then you should attach the seal to stick the upper and lower parts together [to secure the newsletter for mailing]. Then you should put on the stamps." "But why not fold one newsletter, then seal it, then attach the address label, and then put on the stamp? Wouldn't that avoid the wasted effort of picking up and putting down every newsletter four times? Why don't we look at the problem from the standpoint of the newsletter which wants to get mailed in the quickest way with the least effort?" Their emphatic answer: "Because that wouldn't be efficient!"

What was striking was their profound conviction that performing tasks in batches is best—sending the newsletters from "department" to "department" around the kitchen table—and their failure to consider that a rethink of the task might permit continuous flow and more efficient work. What's equally striking when looked at this way is that most of the world conducts its affairs in accord with the thought processes of six- and nine-year-olds!

Taiichi Ohno blamed this batch-and-queue mode of thinking on civilization's first farmers, who he claimed lost the one-thing-at-a-time wisdom of the hunter as they became obsessed with batches (the once-a-year harvest) and inventories (the grain depository).[4] Or perhaps we're simply born with batching thinking in our heads, along with many other "common sense" illusions—for example, that time is constant rather than relative or that space is straight rather than curved. But we all need to fight departmentalized, batch thinking because tasks can almost always be accomplished much more efficiently and accurately when the product is worked on continuously from raw material to finished good. In short, things work better when you focus on the product and its needs, rather than the organization or the equipment, so that all the activities needed to design, order, and provide a product occur in continuous flow.

Henry Ford and his associates were the first people to fully realize the potential of flow. Ford reduced the amount of effort required to assemble a Model T Ford by 90 percent during the fall of 1913 by switching to continuous flow in final assembly. Subsequently, he lined up all the machines needed to produce the parts for the Model T in the correct sequence and tried to achieve flow all the way from raw materials to shipment of the finished car, achieving a similar productivity leap. But he only discovered the *special case.* His method only worked when production volumes were high enough to justify high-speed assembly lines, when every product used exactly the same parts, and when the same model was produced for many years (nineteen in the case of the Model T). In the early 1920s, when Ford towered above the rest of the industrial world, his company was assembling more than two million Model Ts at dozens of assembly plants around the world, every one of them exactly alike.

After World War II, Taiichi Ohno and his technical collaborators, including Shigeo Shingo,[5] concluded that the real challenge was to create continuous flow in small-lot production when dozens or hundreds of copies of a product were needed, not millions. This is the *general case* because these humble streams, not the few mighty rivers, account for the great bulk of human needs. Ohno and his associates achieved continuous flow in low-volume production, in most cases without assembly lines, by learning to quickly change over tools from one product to the next and by "right-sizing" (miniaturizing) machines so that processing steps of different types (say, molding, painting, and assembly) could be conducted immediately adjacent to each other with the object undergoing manufacture being kept in continuous flow.

The benefits of doing things this way are easy to demonstrate. We've recently watched with our own eyes, in plants in North America and Europe, as lean thinkers practiced *kaikaku* (roughly translatable as "radical improvement," in contrast with *kaizen*, or "continuous incremental improvement"). Production activities for a specific product were rearranged in a day from departments and batches to continuous flow, with a doubling of productivity and a dramatic reduction in errors and scrap. We'll report later in this book on the revolutionary rearrangement of product development and order-scheduling activities for these same products to produce the same magnitude of effect in only a slightly longer adjustment period. Yet the great bulk of activities across the world are still conducted in departmentalized, batch-and-queue fashion fifty years after a dramatically superior way was discovered. Why?

The most basic problem is that flow thinking is counterintuitive; it seems obvious to most people that work should be organized by departments in batches. Then, once departments and specialized equipment for making batches at high speeds are put in place, both the career aspirations of employees within departments and the calculations of the corporate accountant (who wants to keep expensive assets fully utilized) work powerfully against switching over to flow.

The reengineering movement has recognized that departmentalized thinking is suboptimal and has tried to shift the focus from organizational categories (departments) to value-creating "processes"—credit checking or claims adjusting or the handling of accounts receivable.[6] The problem is that the reengineers haven't gone far enough conceptually—they are still dealing with disconnected and aggregated *processes* (for example, order-taking for a whole range of products) rather than the entire *flow of value-creating activities for specific products*. In addition, they often stop at the boundaries of the firm paying their fees, whereas major breakthroughs come from looking at the whole value stream. What's more, they treat departments and employees as the enemy, using outside SWAT teams to blast both aside. The frequent result is a collapse of morale among those who survive being reengineered and a regression of the organization to the mean as soon as the reengineers are gone.

The lean alternative is to redefine the work of functions, departments, and firms so they can make a positive contribution to value creation and to speak to the real needs of employees at every point along the stream *so it is actually in their interest to make value flow.* This requires not just the creation of a *lean enterprise* for each product but also the rethinking of conventional firms, functions, and careers, and the development of a lean strategy, as explained in Part III.

Pull

reduction of lead time

The first visible effect of converting from departments and batches to product teams and flow is that the time required to go from concept to launch, sale to delivery, and raw material to the customer falls dramatically. When flow is introduced, products requiring years to design are done in months, orders taking days to process are completed in hours, and the weeks or months of throughput time for conventional physical production are reduced to minutes or days. Indeed, if you can't quickly take throughput times down by half in product development, 75 percent in order processing, and 90 percent in physical production, you are doing something wrong. What's more, lean systems can make any product currently in production in any combination, so that shifting demand can be accommodated immediately.

So what? This produces a onetime cash windfall from inventory reduction and speeds return on investment, but is it really a revolutionary achievement? In fact, it is because the ability to design, schedule, and make exactly what the customer wants just when the customer wants it means you can throw away the sales forecast and simply make what customers actually tell you they need. That is, you can let the customer *pull* the product from you as needed rather than pushing products, often unwanted, onto the customer. What's more, as explained in Chapter 4, the demands of customers become much more stable when they know they can get what they want right away and when producers stop periodic price discounting campaigns designed to

Let's take a practical example: the book you hold in your hand. In fact, your copy is lucky. One half of the books printed in the United States each year are shredded without ever finding a reader! How can this be? Because publishers and the printing and distribution firms they work with along the value stream have never learned about flow, so the customer can't pull. It takes many weeks to reorder books if the bookseller or warehouse runs out of stock, yet the shelf life of most books is very short. Publishers must either sell the book at the peak of reader interest or forgo many sales. Because the publisher can't accurately predict demand in advance, the only solution is to print thousands of copies to "fill the channel" when the book is launched even though only a few thousand copies of the average book will be sold. The rest are then returned to the publisher and scrapped when the selling season is over.

The solution to this problem will probably emerge in phases. In the next few years, printing firms can learn to quickly print up small lots of books and distribution warehouses can learn to replenish bookstore shelves frequently (using a method described in Chapter 4). Eventually, new "right-sized" book-printing technologies may make it possible to simply print out the books the customer wants at the moment the customer asks for them, either in a bookstore or, even better, in the customer's office or home. And some customers may not want a physical copy of their "book" at all. Instead, they will request the electronic transfer of the text from the "publisher" to their own computer, printing out an old-fashioned paper version only if they happen to need it. The appropriate solution will be found once the members of the publishing value stream embrace the fourth principle of lean thinking: *pull.*

Perfection

As organizations begin to accurately specify *value*, identify the entire *value stream*, make the value-creating steps for specific products *flow* continuously, and let customers *pull* value from the enterprise, something very odd begins to happen. It dawns on those involved that there is no end to the process of reducing effort, time, space, cost, and mistakes while offering a product which is ever more nearly what the customer actually wants. Suddenly *perfection*, the fifth and final principle of lean thinking, doesn't seem like a crazy idea.

Why should this be? Because the four initial principles interact with each other in a virtuous circle. Getting value to flow faster always exposes hidden *muda* in the value stream. And the harder you pull, the more the impediments to flow are revealed so they can be removed. Dedicated product teams in direct dialogue with customers always find ways to specify value more

In addition, although the elimination of *muda* sometimes requires new process technologies and new product concepts, the technologies and concepts are usually surprisingly simple and ready for implementation right now. For example, we recently watched while Pratt & Whitney replaced a totally automated grinding system for turbine blades with a U-shaped cell designed and installed by its own engineers in a short time and at a quarter of the capital cost of the automated system being replaced. The new system cuts production costs by half while reducing throughput times by 99 percent and slashing changeover time from hours to seconds so Pratt can make exactly what the customer wants upon receiving the order. The conversion to lean thinking will pay for itself within a year, even if Pratt receives nothing more than scrap value for the automated system being junked.

Perhaps the most important spur to perfection is *transparency,* the fact that in a lean system everyone—subcontractors, first-tier suppliers, system integrators (often called assemblers), distributors, customers, employees—can see everything, and so it's easy to discover better ways to create value. What's more, there is nearly instant and highly positive feedback for employees making improvements, a key feature of lean work and a powerful spur to continuing efforts to improve, as explained in Chapter 3.

Readers familiar with the "open-book management" movement in the United States[7] will recall that financial transparency and immediate feedback on results, in the form of monetary bonuses for employees, are its central elements. Thus, there is a broad consistency between our approach and theirs. However, a major question emerges for open-book managers as finances are made transparent and employees are rewarded for performance. How can performance be improved? Sweat and longer hours are not the answer but will be employed if no one knows how to work smarter. The techniques for flow and pull that we will be describing in the pages ahead are the answer. What's more, when employees begin to feel the immediate feedback from making product development, order-taking, and production flow and are able to see the customer's satisfaction, much of the carrot-and-stick apparatus of open-book management's financial reward system becomes unnecessary.

The Prize We Can Grasp Now

Dreaming about perfection is fun. It's also useful, because it shows what is possible and helps us to achieve more than we would otherwise. However, even if lean thinking makes perfection seem plausible in the long term, most of us live and work in the short term. What are the benefits of lean thinking which we can grasp right away?

Based on years of benchmarking and observation in organizations around the world, we have developed the following simple rules of thumb: Converting a classic batch-and-queue production system to continuous flow with effective pull by the customer will double labor productivity all the way through the system (for direct, managerial, and technical workers, from raw materials to delivered product) while cutting production throughput times by 90 percent and reducing inventories in the system by 90 percent as well. Errors reaching the customer and scrap within the production process are typically cut in half, as are job-related injuries. Time-to-market for new products will be halved and a wider variety of products, within product families, can be offered at very modest additional cost. What's more, the capital investments required will be very modest, even negative, if facilities and equipment can be freed up and sold.

And this is just to get started. This is the *kaikaku* bonus released by the initial, radical realignment of the value stream. What follows is continuous improvements by means of *kaizen* en route to perfection. Firms having completed the radical realignment can typically double productivity again through incremental improvements within two to three years and halve again inventories, errors, and lead times during this period. And then the combination of *kaikaku* and *kaizen* can produce endless improvements.

Performance leaps of this magnitude are surely a bit hard to accept, particularly when accompanied by the claim that no dramatically new technologies are required. We've therefore worked for several years to carefully document specific instances of lean transformations in a wide range of firms in the leading industrial economies. In the chapters ahead, we provide a series of "box scores" on precisely what can be achieved and describe the specific methods to use.

The Antidote to Stagnation

Lean thinking is not just the antidote to *muda* in some abstract sense; the performance leap just described is also the answer to the prolonged economic stagnation in Europe, Japan, and North America. Conventional thinking about economic growth focuses on new technologies and additional training and education as the keys. Thus the overwhelming emphasis of current-day popular writing on the economy is on falling computing costs and the growing ease of moving data around the planet, as exemplified by the World Wide Web. Coupling low-cost, easily accessible data with interactive educational software for knowledge workers will surely produce a great leap in productivity and well-being, right?

The record is not promising. During the past twenty years we've seen the robotics revolution, the materials revolution (remember when cars would have ceramic engines and airplanes would be built entirely of plastic?), the microprocessor and personal computer revolution, and the biotechnology revolution, yet domestic product per capita (that is, the average amount of value created per person) in all the developed countries has been firmly stuck.

The problem is not with the new technologies themselves but instead with the fact that they initially affect only a small part of the economy. A few companies like Microsoft grow from infants to giants overnight, but the great bulk of economic activity—construction and housing, transport, the food supply system, manufacturing, and personal services—is only affected over a long period. What's more, these activities may not be affected at all unless new ways are found for people to work together to create value using the new technologies. Yet these traditional tasks comprise 95 percent or more of day-to-day production and consumption.

Stated another way, most of the economic world, at any given time, is a brownfield of traditional activities performed in traditional ways. New technologies and augmented human capital may generate growth over the long term, but only lean thinking has the demonstrated power to produce green shoots of growth all across this landscape within a few years. (And, as we will see, lean thinking may make some new technologies unnecessary.)

The continuing stagnation in developed countries has recently led to ugly scapegoating in the political world, as segments of the population in each country push and shove to redivide a fixed economic pie. Stagnation has also led to a frenzy of cost cutting in the business world (led by the reengineers), which removes the incentive for employees to make any positive contribution to their firms and swells the unemployment ranks. Lean thinking and the lean enterprise is the solution immediately available that can produce results on the scale required. This book explains how to do it.

Getting Started

Because lean thinking is counterintuitive and a bit difficult to grasp on first encounter (but then blindingly obvious once "the light comes on"), it's very useful to examine the actual application of the five lean principles in real organizations. The material in the remainder of Part I, therefore, provides real instances of lean principles banishing *muda*. The place to start, as always, is with *value* as defined by the customer.

CHAPTER 1

Value

A House or a Hassle-Free Experience?

Doyle Wilson of Austin, Texas, had been building homes for fifteen years before he got serious about quality. "In October of 1991 I just got disgusted. Such a large part of my business was waiting and rework, with expensive warranty claims and friction with customers, that I knew there must be a better way. Then I stumbled across the quality movement."

He read Carl Sewell's book on car dealing, *Customers for Life*,[1] and decided to test his claims by buying a car at Sewell's Dallas dealership. ("I thought that if even a car dealer could make a customer feel good, it should be easy for a homebuilder!") His purchase was such a positive experience that he asked Sewell for advice on quality in home building and was told to read the works of W. Edwards Deming.

Doyle Wilson is the archetypical Texan and never does things halfway. By February of 1992 he had launched a wall-to-wall Total Quality Management campaign at Doyle Wilson Homebuilder. Over the next three years he personally taught his workforce the principles of TQM, began to collect and analyze enormous amounts of data on every aspect of his business, got rid of individual sales commissions ("which destroy quality consciousness"), eliminated the traditional "builder bonus" for his construction superintendents (who were qualifying for the "on-time completion" bonus by making side deals with customers on a "to-be-done-later" list), reduced his contractor corps by two thirds, and required the remaining contractors to attend (and pay for) his monthly quality seminars.

Customer surveys showed a steady rise in satisfaction with the homebuilding experience and sales grew steadily even in a flat market as Wilson took sales from his competitors. In 1995, Doyle Wilson Homebuilder won the National Housing Quality Award (often called the Baldrige Award for quality of the construction industry), and Wilson set a goal of winning the Baldrige Award itself by 1998. Yet he was not satisfied.

"I knew I was making progress in competing with other builders for the new-home buyer, but a simple fact, once it lodged in my mind, wouldn't go away: 78 percent of the homes bought in central Texas each year are 'used' or older homes. I've been making progress in increasing my share of the 22 percent seeking a new home, but what about the 78 percent who bought older homes? Obviously, these buyers are the real market opportunity."

So instead of surveying people who were buying new homes, Wilson began to talk with people who were buying older homes. What he discovered was obvious in retrospect but has required a complete rethinking of his business. Specifically, he found that many buyers of older homes hated the "hassle factor" in negotiating for new construction, the long lead times to get the job done and move in, the inevitable "to-be-done" list after moving in, and the "phony choices" available from builders who promise custom homes but then load on as "standard equipment" many features of little interest to buyers.

Wilson soon realized that that was exactly what he had been asking his customers to go through. By contrast, older-home customers could clearly see what they were getting, buy only what they wanted, and, often, move in immediately. "No wonder I was losing 78 percent of my potential customers!"

To create a hassle-free experience to go with the house itself (these together constituting Wilson's "product"), it was necessary to rethink every step in the process. He has recently opened a one-stop sales center where the customer can see and decide on every option available in a house (for example, the forty different varieties of brick, the three thousand varieties of wallpaper, the four styles of built-in home office), customize a basic design with the help of an Auto-Cad computer system, select features beyond the standard level (for example, extra-thick carpet pads, additional outdoor lighting, and heavier-duty wiring), determine the exact price, work out the mortgage, arrange for insurance, and arrange for the title search. For customers truly in a hurry this can be done during one walk-through of the sales center.

To shrink the lead time from contract signing to moving in from six months to a target of thirty days, he has reorganized his contract-writing and job-release process and is developing a system of pull scheduling for contractors who are assigned new jobs as downstream jobs are completed. He is also introducing standardized work statements, parts lists, and tool kits for every job. Eventually these steps will eliminate the "to-do" list because the new system does not allow the next task to start until the previous task is certified as complete with perfect quality.[2]

Finally, Wilson has created a wide range of basic house designs with a minimum construction standard and asks the customer to specify all materials and systems upgrades (using the computer design system) to a selected base design so the customer only pays for exactly what she or he feels is really needed.

Doing all of this will not be easy, as we'll see when we return to this example in Chapter 3 on flow, but Doyle Wilson has already made the key leap. Instead of concentrating on conventional markets and what he and his contractors were accustomed to making in a conventional way, he has looked hard at *value* as defined by his customers and set off down a new path.

Start by Challenging Traditional Definitions of "Value"

Why is it so hard to start at the right place, to correctly define value? Partly because most producers want to make what they are already making and partly because many customers only know how to ask for some variant of what they are already getting. They simply start in the wrong place and end up at the wrong destination. Then, when providers or customers do decide to rethink value, they often fall back on simple formulas—lower cost, increased product variety through customization, instant delivery—rather than jointly analyzing value and challenging old definitions to see what's really needed.

Steve Maynard, vice president for engineering and product development at the Wiremold Company in West Hartford, Connecticut, was trying to deal with these very problems when he reorganized Wiremold's product development system in 1992. For many years previously, Wiremold had developed new products—consisting of wire guides for office and industrial users and surge protectors for PCs and other business electronics—through a conventional departmentalized process. It started with marketing, which commissioned surveys comparing Wiremold's products with the offerings of competitors. When an "opportunity" was identified, usually a gap in the market or a reported weakness in a competitor's offering, a design was developed by product engineering, then tested by the prototype group. If it worked according to specification, the design proceeded to the engineers designing the machines to make the products and eventually went into production.

This system produced designs which lacked imagination and which customers often ignored. (The designs also took too much time and effort to develop and cost too much to make, but these are a different type of problem we'll discuss in Chapter 3.) Simply speeding up this process through simultaneous engineering and then broadening product variety would just have brought more bad designs to market faster. Pure *muda*.

Steve Maynard's solution was to form a team for each product to stick with that product during its entire production life. This team—consisting of a marketer, a product engineer, and a tooling/process engineer—proceeded to enter into a *dialogue* with leading customers (major contractors) in which all of the old products and solutions were ignored. Instead, the customer and the producer (Wiremold) focused on the value the customer really needed.

For example, traditional Wiremold wire guides (which channel wiring through hostile factory environments and provide complex arrays of outlets in high-use areas like laboratories and hospitals) had been designed almost entirely with regard to their ruggedness, safety, and cost per foot as delivered to the construction site. This approach nicely matched the mentality of Wiremold's product engineers, who dominated the development process and who found a narrow, "specification" focus very reassuring.

As the new dialogue began, it quickly developed that what customers also wanted was a product that "looked nice" and could be installed at the construction site very quickly. (Wiremold had never employed a stylist and knew relatively little about trends in the construction process.) Customers were willing to make substantial trades on cost per foot to get better appearance (which increased the bid price of construction jobs) and quicker installation (which reduced total cost).

Within two years, as all of Wiremold's product families were given the team treatment, sales for these very conventional products increased by more than 40 percent and gross margins soared. Starting over with a joint customer-producer dialogue on value paid a major dividend for Wiremold quite aside from savings in product development and production costs.

While Wiremold and Doyle Wilson Homebuilder and every other firm needs to be searching for fundamentally new capabilities that will permit them to create value in unimagined dimensions, most firms can substantially boost sales immediately if they find a mechanism for rethinking the value of their core products to their customers.

Define Value in Terms of the Whole Product

Another reason firms find it hard to get value right is that while value creation often flows through many firms, each one tends to define value in a different way to suit its own needs. When these differing definitions are added up, they often don't add up. Let's take another nightmarish (but completely typical) travel example.

One of us (Jones) recently took his family on an Easter holiday in Crete from his home in Herefordshire in the United Kingdom. What was wanted was a total, hassle-free package of transport to the airport, a flight to Crete, was a product pieced together by the user and involving nineteen different operating organizations:

The *travel company* (to book the air tickets and the villa), the *taxi firm* (which doesn't deal with the travel company) handling the long trip from Hereford to London Gatwick—no airline flies nonstop between Birmingham (the nearest airport) and Crete at Easter time, the *ground staffs* at both airports (independent contractors to the airline), the *security staffs* at both airports (more independent contractors), the two *customs staffs* (to check your documents at both ends and to keep themselves occupied doing so), the two *airport authorities* (who love long layovers because spending per passenger goes up), the *airline* (which has been deintegrating and performs less of the support activities for its operations itself), the *air-traffic authorities* in five countries along the route of flight (who follow the standard form for governments by being undercapitalized and optimized for delays), the *bank* exchanging currency at Gatwick airport, the *bus company* to convey the family to the villa in Crete, and the *villa.*

The trip was reasonably routine but look at what the Jones family did to "process" itself through the system:

1. Call the travel company to make the booking.
2. Receive the tickets by mail.
3. Call the taxi company to make the booking.
4. Wait for the taxi.
5. Load the luggage (8:00 A.M. GMT).
6. Drive to the airport (three and a quarter hours), arriving two hours before the scheduled flight time as required by the airline.
7. Unload the luggage.
8. Wait in the currency exchange queue (to change English pounds into Greek drachma).
9. Wait in the check-in line.
10. Wait in the security line.
11. Wait in the customs line.
12. Wait in the departure lounge.
13. Wait in the boarding line.
14. Wait in the airplane (two-hour air-traffic delay).
15. Taxi to the runway.
16. Fly to Crete (three hours).
17. Wait in the airplane (taxi and deboarding).
18. Wait in the baggage-claim line.
19. Wait in the immigration line.
21. Wait in the customs line.
22. Load luggage onto the bus.

24. Travel by bus to the villa (almost forty-five minutes).
25. Unload luggage and carry to villa.
26. Wait to check in at the villa (9:00 P.M. GMT).

The box score:

Total travel time: 13 hours
Time actually going somewhere: 7 hours (54 percent of the total)[3]
Queuing and wait time: 6 hours
Number of lines: 10
Number of times luggage was picked up and put down: 7
Number of inspections (all asking the same questions): 8
Total processing steps: 23

The problem here is not that there were too many firms involved. Each was appropriately specialized for its current task. The problem instead is that each firm was providing a partial product, often only looking inward toward its own operational "efficiency" while no one was looking at the whole product through the eyes of the customer. The minute the focus is shifted to the whole as seen by the customer, obvious questions emerge:

Could one person at check-in handle the security, customs, and check-in tasks? (Letting you walk past them into the boarding area or even onto the plane.) Better yet, could the ticket sent by your travel agent include your baggage tags, boarding passes, taxi voucher, bus tickets, and villa registration, so you just drop these off as you walk through each point? (Or perhaps travelers could create their own ticket using their personal computer linked to reservations systems. They could simply swipe their credit card through a card reader at each point, eliminating paperwork altogether along with the travel agent.) Could the customs authorities in Crete have your passport scanned at check-in in London and use the hours you are en route to figure out whether you ought to be admitted? (Then, unless there is a problem, you could just walk off the plane without visiting immigration and customs at all.) And why (does *anyone* know?) do you need to arrive at the airport two hours before departure? In short, the appropriate definition of the product changes as soon as you begin to look at the whole through the eyes of the customer.

The Critical Need for Lean Firms to Rethink Value

If you take a few moments to reflect on almost any "product"—a good, a service, or more likely both in combination—you will begin to see the same issue of the appropriate way to define it. Doing this will generally require producers to talk to customers in new ways and for the many firms along a value stream to talk to each other in new ways. (We'll see many more examples of this need in the pages ahead—for example, the need for car companies to stop selling a product and car dealers to stop selling services, both to be replaced by a new product [personal mobility] provided jointly to the user.)

It's vital that producers accept the challenge of redefinition, because this is often the key to finding more customers, and the ability to find more customers and sales very quickly is critical to the success of lean thinking. This is because lean organizations, as we will demonstrate shortly, are always freeing up substantial amounts of resources. If they are to defend their employees and find the best economic use for their assets as they strike out on a new path, they need to find more sales right now. Beginning with a better specification of value can often provide the means.

Then, once the initial rethinking of value is done (in what might be called *kaikaku* for value), lean enterprises must continually revisit the value question with their product teams to ask if they have really got the best answer. This is the value specification analog of *kaizen* which seeks to continually improve product development, order-taking, and production activities. It produces steady results along the path to perfection.

The Final Element in Value Definition: The Target Cost

The most important task in specifying value, once the product is defined, is to determine a *target cost* based on the amount of resources and effort required to make a product of given specification and capabilities *if* all the currently visible *muda* were removed from the process. Doing this is the key to squeezing out the waste.

Conventional firms set target selling prices based on what they believe the market will bear. They then work backwards to determine acceptable costs to ensure an adequate profit margin, and they must do this any time they begin to develop a new product. So what's different here? Lean enterprises look at the current bundles of pricing and features being offered customers by conventional firms and then ask how much cost they can take out by full application of lean methods. They effectively ask, What is the *muda*-free cost of this product, once unnecessary steps are removed and value is made to flow? This becomes the target cost for the development, order-taking, and production activities necessary for this product.[4]

Because the target is certain to be far below the costs borne by competitors, the lean enterprise has choices: reduce prices (another way to increase sales volume and utilize freed-up resources); add features or capabilities to the product (which should also increase sales); add services to the physical product to create additional value (and jobs); expand the distribution and service network (again increasing sales, although with a time lag); or take profits to underwrite new products (which will increase sales in the longer term).

Once the target cost is set for a specific product, it becomes the lens for examining every step in the value stream for product development, order-taking, and production (this latter being called operations in the case of a service like insurance or transportation). As we will see in the next chapter, the relentless scrutiny of every activity along the value stream—that is, asking whether a specific activity really creates any value for the customer—becomes the key to meeting the aggressive cost target.

CHAPTER 2

The Value Stream*

The View from the Aisle

An excellent spot for observing the value stream is the aisle of the supermarket, for it is here that a thousand streams empty into the arms of the customer. Not only does the flow of the physical product culminate in the supermarket aisle, as pulled forward by the decisions of the shopper, but also the process of product development as new products are launched. Indeed, Taiichi Ohno found this vantage point in the modern supermarket so stimulating that it inspired him in 1950 to invent the new system of flow management we now call Just-in-Time (JIT).[1]

In the past two years we have been putting ourselves in the aisle, in collaboration with the British grocery chain Tesco[2] and a number of its suppliers, to think through the value stream for specific products in a search for *muda*. To do this we have started to map out every step—each individual action—involved in the process of physical production and order-taking for specific products. Recently we have started to think about product development as well.

Our method is based on a simple premise. Just as activities that can't be measured can't be properly managed, the activities necessary to create, order, and produce a specific product which can't be precisely identified, analyzed, and linked together cannot be challenged, improved (or eliminated altogether), and, eventually, perfected. The great majority of management attention has historically gone to managing aggregates—processes, departments, firms—overseeing many products at once. Yet what's really needed is to manage whole value streams for specific goods and services.

Our initial objective in creating a value stream "map" identifying every

---

* This chapter is based largely on a case study developed by Nick Rich of the Lean Enterprise Research Centre, Cardiff Business School. We are grateful for his help.

action required to design, order, and make a specific product is to sort these actions into three categories: (1) those which actually create value as perceived by the customer; (2) those which create no value but are currently required by the product development, order filling, or production systems (Type One *muda*) and so can't be eliminated just yet; and (3) those actions which don't create value as perceived by the customer (Type Two *muda*) and so can be eliminated immediately. Once this third set has been removed, the way is clear to go to work on the remaining non-value-creating steps through use of the flow, pull, and perfection techniques described in the chapters ahead.

The Value Stream for a Carton[3] of Cola

The only way to make this method clear is to describe a typical value stream analysis.[4] We'll use a product chosen more or less at random in the beverages aisle at Tesco, a cardboard carton of eight cans of cola. We should, however, tell you at the outset that what we will find is fairly horrific—a lengthy set of actions extending over three hundred days, most of which consume resources but create no value and are therefore *muda*. You should understand that looking at any of the thirty thousand other items in the typical Tesco store would produce very much the same result. The cola example is neither better nor worse than the norm.

You should also bear in mind that the firms arrayed along the cola value stream are all competently managed in terms of mass-production thinking. The problem is not the competence of managers operating the system in accord with an agreed logic. The problem is the logic itself.

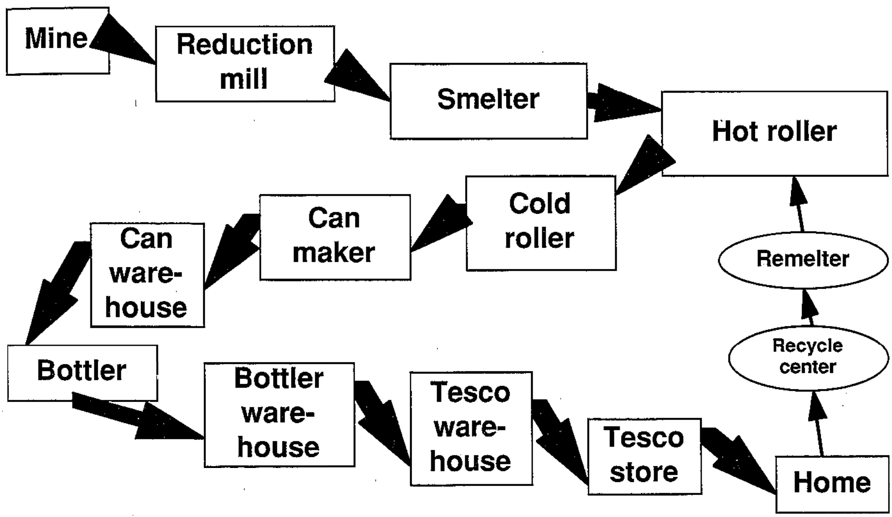
Producing Cola

Even the mightiest river has modest headwaters. For cola one of these is literally water, supplied in the United Kingdom by the local Water Authorities. Other basic ingredients are the "essence" (in plain language, the taste) used in tiny amounts and supplied as a concentrate by the parent cola company,[5] beets for sugar, corn for caramels (to provide the "cola" color and additional taste), fir trees for cardboard to make the carton, and bauxite or recycled cans to create aluminum for the can.[6]

Because the can rather than the actual beverage is by far the most complex aspect of a carton of cola[7]—and the one with the longest production lead time—we'll initially focus our analysis on the flow of aluminum for the can, treating sugar, caramels, essences, and cartons as tributaries joining the stream farther down the valley.

As shown in the value stream map in Figure 2.1, the first step is to mine bauxite in Australia. Although the ore could in principle be mined in small amounts and sent along to the next step within a few minutes of the receipt of an order, the mining machinery is truly massive and the actual process involves scooping out millions of tons of bauxite at a go in accord with a long-term production forecast. The mountain of ore is then transferred to massive trucks for shipment to a nearby chemical reduction mill where the bauxite is reduced to powdery alumina.

FIGURE 2.1: VALUE STREAM FOR COLA CANS



This process, which turns four tons of bauxite into two tons of alumina, requires about thirty minutes. When enough alumina is accumulated to fill an ultralarge ore carrier (over two weeks or so; about 500,000 tons or enough for 10 million cans), it is shipped by sea—a four-week trip—to Norway or Sweden, countries with cheap hydroelectric power, for smelting.

After about a two-month wait at the smelter, the application of an enormous amount of energy (twenty times that needed to melt down and recycle old cans) reduces two tons of alumina to one ton of aluminum in about two hours. Again, scale in smelting dictates that large amounts of aluminum be created in each batch, with the molten aluminum poured into dozens of ingots one meter on each side and ten meters long. These are then carefully cooled and stored for about two weeks before shipment by truck, boat, and truck to a hot rolling mill in Germany or Sweden.

After about two weeks of storage at the hot rolling mill, the ingot is heated to five hundred degrees centigrade and run through a set of heavy rollers three times to reduce the thickness from one meter to three millimeters. The actual rolling process takes about one minute, but the machinery is extremely complex and difficult to change from one specification of product to another, so management has found it best to wait until there are orders in hand for a large amount of material of a given specification and then to process these orders all at once. When this is done for the specification of aluminum needed for cola cans, the aluminum sheet emerging from the rolling mill is wound onto a ten-ton coil and taken to a storage area, where it sits for about four weeks.

When needed for the next step, the coil is taken from storage and shipped by truck to a cold rolling mill, either in Germany or Sweden, where it is stored for about another two weeks. Cold rolling (at 2100 feet of aluminum sheet per minute—about 25 miles an hour) squeezes the aluminum sheet from 3 millimeters to .3 millimeter, the thickness needed by can makers. Because the cold rolling equipment is also extremely expensive and difficult to change over to the next product, the managers of the cold rolling mills have also found it most economical to accumulate orders for products of a given specification and do them all at once. The thin sheet emerging from the cold roller is then slit into narrower widths, wound onto ten-ton coils, and stored for about a month on average.

When needed for can making, the aluminum coils are shipped by truck, by sea, and again by truck to the can maker in England, where the coils are unloaded and stored, again for about two weeks. When needed, the coils are taken from storage to the can making machinery and run through a blanking machine which punches circular discs out of the aluminum sheet at the rate of four thousand per minute. The discs are then fed automatically into "wall drawing" machines, which punch the disc three times in succession to create a can without a top, at the rate of three hundred cans per minute per machine. (Thirteen forming machines are downstream from each blanking machine.)

From the forming machines, the cans travel by conveyor through a washer, a dryer, and a paint booth applying a base coat and then a top coat consisting of the cola color scheme plus consumer information in different languages and varying promotional messages. The cans then travel through lacquering, necking and flanging (to prepare the cans to receive their tops after filling), bottom and inside spraying (to prevent discoloration and any aluminum taste from getting into the cola), and on to final inspection.

The can making machinery just described (really just one big interconaluminum into a finished, painted can—with no human intervention—in less than ten seconds of actual processing time. However, it is also extremely expensive to change over from one type of can to the next and one paint scheme to the next, so management tries to produce large lots of each type. From the can maker's standpoint this is clearly the most economical approach, and it also meshes with the practice of the smelter, hot roller, and cold roller of processing specific types of aluminum in large batches.

After inspection, the cans proceed to an automated palletizing machine which loads the empty cans on pallets, eight thousand to each pallet, and sends them to a massive warehouse for storage until needed, usually four weeks. In the warehouse, they are stored by type of can because the bottling firm eventually filling the cola cans needs a variety of cans with different labels for beverages besides plain cola (for example, diet cola, caffeine-free cola, cherry cola). And even for plain cola, the bottler must support many different packaging configurations and promotional campaigns. Each package and many marketing campaigns require different information to be painted on the cans.[8]

From the can maker's warehouse, the cans are trucked to the bottler's warehouse, where they are stored again, although this time only for about four days. They are then depalletized and loaded into massive can filling machines, where they are washed and filled. It is at this point that the major tributary streams converge in a massive tank adjacent to the filling machine.

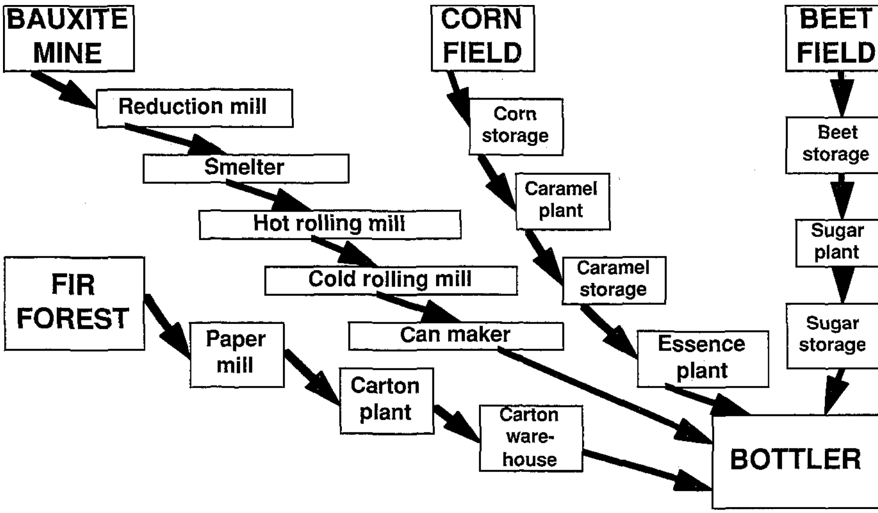
In this step, water, caramels, sugar, and essence are carefully mixed, and carbon dioxide (the fizz) is added to create cola. (Figure 2.2 shows the confluence of the tributaries.) The value streams for these items also require detailed analysis by Tesco, the bottler, and their suppliers, but the method for value stream analysis is best illustrated by sticking to the longest stream.

After the cola is poured into the cans (at the rate of fifteen hundred cans per minute), the cans are sealed with an aluminum can end containing the familiar "pop top," supplied through a separate but very similar process by the can maker. The cans are then date stamped and packed into cartons of varying numbers of cans, eight in the present case. Each type of carton has its own paint scheme and promotional information.

The mixing and filling process, which brings all of the tributary value streams together, requires only one minute to proceed from washing to packing, but it is expensive and time-consuming to change over. In addition, putting cola in a few cans and then a clear soda in the next can requires purging the whole fill system, so the bottler has found it most economical to run large lots of each type of beverage through its complex equipment.[9]

At the end of the filling/packing line, the cartons are palletized, stretch-wrapped (using equipment you will learn a bit more about in Chapter 6), and taken to the bottler's central warehouse serving all customers in the

FIGURE 2.2: CONFLUENCE OF COLA VALUE STREAMS



At the warehouse, the pallets are sorted and placed in designated areas by type. (A process called "stocking.") They are then "picked" as needed and loaded onto one of the bottler's trucks for conveying to one of Tesco's regional distribution warehouses around the U.K.

Once at the Tesco warehouse things move much faster. Incoming pallets are stored for about three days before cases are taken from the pallets and placed in roll cages going overnight to each store. Once at the retail store, the roll cages are taken from the receiving dock to a storage area in the rear or directly to the shelves, and the cola is sold in about two days.

When the cola is taken home it is typically stored again, at least for a few days, perhaps in the basement if the shopper has bought a number of cartons to take advantage of a special promotional offer. Then it's chilled and, finally, consumed. The last step probably requires about five minutes, after nearly a year along the stream.

A final important step, also shown in Figure 2.1, is recycling the can to reintroduce it into the production process at the smelting stage. Currently, only 16 percent of aluminum cans in the U.K. are recycled (and shipped back to Norway), but the percentage is rising. If the percentage of cans recycled moved toward 100 percent, interesting possibilities would emerge for the whole value stream. Mini-smelters with integrated mini-rolling mills might be located near the can makers in England, eliminating in a flash most of the time, storage, and distances involved today in the steps above the can maker. (These activities would suddenly convert from type 1 in our typology—*muda* but unavoidable—to type 2—*muda* that can be completely eliminated right away.) The slow acceptance of recycling is surely due in part to the failure to analyze costs in the whole system rather than just for the recycling step in isolation.

When laid out this way, action by action, so it's possible to see every step for a specific product, the value stream for physical production is highly thought-provoking. First, as shown in Table 2.1, the amount of time when value is actually being created (3 hours) is infinitesimal in relation to the total time (319 days) from bauxite to recycling bin. More than 99 percent of the time the value stream is not flowing at all: the *muda* of waiting. Second, the can and the aluminum going into it are picked up and put down thirty times. From the customer's standpoint none of this adds any value: the *muda* of transport. Similarly, the aluminum and cans are moved through fourteen storage lots and warehouses, many of them vast, and the cans are palletized and unpalletized four times: the *muda* of inventories and excess processing. Finally, fully 24 percent of the energy-intensive, expensive aluminum coming out of the smelter never makes it to the customer: the *muda* of defects (causing scrap).

TABLE 2.1: THE VALUE STREAM OF A CARTON OF COLA

	INCOMING STORAGE*	PROCESSING TIME	FINISHED STORAGE	PROCESS RATE	CUM. DAYS	CUM.† SCRAP
Mine	0	20 min	2 weeks	1000 t/hr	319	0
Reduction mill	2 weeks	30 min	2 weeks		305	0
Smelter	3 months	2 hrs	2 weeks		277	2
Hot rolling mill	2 weeks	1 min	4 weeks	10 ft/min	173	4
Cold rolling mill	2 weeks	<1 min	4 weeks	2100 ft/m	131	6
Can maker	2 weeks	1 min	4 weeks	2000/min	89	20
Bottler	4 days	1 min	5 weeks	1500/min	47	24
Tesco RDC	0	0	3 days	—	8	24
Tesco store	0	0	2 days	—	5	24
Home storage	3 days	5 min	—	—	3	[90]
Totals	5 months	3 hours	6 months		319	24

* Includes transport time from previous step.

† Cumulative scrap is the percentage of the original aluminum scrapped. The jump in scrap at the can maker is due to the loss of about 14 percent of the material in the punch. The loss at the bottler is mainly from damaged cans rejected as they are loaded in the filling machinery. Because the cans are stored empty with no internal pressure, they are easy to damage in handling.

The jump in scrap rate at the home of the customer, shown in brackets, is the consequence of recycling only 16 percent of the 76 percent of the original aluminum which reaches the customer.

The Root Cause of *Muda*

The simplest way to think about this situation is that a can of cola is very small and cola is consumed by the individual customer in small amounts, yet all of the apparatus used to make cola and get it to the customer is very large, very hard to change over, and designed to operate efficiently at very high speeds. The boats, warehouses, and processing machines we have been describing are truly massive and we can see that the primary objective of technologists in the beverage industry has been to scale up and speed up this equipment while removing direct labor, in a classic application of the ideas of mass production.[10]

However, what appears to be efficient to individual companies along the stream—for example, purchase of one of the world's fastest canning machines, operating at fifteen hundred cans per minute, to yield the world's lowest fill cost per can—may be far from efficient when indirect labor (for technical support), upstream and downstream inventories, handling charges, and storage costs are included. Indeed, this machine may be much more expensive than a smaller, simpler, slower one able to make just what the next firm down the stream needs (Tesco in this case) and to produce it immediately upon receipt of the order rather than shipping from a large inventory.

For the moment, let's just reemphasize the critical leap in embracing value stream thinking: Stop looking at aggregated activities and isolated machines —the smelter, the rolling mill, the warehouse, and the can filling machine. Start looking at all the specific actions required to produce specific products to see how they interact with each other. Then start to challenge those actions which singly and in combination don't actually create or optimize value for the customer.

Ordering Cola

If it takes 319 days to bring a cola from bauxite to Tesco (and a similar amount of time to make most of the other items along Tesco's aisles), there is a clear problem in ordering. Either orders must be completely uniform over time so the producers all along the stream can operate stable schedules with little inventory, or the upstream producers must maintain large inventories at every stage to deal with shifts in demand, or Tesco's customers must learn to live with shortages. None of these is desirable because all create *muda*.

In fact, we encountered Tesco because this firm has made remarkable progress in recent years in streamlining its own ordering system to avoid these choices. It has dramatically reduced "stock-outs" (a situation of not having a product the customer wants) while also slashing its own in-store and warehouse inventories by more than half. Because Tesco was already one of the most efficient grocers in the world when it started this process, it appears that its current inventories are only half the U.K. average, a quarter the European average, and an eighth the North American average.

However, Tesco has recently realized that to move even further in reducing inventories, stock-outs, and costs on a total system basis (where more than 85 percent of the costs of a typical product like cola are outside Tesco's corporate control), it will need to improve responsiveness and ordering accuracy all the way up its value stream, running across seven firms in this particular case.[11]

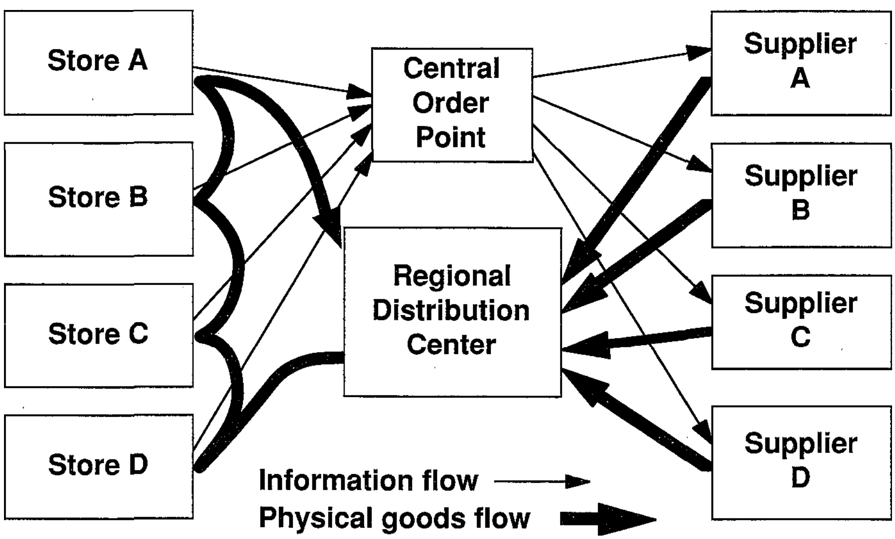
To understand why Tesco reached this conclusion, let's look at their current order-taking system, which is probably the most advanced in the world. Tesco installed a Point-of-Sale (POS) bar-code scanning system in the checkout lanes of all of its stores in the mid-1980s. This permitted each store to maintain a "perpetual inventory" of exactly how much of every item it had on hand and to make more accurate orders to suppliers. This was possible because every time a customer in the aisle took a carton of cola past the checkout, the system noted this fact along with the recent rate of sales and the number of cartons remaining. Replenishment orders could be automatically generated.

A few years later, Tesco transferred decision making on what each store would purchase and when from the store manager, who had been ordering direct from each supplier, to a centralized system where Tesco placed orders combined from all stores to suppliers. At the same time, it opened a dozen Regional Distribution Centers (RDCs) in England so that suppliers for more than 95 percent of all sales volume (the exceptions being milk, sugar, and bread) would ship to the RDC rather than the store. Instead of sending a small truck, partially loaded, to each store, each supplier could send a large truck to each RDC and Tesco could send another large truck to the retail store each night.

In 1989, Tesco took a revolutionary step for the grocery industry by moving toward daily orders (rather than weekly or even monthly) for all fresh products and for many long-shelf-life items. Today, when each store takes inventory at the end of each day, the Tesco ordering system calculates the quantity needed to restore normal stocks plus any special demand likely to be caused by the day of the week, the time of year, the weather, or a sales promotion. After a quick review by the store manger, to check for glitches in the assumptions, this information is dispatched to Tesco's central computer. There, the requirements from all stores in each region are accumulated and orders are dispatched electronically to each supplier during the night.[12] The suppliers are given a precise time (within fifteen minutes) on a precise day[13] to have the precise amount of goods delivered to a specific receiving dock in each RDC. Regional Distribution Center

When the goods arrive at the RDC, they are taken to an area on the floor assigned to each store and consolidated as a load to be taken that night from the RDC to the store, arriving early in the morning. Thus, orders made by each Tesco store on Monday night result in replenishment goods from suppliers reaching each store before it opens on Wednesday morning,[14] effectively creating a twenty-four-hour continuous replenishment system. (The system is shown in Figure 2.3.)

FIGURE 2.3: TESCO REORDER SYSTEM



As a result of this system of daily replenishment, Tesco has increased the "service level" to its retail stores (the percentage of supplier shipments which arrive exactly on time in good condition in exactly the right amount) from 92 to 98.5 percent. At the same time, the stocks on hand of the average good (in the retail stores plus the RDCs) fell from 21 to 12.8 days. For "fast movers" like cola, accounting for more than half of Tesco's total sales, inventories at the RDC and the retail store combined are now only 3 to 5 days.

However, as Tesco did this, they learned the limits of what can be accomplished by one firm alone. Specifically, first-tier suppliers like its bottlers have been fulfilling Tesco orders nightly, just-in-time, but from massive finished goods inventories. Their production methods—with high-speed machines, long changeover times, and large batches—have given them no real choice.[15] Meanwhile, the firms farther up the value stream from the bottler, also using massive high-speed machines with long changeovers to produce large batches, have not yet even taken the step of delivering just-in-time from finished goods inventories. Because the bottler cannot get rapid response from its upstream suppliers to changing levels of demand, it continues to order batches of goods at weekly, monthly, or even quarterly intervals (in the case of some raw materials).

If Tesco wants to shrink costs and improve the reliability of the 85 percent of the value stream it does not directly control, it's obvious that the upstream firms must collectively rethink their operating methods, and this is how Tesco and the Lean Enterprise Research Centre joined forces. While it is still in the early stages, the process of jointly conducting the analysis just described should gradually change Tesco, the bottler, the can maker, the cold roller, the hot roller, the smelter, and the bauxite miner from seven isolated adversaries into a team of collaborators, indeed into a *lean enterprise.*

Creating Cola

The final element in the cola story is the value stream for product development. Historically, in the grocery business, first-tier suppliers like the bottler or the branded purveyor of goods have been responsible for the great bulk of product innovations and introductions. Yet only a brief effort to list the activities in the value stream culminating in the launch of a new product raises many questions.

Typically, a firm like the bottler is continually looking for new products to defend its current market share, to expand its scope of offerings (and justify more shelf space at Tesco), and to substitute products with higher margins for old standbys like cola. In the industry, the typical product development cycle is about one year and consists of a number of product clinics followed by larger product trials culminating in the decision for a full-scale launch.

Although the actual steps involved are very simple and typically involve very little true "research and development," they are conducted sequentially so that if one looks down on a product concept from a bird's-eye view it is quickly apparent that during most of the development period the concept is sitting still, awaiting feedback from the group which conducts the clinics on all of the firm's products or awaiting its place on the schedule of the department which conducts small-scale market trials for all products. Then, when the decision to launch is made, there is more waiting while the production system is adapted to accept the new product, new packaging materials are developed, and the marketing campaign is planned.

The end result of this system is that new products—which are often "new" only in the sense of having reformulated ingredients (for example, caffeine-free and cherry cola)—cost an average of $15 million to launch (half of this going to advertising) and . . . usually fail in the marketplace.[16]

The result for Tesco is large amounts of shelf space tied up with "new" products that don't sell and are launched at the same time in the stores of its direct competitors. The obvious question is: How can it take a year's development time and a $15 million expenditure to introduce a "new" product which isn't new and which no one wants?

Simply reducing development time and expense, while highly desirable, will not be enough to have much effect on this value stream, so Tesco has started to rethink the product development process on a more fundamental level in terms of value. Perhaps, just as the individual steps in the value stream are incomprehensible in isolation, customers do not really want to shop for isolated items. Would it perhaps be better for Tesco and its bottler to jointly undertake the development of the full complement of beverages necessary to keep Tesco customers happy, and for Tesco to develop longer-term relations with its customers so they would not be strangers? Toward this end Tesco has just launched a frequent-shopper program that will gather purchase pattern data on every regular customer and should permit a more coherent value stream in product development.

Putting Value Stream Analysis to Work

Having looked at the specific steps involved in the value stream for one specific product, we are ready to put our findings to work more broadly. In the cola case, unlike the Pratt & Whitney example cited in the Introduction, we do not see any steps in the third category which can be immediately eliminated because they are simply redundant. Instead, we see a large number of steps in the second category. They clearly add no value—they're *muda*—and they therefore become targets for elimination by application of lean techniques.

Note that in performing this analysis we are not "benchmarking" by comparing Tesco's cola value stream with those of its competitors. Although we gave a boost to the benchmarking industry with our previous book, *The Machine That Changed the World*, which described the most comprehensive benchmarking ever attempted in a gigantic global industry, we now feel that benchmarking is a waste of time for managers that understand lean

Lean benchmarkers who discover their performance is superior to their competitors' have a natural tendency to relax (the risk Tesco would run today in benchmarking its internal operations) while mass producers discovering that their performance is inferior often have a hard time understanding exactly why (for example, General Motors and Volkswagen in the 1980s). They tend to get distracted by easy-to-measure or impossible-to-emulate differences in factor costs, scale, or "culture," when the really important differences lie in the harder-to-see ways value-creating activities are organized.

Our earnest advice to lean firms today is simple: To hell with your competitors; compete against *perfection* by identifying all activities that are *muda* and eliminating them. This is an absolute rather than a relative standard which can provide the essential North Star for any organization. (In its most spectacular application, it has kept the Toyota organization in the lead for forty years.) However, to put this admonition to work you must master the key techniques for eliminating *muda*. It all begins with flow.

CHAPTER 3

Flow

The World of Batch-and-Queue

What happens when you go to your doctor? Usually, you make an appointment some days ahead, then arrive at the appointed time and sit in a waiting room. When the doctor sees you—usually behind schedule—she or he makes a judgment about what your problem is likely to be. You are then routed to the appropriate specialist, quite possibly on another day, certainly after sitting in another waiting room. Your specialist will need to order tests using large, dedicated laboratory equipment, requiring another wait and then another visit to review the results. Then, if the nature of the problem is clear, it's time for the appropriate treatment, perhaps involving a trip to the pharmacy (and another line), perhaps a trip back to the specialist for a complex procedure (complete with wait). If you are unlucky and require hospital treatment, you enter a whole new world of specialized functions, disconnected processes, and waiting.

If you take a moment to reflect on your experience, you discover that the amount of time actually spent on your treatment was a tiny fraction of the time you spent going through the "process." Mostly you were sitting and waiting ("patient" is clearly the right word), or moving about to the next step in the diagnosis and treatment. You put up with this because you've been told that all this stopping and starting and being handed off to strangers is the price of "efficiency" in receiving the highest-quality care.

We've already looked briefly at another service, a trip involving an airline. And most of the time the experience is even worse than the Joneses' family trip to Crete because rather than taking a direct flight you must go through a hub for sortation. In the end, the time you spend actually moving along the most direct route is likely to be little more than half the total time required to get from door to door. Yet most travelers put up with this system without dreaming of anything better. After all, it's extremely safe, and travelers are told that it's highly efficient because it fully utilizes expensive airplanes and airports.

Health care and travel are usually called "personal services," in contrast with "products" like VCRs, washing machines, Wiremold's wire guides, and Tesco's beverages. Actually, the major difference is that in the case of health care and travel, you the customer are being acted upon—you are necessarily part of the production process. With goods, by contrast, you wait at the end of the process, seemingly beyond harm's reach. However, there is no escaping the consequences of the way the job gets done even if you are not directly involved.

Let's take just one example for a common good, the single-family home. Henry Ford dreamed about mass-producing homes using standard but modularized designs with the modules built in factories to slash design and production costs while still providing variety. A number of entrepreneurs actually created modular designs and briefly set up production lines in the United States to make the modules for prefabricated houses immediately after World War II.[1] And Toyota has had modest success in Japan since the 1960s in offering a wide range of floor plans and exterior appearances using a few basic modules fabricated on a production line and assembled almost instantly at the construction site.

Yet, almost all of the world's new single-family homes are still built largely at the construction site by cutting and fastening a welter of materials to create the basic structure and then installing thousands of individual components, from plumbing fixtures to kitchen appliances to wall sockets.

If you go to your home builder and then to the construction site and take a seat to watch the action, you will mostly note inaction. For example, when Doyle Wilson started to measure what occurred in his office and at the work site as part of his TQM effort, he discovered that five-sixths of the typical construction schedule for a custom-built home was occupied with two activities: *waiting* for the next set of specialists (architects, cost estimators, bill-of-material drafters, landscape architects, roofers, sheetrockers, plumbers, electricians, landscapers) to work a particular job into their complex schedules, and *rework* to rip out and correct the work just done that was either incorrect from a technical standpoint or failed to meet the needs and expectations of the home buyer.

As the buyer at the end of the process, you pay for all the waiting and rework—grumbling, of course—but it is a custom product, after all, and you've heard many stories from your friends about even worse problems with their homes, so you tend to accept the predominant system and its problems as unavoidable and inherent to the nature of the activity.

In fact, all of these activities—the creation, ordering, and provision of any good or any service—can be made to flow. And when we start thinking about ways to line up all of the essential steps needed to get a job done into a steady, continuous flow, with no wasted motions, no interruptions, no batches, and no queues, it changes everything: how we work together, the kinds of tools we devise to help with our work, the organizations we create to facilitate the flow, the kinds of careers we pursue, the nature of business firms (including nonprofit service providers) and their linkages to each other and society.

Applying flow to the full range of human activities will not be easy or automatic. For starters, it's hard for most managers to even see the flow of value and, therefore, to grasp the value of flow. Then, once managers begin to see, many practical problems must be overcome to fully introduce and sustain flow. However, we do insist that flow principles can be applied to any activity and that the consequences are always dramatic. Indeed, the amount of human effort, time, space, tools, and inventories needed to design and provide a given service or good can typically be *cut in half* very quickly, and steady progress can be maintained from this point onward to cut inputs in half again within a few years.

The Techniques of Flow

So, how do you make value flow? The first step, once value is defined and the entire value stream is identified, is to focus on the actual object—the specific design, the specific order, and the product itself (a "cure," a trip, a house, a bicycle)—and never let it out of sight from beginning to completion. The second step, which makes the first step possible, is to ignore the traditional boundaries of jobs, careers, functions (often organized into departments), and firms to form a lean enterprise removing all impediments to the continuous flow of the specific product or product family. The third step is to rethink specific work practices and tools to eliminate backflows, scrap, and stoppages of all sorts so that the design, order, and production of the specific product can proceed continuously.

In fact, these three steps must be taken together. Most managers imagine that the requirements of efficiency dictate that designs, orders, and products go "through the system" and that good management consists of avoiding variances in the performance of the complex system handling a wide variety of products. The real need is to get rid of the system and start over, on a new basis. To make this approach clear and specific, let's take as a concrete example the design, ordering, and production of a bicycle.

From Batch to Flow in Bicycles

We've chosen this example partly because the bicycle itself is simple and lacks glamour. You will not be distracted by novel product designs or exotic technologies. We've also chosen it because we happen to know something about the bicycle industry, one of us having resolved to test the methods we describe in this book by taking an ownership position in a real bicycle company. Finally, we have chosen bicycle manufacture because it is a deeply disintegrated industry, with most final-assembler firms making only the frame while buying the components—wheels, brakes, gears, seats, handlebars, plus raw materials in the form of frame tubing—from a long list of supplier companies, many larger than the final assemblers themselves. The problems of value stream integration are present in abundance.

DESIGN

Product design in the bicycle industry was historically a classic batch-and-queue affair in which the marketing department determined a "need," the product engineers then designed a product to serve the need, the prototype department built a prototype to test the design, the tooling department designed tools to make a high-volume version of the approved prototype, and the production engineering group in the manufacturing department figured out how to use the tools to fabricate the frame and then assemble the component parts into a completed bike. Meanwhile, the purchasing department, once the design was finalized, arranged to buy the necessary component parts for delivery to the assembly hall.

A design for a new product, usually only one of many under development at a given time, moved from department to department, waiting in the queue in each department. Frequently it went back for rework to a previous department or was secretly reengineered at a point downstream to deal with incompatibilities between the perspectives of, say, the tool designers and the product designers who handled the design in the previous step. There was no flow.

In the late 1980s and early 1990s, most firms switched to "heavyweight" program management with a strong team leader and a few dedicated team members, but without changing the rest of the system. The product "team" was really just a committee with a staff that sent the great bulk of the actual development work back to the departments, where it still waited in queues. What's more, there was no effective methodology for carrying designs through the system without lots of rework and backflows. Even worse, no one was really responsible for the final results of development efforts because the accounting and reward systems never linked the success of a product through its production life with the original efforts of the design team. There was, therefore, a bias toward ingenious designs with admirable technical features which customers liked but which failed to return a profit due to excess costs and launch delays.

The lean approach is to create truly dedicated product teams with all the skills needed to conduct value specification, general design, detailed engineering, purchasing, tooling, and production planning in one room in a short period of time using a proved team decision-making methodology commonly called Quality Function Deployment (QFD).[2] This method permits development teams to *standardize work* so that a team follows the same approach every time. Because every team in a firm also follows this approach, it's possible to accurately measure throughput time and to continually improve the design methodology itself.

With a truly dedicated team in place, rigorously using QFD to correctly specify value and then eliminate rework and backflows, the design never stops moving forward until it's fully in production. The result, as we will demonstrate in the examples in Part II, is to reduce development time by more than half and the amount of effort needed by more than half while getting a much higher "hit rate" of products which actually speak to the needs of customers.

In our experience, dedicated product teams do not need to be nearly as large as traditional managers would predict, and the smaller they can be kept the better all around. A host of narrowly skilled specialists are not needed because most marketing, engineering, purchasing, and production professionals actually have much broader skills than they have (1) ever realized, (2) ever admitted, or (3) ever been allowed to use. When a small team is given the mandate to "just do it," we always find that the professionals suddenly discover that each can successfully cover a much broader scope of tasks than they have ever been allowed to previously. They do the job well and they enjoy it.

Moving most of the employees formerly in marketing, engineering, and production groups into dedicated teams for specific products does create problems for the functional needs of each firm along the value stream, a point we will address in Part III. Similarly, the need to include employees of key component and material supply firms as dedicated members of the product team raises difficult questions of where one firm stops and the next begins, the second major topic of Part III.

Order-Taking

The historic practice in the bicycle industry has been to task the Sales these range from the giant mass-marketers like Wal-Mart at one extreme to thousands of tiny independent bicycle shops at the other. When the orders are fully processed—to make sure that they are internally consistent and that the buyer is credit-worthy—they are sent to the Scheduling Department in Operations or Manufacturing to work into the complex production algorithm for a firm's many products. A shipment date is then set for communication back to Sales and on to the customer.

To check on the progress of orders, particularly in the event of late delivery, the customer calls Sales, which then calls Scheduling. When orders are really late and important customers threaten to cancel, Sales and Scheduling undertake some form of expediting by going directly into the physical production system in both the assembler firm and the supply base to move laggard orders forward. This is done by jumping them to the head of each queue in physical production.

Under the influence of the reengineering movement in the early 1990s, a number of firms integrated Sales and Scheduling into a single department so that the orders themselves can be processed much more quickly—often by one person tied in to the firm's electronic information management system so that orders never need to be handed off, placed in waiting lines, or put down. (They now flow.) As a result, orders can be scheduled for production in a few minutes rather than the days or even weeks previously required; at the same time, order information can be transmitted electronically to suppliers. Similarly, expediting procedures are tightened up to eliminate the confusion which often arose between Sales and Scheduling.

These innovations certainly helped, but a fully implemented lean approach can go much further. In the lean enterprise, Sales and Production Scheduling are core members of the product team, in a position to plan the sales campaign as the product design is being developed and to sell with a clear eye to the capabilities of the production system so that both orders and the product can flow smoothly from sale to delivery. And because there are no stoppages in the production system and products are built to order, with only a few hours elapsed between the first operation on raw materials and shipment of the finished item, orders can be sought and accepted with a clear and precise knowledge of the system's capabilities. *There is no expediting.*

A key technique in implementing this approach is the concept of *takt* time,[3] which precisely synchronizes the rate of production to the rate of sales to customers. For example, for a bicycle firm's high-end titanium-framed bike, let's assume that customers are placing orders at the rate of forty-eight per day. Let's also assume that the bike factory works a single eight-hour shift. Dividing the number of bikes by the available hours of production tells the production time per bicycle, the *takt* time, which is ten

Obviously, the aggregate volume of orders may increase or decrease over time and *takt* time will need to be adjusted so that production is always precisely synchronized with demand. The point is always to define *takt* time precisely at a given point in time in relation to demand and to run the whole production sequence precisely to *takt* time.

In the lean enterprise, the production slots created by the *takt* time calculation—perhaps ten per hour for high-end bicycles (for a *takt* time of six minutes) and one per minute for low-end models (for a *takt* time of sixty seconds)—are clearly posted. This can be done with a simple whiteboard in the product team area at the final assembler but will probably also involve electronic displays (often called *andon* boards) in the assembler firm and electronic transmission for display in supplier and customer facilities as well. Complete display, so everyone can see where production stands at every moment, is an excellent example of another critical lean technique, *transparency* or *visual control.*[4] Transparency facilitates consistently producing to *takt* time and alerts the whole team immediately to the need either for additional orders or to think of ways to remove waste if *takt* time needs to be reduced to accommodate an increase in orders.[5]

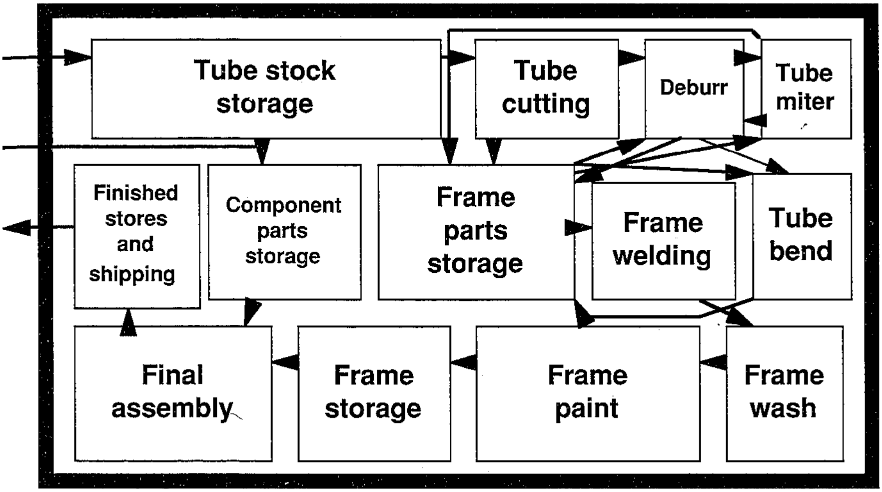
Raising awareness of the tight connection between sales and production also helps guard against one of the great evils of traditional selling and order-taking systems, namely the resort to bonus systems to motivate a sales force working with no real knowledge of or concern about the capabilities of the production system. These methods produce periodic surges in orders at the end of each bonus period (even though underlying demand hasn't changed) and an occasional "order of the century" drummed up by a bonus-hungry sales staff, which the production system can't possibly accommodate. Both lead to late deliveries and bad will from the customer. In other words, they magically generate *muda.*

Production

The historic practice in the bicycle industry was to differentiate production activities by type and to create departments for each type of activity: tube cutting, tube bending, mitering, welding, washing and painting for the frame and handle bars, and final assembly of the complete bike. Over time, higher-speed machines with higher levels of automation were developed for tasks ranging from cutting and bending to welding and painting. Assembly lines were also installed to assemble a mix of high-volume models in dedicated assembly halls.

All bike makers produced a wide range of models using the same production equipment, and part fabrication tools typically ran at much higher speeds (expressed as pieces per minute) than the final assembly line. Because changing over part fabrication tools to make a different part was typically quite time-consuming, it made sense to make large batches of each part before changing over to run the next part. The typical final assembly plant layout and materials flow looked as shown in Figure 3.1.

FIGURE 3.1: BICYCLE PLANT LAYOUT AND FLOW



As batches of parts were created, an obvious problem arose: how to keep track of the inventory and make sure that the right parts were sent to the right operation at the right time. In the early days of the bicycle industry—an activity dating back to the 1880s and a key precursor to the auto industry—scheduling was handled by means of a master schedule and daily hand-written orders to each department to make the parts final assembly would need.

After nearly a hundred years, these manual scheduling methods were replaced in the 1970s by computerized Material Requirements Planning systems, or MRPs. A good MRP system was at least 99 percent accurate in keeping track of inventory, ordering materials, and sending instructions to each department on what to make next. As a group, these systems were a clear improvement on older manual systems for controlling batch-and-queue operations and became progressively more complex over time. Eventually capacity planning tools were added to evaluate the capacity of machines at every step in the production process and to guard against the emergence of bottlenecks and capacity constraints.

MRP, however, had a number of problems. If even one part was not properly logged into the system as it proceeded from one production stage to the next, errors began to accumulate that played havoc with the reorder "triggers" telling a department when to switch over to the next type of part. As a result, downstream manufacturing operations often had too many parts (the *muda* of overproduction) or too few parts to meet the production schedule (producing the *muda* of waiting).

A worse problem was that total lead times in batch-and-queue systems were usually quite lengthy—typically a few weeks to a few months between the point in time when the earliest upstream part was produced and the moment when a bike containing that part was shipped to the retailer. This would have been fine if orders had been perfectly smooth, but in fact orders received by the bike manufacturer changed all the time, partly due to the bonus-driven selling system, partly due to the substantial inventories in the retail channel, and partly due to seasonal demand patterns, particularly for low-end bikes. What's more, there were often engineering changes in bicycle designs, even for mature products, meaning that a considerable fraction of the parts piled up alongside the value stream were suddenly either completely obsolete or in need of rework.[6]

MRP systems which were very simple in concept therefore became exceedingly complex in practice. In the bicycle industry, every firm's MRP system was supplemented by a backup system of expediters moving through the production system to move parts in urgent shortage downstream to the head of the queue in every department and at every machine. Their efforts, while essential to avoiding cancellations or large penalties on overdue orders, played havoc with the internal logic of the MRP system—often causing it to generate absurd orders—and with inventory accuracy as well. In the end, most MRP applications were better than manual systems, but they operated day to day at a level of performance far below what was theoretically possible and what had been widely expected when MRP was first introduced.

Just-in-Time, an innovation pioneered at Toyota in the 1950s and first embraced by Western firms in the early 1980s, was designed to deal with many of these problems. This technique was envisioned by Taiichi Ohno as a method for facilitating smooth flow, but JIT can only work effectively if machine changeovers are dramatically slashed so that upstream manufacturing operations produce tiny amounts of each part and then produce another tiny amount as soon as the amount already produced is summoned by the next process downstream. JIT is also helpless unless downstream production steps practice level scheduling (*heijunka* in Toyota-speak) to smooth out the perturbations in day-to-day order flow unrelated to actual customer demand. Otherwise, bottlenecks will quickly emerge upstream and buffers ("safety stocks") will be introduced everywhere to prevent them.

The actual application of JIT in the bicycle industry largely ignored the need to reduce setup times and smooth the schedule. Instead, it concentrated on suppliers, making sure that they only delivered parts to the final assemblers "just in time" to meet the erratic production schedule. In practice, most suppliers did this by shipping small amounts daily or even several times a day from a vast inventory of finished goods they kept near their shipping docks. Some final assemblers even specified the existence of these safety stocks and periodically sent around their purchasing staffs to inspect them. In the end, "just in time" was little more than a once-and-for-all shift of massive amounts of work-in-process from the final assembler to the first-tier supplier and, in turn, from first-tier supplier to firms farther upstream.

To get manufactured goods to flow, the lean enterprise takes the critical concepts of JIT and level scheduling and carries them all the way to their logical conclusion by putting products into continuous flow wherever possible. For example, in the case of the bicycle plant shown in Figure 3.1, flow thinking calls for the creation of production areas by product family, which includes every fabrication and assembly step. (Product families can be defined in various ways, but in this industry they would logically be defined by the base material used for the frame, specifically titanium, aluminum, steel, or carbon-fiber. This classification makes sense because the fabrication steps and processing techniques are quite different in each case.)

Better yet, if noise problems can be managed, the lean enterprise groups the product manager, the parts buyer, the manufacturing engineer, and the production scheduler in the team area immediately next to the actual production equipment and in close contact with the product and tool engineers in the nearby design area dedicated to that product family. The old-fashioned and destructive distinction between the office (where people work with their minds) and the plant (where people work with their hands) is eliminated.

(We're often struck that in the old world of mass production, the factory workforce really had no need to talk to each other. They were supposed to keep their heads down and keep working and professionals rarely went near the scene of the action. So production machinery could make a lot of noise. The isolated workers simply donned their ear protection and shut out the world. In the lean enterprise, however, the workforce on the plant floor needs to talk constantly to solve production problems and implement improvements in the process. What's more, they need to have their professional support staff right by their side and everyone needs to be able to see the status of the entire production system. Many machine builders are still oblivious to the fact that a lean machine needs to be a quiet machine.)

In the continuous-flow layout, the production steps are arranged in a sequence, usually within a single cell, and the product moves from one step to the next, one bike at a time, with *no* buffer of work-in-process in between, using a range of techniques generically labeled "single-piece flow." To achieve single-piece flow in the normal situation when each product family includes many product variants—in this case, touring and mountain bike designs in a wide range of sizes—it is essential that each machine can be converted almost instantly from one product specification to the next. It's also essential that many traditionally massive machines—paint systems being the most critical in the bike case—be "right-sized" to fit directly into the production process. This, in turn, often means using machines which are simpler, less automated, and slower (but perhaps even more accurate and "repeatable") than traditional designs. We will look in detail in Chapter 8 at the Pratt & Whitney example of simplified blade grinding machinery that we mentioned in the Introduction.

This approach seems completely backward to traditional managers who have been told all their lives that competitive advantage in manufacture is obtained from automating, linking, and speeding up massive machinery to increase throughput and remove direct labor. It also seems like common sense that good production management involves keeping every employee busy and every machine fully utilized, to justify the capital invested in the expensive machines. What traditional managers fail to grasp is the cost of maintaining and coordinating a complicated network of high-speed machines making batches. This is the *muda* of complexity.

Because conventional "standard-cost" accounting systems make machine utilization and employee utilization their key performance measures while treating in-process inventories as an asset—even if no one will ever want them—it's not surprising that managers also fail to grasp that machines rapidly making unwanted parts during 100 percent of their available hours and employees earnestly performing unneeded tasks during every available minute are only producing *muda.*

To get continuous-flow systems to flow for more than a minute or two at a time, every machine and every worker must be completely "capable." That is, they must always be in proper condition to run precisely when needed and every part made must be exactly right. By design, flow systems have an everything-works-or-nothing-works quality which must be respected and anticipated. This means that the production team must be cross-skilled in every task (in case someone is absent or needed for another task) and that the machinery must be made 100 percent available and accurate through a series of techniques called Total Productive Maintenance (TPM). It also means that work must be rigorously *standardized* (by the work team, not by chines must be taught to monitor their own work through a series of techniques commonly called *poka-yoke*, or mistake-proofing, which make it impossible for even one defective part to be sent ahead to the next step.[7]

A simple example of a *poka-yoke* is installing photo cells across the opening of each parts bin at a workstation. When a product of a given description enters the area the worker must reach into the boxes to get parts, breaking the light beam from the photo cells on each box. If the worker attempts to move the product on to the next station without obtaining the right parts, a light flashes to indicate that a part has been left out.

These techniques need to be coupled with *visual controls*, as mentioned earlier, ranging from the 5Ss[8] (where all debris and unnecessary items are removed and every tool has a clearly marked storage place visible from the work area) to status indicators (often in the form of *andon* boards), and from clearly posted, up-to-date standard work charts to displays of key measurables and financial information on the costs of the process. The precise techniques will vary with the application, but the key principle does not: Everyone involved must be able to see and must understand every aspect of the operation and its status at all times.

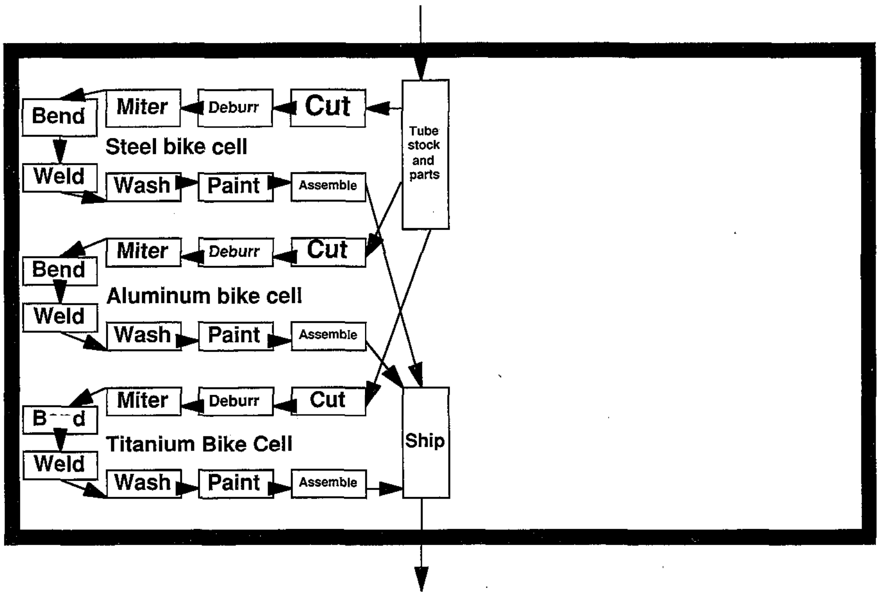
Once the commitment is made to convert to a flow system, striking progress can be made very quickly in the initial *kaikaku* exercise. However, some tools (for example, massive paint booths with elaborate emission control equipment) will be unsuited for continuous-flow production and won't be easy to modify quickly. It will be necessary to operate them for an extended period in a batch mode, with intermediate buffers of parts between the previous and the next production step. The key technique here is to think through tool changes to reduce changeover times and batch sizes to the absolute minimum that existing machinery will permit.[9] This typically can be done very quickly and almost never requires major capital investments. Indeed, if you think you need to spend large sums to convert equipment from large batches to small batches or single pieces, you don't yet understand lean thinking.

The original small-lot, quick-change techniques pioneered at Toyota in the 1960s are a striking achievement, but we caution readers not to take quick-change machines still producing batches, however small, as an end in themselves. Any changeover requiring any loss in production time and any machine which must run at a rate far out of step with the rest of the production sequence can still create *muda*. The end objective of flow thinking is to totally eliminate all stoppages in an entire production process and not to rest in the area of tool design until this has been achieved.

Ziel

Let's tie all of these techniques together by showing what a lean bicycle production process looks like, as shown in Figure 3.2. First, note that the half empty, in large part because all of the in-process storage areas have disappeared. Although the diagram cannot show this, the human effort needed to produce a bicycle has been cut in half as well, and time through the system has been reduced from four weeks to four hours. (We'll talk in Part II about what to do with people no longer needed for their traditional tasks as *muda* is eliminated. Protecting their jobs by finding them other productive tasks is a central part of any successful lean transition.)

FIGURE 3.2: LEAN BICYCLE PLANT LAYOUT AND FLOW



The diagram does show that single large machines have been broken down into multiple small machines, in particular the washing systems and paint booths, so that bikes can proceed continuously, one at a time, from tube cutting to mitering to bending to welding to washing to painting to final assembly without ever stopping. In this arrangement the inventory between workstations can be zero and the size of the work team can be geared to the production volume of the cell, with high-volume cells having more workers than low-volume cells. Finally, note that the track assembly operations have been eliminated. When production is broken into product families, it is often the case that no family accounts for the kind of volume needed for track assembly. Remarkably, manual advancing of the product through assembly is often cheaper.

Because the work flow has been so drastically simplified, the MRP system and the accompanying expediters are no longer needed to get parts from step to step. (MRP still has a use for long-term capacity planning for the assembler firm and its suppliers.) When the sequence is initiated at the end of final assembly, work progresses from each station to the next in accordance with *takt* time and at the same rate as final assembly.

The entire product team including the team leader, the production engineer, the planner/buyer, the TPM/maintenance expert, and the operators (collectively the heart of the lean enterprise) can be located immediately adjacent to the machinery for each product cell. Because the process machinery currently available for these operations in the bicycle industry either makes very little noise inherently—for example, paint—or can be shielded so that very little noise escapes into the team area—the mitering step—it's possible to lay out activities so everyone can see the whole operation and its status at a quick glance.

A final point about the cells which is hard to illustrate with a diagram is that the work in each step has been very carefully balanced with the work in every other step so that everyone is working to a cycle time equal to *takt* time. When it's necessary to speed up or slow down production, the size of the team may be increased or shrunk (contracting or expanding job scope), but the actual pace of physical effort is never changed. And when the specification of the product changes, the right-sized machines can be added or subtracted and adjusted or rearranged so that continuous flow is always maintained.

Right Location

Only one more flow technique needs mentioning, which is to locate both design and physical production in the appropriate place to serve the customer. Just as many manufacturers have concentrated on installing larger and faster machines to eliminate direct labor, they've also gone toward massive centralized facilities for product families (sometimes called "focused factories") while outsourcing more and more of the actual component part making to other centralized facilities serving many final assemblers. To make matters worse, these are often located on the wrong side of the world from both their engineering operations and their customers (Taiwan in the bicycle case) to reduce the cost per hour of labor.

The production process in these remotely located, high-scale facilities may even be in some form of flow, but launching products and improving the process machinery is much harder (because the core engineering skills are on the other side of the world), and the flow of the product stops at the end of the plant. In the case of bikes, it's a matter of letting the finished product sit while a whole sea container for a given final assembler's warehouse in North America is filled, then sending the filled containers to the port, where they sit some more while waiting for a giant container ship. After a few weeks on the ocean, the containers go by truck to one of the bike firm's regional warehouses, where the bikes wait until a specific customer order needs filling, often followed by shipment to the customer's warehouse for more waiting. In other words, there's no flow except along a tiny stretch of the total value stream inside one isolated plant.

The result is high logistics costs and massive finished unit inventories in transit and at retailer warehouses. Another consequence is obsolete goods, eventually sold at large discounts, created by the need to place orders based on forecasts months in advance of demonstrated demand. When carefully analyzed, these costs and revenue losses are often found to more than offset the savings in production costs from low wages, savings which can be obtained in any case by locating smaller flow facilities incorporating more of the total production steps much closer to the customer. (We'll return to this point in Chapter 10 on Japan, because wrong location, rather than high wages, lies at the heart of Japan's current competitive dilemma.)

Applying Flow Thinking to Any Activity

Flow thinking is easiest to see in conventional, discrete-product manufacturing, which is where flow techniques were pioneered. However, once managers learn to see it, it's possible to introduce flow in any activity and the principles are in every case the same: Concentrate on managing the value stream for the specific service or good, eliminate organizational barriers by creating a lean enterprise, relocate and right-size tools, and apply the full complement of lean techniques so that value can flow continuously. At the end of this volume, in Chapter 13, we'll apply lean thinking to a wide range of activities besides traditional manufacturing.

Flow in Work; Work as Flow

So far, we have been talking about the flow of value as if the needs of the customer and the investor are the only ones which count. However, we all know from our daily lives that our experience as producers (that is, as employees and workers) is often far more significant than our activities as consumers or investors. What does the transition to flow mean for the experience of work?

Let's begin with a brief look at the recent research findings of the Polish-born psychologist Mihaly Csikszentmihalyi, now at the University of Chicago. He has spent the last twenty-five years reversing the usual focus of psychology. Instead of asking what makes people feel bad (and how to change it) he has explored what makes people feel good, so that positive attributes of experience can be built into daily life.

His method has been to attach beepers, which sound at random intervals, to his research subjects. When the beeper sounds, the subject is asked to record in a notebook what she or he was doing and how they were feeling. After sifting decades of notebook data from thousands of subjects around the world, he has reached some very simple conclusions.

The types of activities which people all over the world consistently report as most rewarding—that is, which make them feel best—involve a clear objective, a need for concentration so intense that no attention is left over, a lack of interruptions and distractions, clear and immediate feedback on progress toward the objective, and a sense of challenge—the perception that one's skills are adequate, but just adequate, to cope with the task at hand.

When people find themselves in these conditions they lose their self-consciousness and sense of time. They report that the task itself becomes the end rather than a means to something more satisfying, like money or prestige. Indeed, and very conveniently for us, Csikszentmihalyi reports that people experiencing these conditions are in a highly satisfying psychological state of *flow*.[10]

Csikszentmihalyi's classic flow experience is rock climbing, where the need for concentration is obvious and the task itself is clearly the end, not a means. Participation sports less dangerous than rock climbing, interactive games, and focused intellectual tasks (such as writing books!) are often mentioned by Csikszentmihalyi's respondents as flow experiences. However, traditional work-related tasks are only rarely mentioned despite the fact that work is rated the most important overall life activity. This is for a good reason. Classic batch-and-queue work conditions are hardly conducive to psychological flow. The worker can see only a small part of the task, there is often no feedback (much less immediate feedback), the task requires only a small portion of one's concentration and skills, and there are constant interruptions to deal with other tasks in one's area of responsibility.

By contrast, work in an organization where value is made to flow continuously also creates the conditions for psychological flow. Every employee has immediate knowledge of whether the job has been done right and can see the status of the entire system. Keeping the system flowing smoothly with no interruptions is a constant challenge, and a very difficult one, but the product team has the skills and a way of thinking which is equal to the challenge. And because of the focus on perfection, to be further explored in

Chapter 5, the whole system is maintained in a permanent creative tension which demands concentration.

Flow Is Not Enough

We've now seen striking examples of what happens when the value stream flows smoothly. What's more, there is absolutely no magic involved. Any organization can introduce flow in any activity. However, if an organization uses lean techniques only to make unwanted goods flow faster, nothing but *muda* results. How can you be sure you are providing the services and goods people really want when they really want them? And how can you tie all the parts of a whole value stream together when they can't be conducted in one continuous-flow cell in one room? Next you need to learn how to *pull.*

CHAPTER 4

Pull

Pull in simplest terms means that no one upstream should produce a good or service until the customer downstream asks for it, but actually following this rule in practice is a bit more complicated. The best way to understand the logic and challenge of pull thinking is to start with a real customer expressing a demand for a real product and to work backwards through all the steps required to bring the desired product to the customer. Bob Scott's bumper for his out-of-production 1990 Toyota pickup truck provides a mundane but perfectly typical example.

In August 1995, Bob Scott backed his pickup into a pole near Glenside, Pennsylvania, and bent his rear bumper to a point where it couldn't be straightened. He was determined that his truck look sharp—it was originally ordered with the "deluxe" chrome bumper at extra cost—and the severity of the dent also meant that the trailer hitch on the bumper was no longer safe to use. He needed a new bumper.

When Bob Scott took his pickup to Sloane Toyota in Glenside to get a new bumper installed, he touched off a pull sequence just at the point Toyota was taking a major step in its decade-long effort to synchronize the effort of its dealers, its parts distribution system, and its suppliers so customers could truly pull the flow of value all the way through a highly complex production and service system.

The Bad Old Days of Production

If Bob Scott had wrecked his bumper a year or two earlier, nothing would have happened immediately. When he tried to pull, Sloane Toyota wouldn't have had the right bumper on hand for his out-of-production vehicle. Using a traditional stocking system, it's simply impractical for a car dealer to keep on hand a wide range of replacement parts for older vehicles. With about ten thousand part numbers per vehicle, the carrying cost of the inventory would be staggering.

Instead, Sloane Toyota would have needed several days to get a bumper shipped by truck from a Toyota parts warehouse or used expensive overnight freight in order to get it delivered the next day. Bob Scott would have lost the use of his vehicle for some period of time or paid a premium if he wanted it the next day, and in either case would have been an unhappy customer.

Yet even as he waited, there would have been stacks of the precise bumper needed, indeed mountains of them, in Toyota's parts warehouses and at the bumper maker because no satisfactory method was in place for pulling. To see why this was so and to understand what is being done to implement a true pull system all the way along the value stream, let's go back in time and very nearly to the headwaters of the stream, to the Bumper Works factory in Danville, Illinois, which made the bumper Bob Scott cracked.

Shahid Khan, the president of Bumper Works's parent firm, Flex-N-Gate Corporation, is practically a cliché of the American dream. He came to the United States from Pakistan when he was sixteen to go to engineering school at the University of Illinois in Urbana. To put himself through school, he got a job running a massive stamping press in the down-and-out Bumper Works factory in nearby Danville. When he graduated he became the engineering director of Bumper Works and then, by the time he was twenty-eight, he had raised the funds to buy the company.

When Khan entered Bumper Works in 1970, he also entered the world of batch-and-queue. Bumper Works made chromed and painted steel bumpers in a variety of styles for customizing pickup trucks at the car dealership. It made large batches of each type of bumper—typically a month's worth—before shifting production to the next model and sold the bumpers to new-car dealers and crash-repair body shops through a complex wholesale distribution system.

Because large batches were considered normal in this world, it was not important that it took sixteen hours to change over Bumper Works's stamping presses. Because large batches of raw materials were considered unavoidable, Bumper Works had a warehouse at the end of its plant to receive flat sheets of steel by the ton from the steel company. And because the chroming company performing the key step in the middle of the production process also worked in a batch mode, Bumper Works piled up partially made bumpers in its intermediate goods warehouse until there was an enormous batch and then shipped them to the chromer all at once.

When the chromer shipped them back, all in a batch, they were run through a final assembly operation (to install inner reinforcing bars, attachment brackets, and cosmetic coverings), stored once more in a finished goods warehouse, and sent in a batch to the customer according to a predetermined schedule.

As Shahid Khan grew his business in the 1980s, he began to supply replacement bumpers to the service parts organizations of the American Big Three auto companies and he did very well. His batch thinking and their batch thinking were a match. However, Kahn had always set his standards very high, so in 1984 he approached Toyota about supplying bumpers for the pickups they were importing from Japan. This would give him their "crash" parts business as well.

In 1985, Bumper Works was signed on as a supplier for a small volume of Toyota business, and in 1987 won a sole-source contract for the bumpers on the new version of Toyota's small pickup (the model Bob Scott bought). By 1989, Bumper Works was Toyota's sole bumper supplier for North American needs.

There was only one problem: Bumper Works's production system was still a classic case of batch-and-queue. Toyota took Shahid Khan and his senior managers on their first trip to Japan late in 1989 and walked them through showcase lean suppliers, but as Khan remembers, "The light didn't come on; I really couldn't figure out how they could stay in business using the strange practices I saw." So in May 1990, Toyota told Khan they were dispatching a lean *sensei*,[1] a master of the Toyota system, as Khan's personal tutor.

In fact, Toyota sent a number of *sensei* from its Operations Management Consulting Division, the group established in 1969 by Taiichi Ohno to promote lean thinking within Toyota and in the firms in its supplier group.[2] They stayed for months at a time, and by the end of 1992, they had totally transformed Bumper Works—a unionized, grimy operation using old tools in old facilities—into one of the best examples of lean production in North America.

Lean Production for Pull

The first thing the Toyota *sensei* noted at Bumper Works was the massive inventories and batches. Nothing flowed. Immediately right-sizing the massive stamping presses to permit single-piece flow was not possible, so the only solution was to drastically reduce their changeover times and shrink batch sizes. Changeover times were already down from sixteen hours in the mid-1980s to around two hours, but this was not nearly enough.

The Toyota *sensei* applied their standard formula that machines should be available for production about 90 percent of the time and down for changeovers about 10 percent of the time. Then they looked at the range of products Bumper Works would need to make every day. They concluded that the large presses would need to be changed over in twenty-two minutes or less and the small presses in ten minutes or less. (In fact, the numbers were soon down to sixteen minutes and five minutes, respectively.)

Next, the plant was physically reorganized so flat sheets flowed directly from the receiving dock to the blanking machine, which cut the steel into rectangular shapes just larger than a bumper. The blanks then went immediately to the adjacent cell of three stamping presses, where they were given their shape. Next, they were shipped at frequent intervals to the outside-the-plant chroming operation and returned to the welding shop adjacent to the stamping presses. There, the inner and outer parts of the bumper plus the brackets for attaching the bumper to the vehicle were welded together. Finally, the bumpers went straight to the shipping dock just in time for scheduled shipment. *But they flowed only when pulled by the next step.* That is, the blanking machine did nothing until it received a signal from the stamping machines and the stamping machines made nothing until instructed to do so by the welding booth. Each activity pulled the next. The shipping schedule and *takt* time became the pacemaker for the entire operation.

Because most of Bumper Works's customers, as of 1992, were still ordering massive batches—one-month lots to be delivered by the last day of the month—Bumper Works decided to prepare for the future by creating its own daily schedule using a technique Toyota calls *level scheduling.* Shahid Khan's production manager would take the orders for the next month, let's say 8,000 of Bumper A, 6,000 of Bumper B, 4,000 of Bumper C, and 2,000 of Bumper D. She would add them up (to get 20,000) and divide by the number of working days in the month (say, twenty) to discover that Bumper Works would need each day to make 400 of Bumper A, 300 of Bumper B, 200 of Bumper C, and 100 of Bumper D (with a *takt* time of .96 minutes). This would require four changeovers of the blanking and stamping machines, totaling 88 minutes (9 percent of the 960 minutes of two-shift working time) at the maximum allowable changeover speed of 22 minutes.

The daily schedule was given to the welding booth to start the process. As the booth used up its reserve of inner and outer panels and brackets for Bumper A, the welders would slide the empty parts tub and its associated *kanban,* or signal card, down the short slide to the stamping machines. This provided the only signal needed to stamp more parts for Bumper A. Then, as the stamping press used up its blanks for Bumper A, the empty parts tub was sent back down the slide to the blanking machine, providing the only signal needed to make more blanks for Bumper A.

The in-plant MRP system that had been sending orders to every machine—but which never quite worked right so that expediting was always necessary to keep production going—was no longer needed. The new, simple system of pull and visual control always worked once the inevitable start-up problems were resolved. Bumper Works's new operating doctrine could be summarized simply as "Don't make anything until it is needed; then make it very quickly."

But there was a problem right in the heart of the new system. The steel bumpers once welded required a coat of chrome before they could go to final assembly. This was a complex process conducted by specialist firms operating in batch mode. Shahid Khan's chromer, Chrome Craft in Highland Park, Michigan (near Detroit), was the best supplier Bumper Works had found but was not in step with the new approach. Bumpers disappeared into Chrome Craft and didn't reappear for weeks. What was more, getting a rapid turnaround on expedited orders was impossible.

Khan and the Toyota *sensei* were soon on their way to Chrome Craft, where president and owner Richard Barnett watched with some amazement as rapid changeovers were implemented on Barnett's bumper polishing machines so that small lots could be pulled from the loading dock, taken through the necessary polishing process, and run through the long line of chroming tanks. (Chrome Craft was doing bumpers for other manufacturers and had dozens of bumper types going through its plant.)

By arranging for quick unloading and loading of the Bumper Works truck, it became possible to bring a load of bumpers in at 7:00 A.M. while picking up the load just completed, then to return at 3:00 P.M. to pick up the freshly chromed versions of the bumpers dropped off at 7:00 A.M. By 1995, a bumper's time in the Chrome Craft plant had fallen from fifteen days, on average, to less than a day. What was more, at the end of every shift the entire output of Toyota bumpers was being trucked out of the plant, leaving zero in-process inventories. Chrome Craft's inventory "turns" on Toyota bumpers had zoomed from about twenty to about five hundred per year.

Even this achievement is by no means the limit. In mid-1995, Chrome Craft helped install a right-sized chroming operation in a new Flex-N-Gate plant in Indiana making bumpers for the American Big Three. This brings time-in-process down from twenty-four hours (consisting of two eight-hour truck rides from Bumper Works to Chrome Craft and back, plus eight hours at Chrome Craft) to about eight hours.

As Bumper Works learned how to pull value through its system, it became capable of responding practically instantly to customer orders. Because of its quick changeover ability, Bumper Works could start welding a given type of bumper within about twenty minutes of receiving an order and it could easily vary its entire production as demand changed. All that was needed was to drop off a new set of order cards at the welding booth. Similarly, the time elapsed between the arrival of a flat sheet of steel on Bumper Works's loading dock and the shipment of a finished bumper to the customer fell from an average of four weeks to forty-eight hours. Quality also zoomed, as it always does when flow and pull thinking are put in place together. As of mid-1995, Bumper Works hadn't shipped a bad bumper to Toyota in five years.

The new system gave Bumper Works and Chrome Craft the ability to make small lots of bumpers at short notice—for example, a few replacement bumpers of the type Bob Scott needed—but Khan's customers did not know how to take advantage of his new capabilities. Until very recently, even Toyota was still ordering large batches, then erratically changing its orders as shortages developed in the distribution system. Another step was needed to create a smoothly pulling value stream.

The Bad Old Days of Distribution

When Toyota introduced its Corona model in America in 1965, it suddenly began to sell large numbers of cars. These needed service parts, everything from new bumpers to replace those crumpled in accidents (like Bob Scott's) to oil filters and spark plugs for periodic maintenance. Because of the long shipping time from Japan, Toyota needed large stocks of parts in North America and soon built a network of warehouses—called Parts Distribution Centers, or PDCs—stretching from Los Angeles to Boston.

In 1965, the Toyota Production System (TPS) was just being implemented in Toyota's supplier plants in Toyota City. No one had given any thought to applying TPS principles to Toyota's Japanese service parts warehouses, much less in faraway American warehouses. As a result, the eleven PDCs Toyota built in the United States were laid out like every warehouse in America. Each had vast bins stacked to a high ceiling, thousands and thousands of them, one for each type of part. The bins were lined up in long rows to create endless aisles in a massive square of a building.

The PDCs received parts from Japan in sealed containers, typically in large batches coming off massive container ships at weekly intervals. When the containers arrived at the PDC, they were opened in the receiving area and the parts were given to "stockers" with carts who walked up and down the aisles placing the parts in the proper bin. Because fifteen days were required back in Japan to assemble an order, another thirty-eight days were required for ocean shipping, and five days were needed at the PDC to bin the parts, the PDC needed to order parts at least fifty-eight days ahead of probable need to ensure uninterrupted supplies to Toyota dealers.

Toyota dealers, like Sloane Toyota, placed orders for parts once a week, by estimating likely increases or decreases in demand before the next weekly order. Because these forecasts were often wrong, they produced what Toyota calls "created demand": that is, dramatic waves of orders traveling back up the value stream that are unrelated to actual demand expressed by real customers like Bob Scott. When the weekly orders were received at the PDC, a "picker" was dispatched to collect the appropriate parts from the appropriate bins in the appropriate aisles and take them to shipping. The parts were then delivered to the dealer by common carrier truck service the next day.

Because Toyota accepted the notion that large batches, expressed as "economic order quantities," were efficient due to savings in shipping costs, and because overnight shipment of parts was expensive, it encouraged its dealers to order large amounts of each part whenever they replenished. To make this attractive, Toyota paid the freight for the large weekly batches and allowed dealers to send back up to 5 percent of the value of a weekly order for a credit if they ordered too many parts of a certain type—for example, for a special service promotion which failed to meet its target.

In the event that the dealer didn't have a part in stock—for example, a bumper for Bob Scott's pickup—a VOR or "vehicle off road" order system was able to locate and deliver the exact part needed to the dealership before noon the next day. This system looked electronically into the inventory at the nearest PDC; then, at all PDCs; and finally, at Toyota's national warehouse in Torrance, California, to find the part, print a shipping order, get the order to the pickers in the appropriate warehouse, and get it shipped. To cover the cost of this premium service, Toyota required the dealers or the customer, like Bob Scott, to pay the express freight charge for getting the parts there quickly. In this way dealers could keep large numbers of the most frequently used parts on hand while ordering special needs overnight.

At the PDC, the bins for each type of part were large and the shipping containers were larger. And the container ships were truly massive. Air freight to supply parts in the event of a shortage was very expensive, so it seemed like common sense to order large batches of a given part whenever stocks at the PDC began to run low. In addition, Toyota's scheduling computer, reaching all the way back to the factories in Japan, was programmed to anticipate certain events—the onset of winter, when more bumpers are crumpled, or sales promotions, when a large number of oil filters and spark plugs are needed in a short time as dealers offer a "special" on routine service. Extra orders were added to ensure adequate supplies for these predictable surges in demand.

By the time Toyota's warehouse network was fully in place in the early 1970s, the typical PDC had a six-month supply of the typical part. In addition, a special area of the national warehouse in Torrance housed very low-volume, rarely ordered parts, often for very old Toyotas. The months of supply on hand in this warehouse was difficult to calculate because some parts might never be ordered. Shortages still occurred, for reasons which always seemed mysterious, and some air freight across the Pacific was still necessary, but in general the system ran pretty well and permitted Toyota to achieve the highest "fill rate" (or percentage of parts available from the PDCs on demand) in the North American auto industry, at 98 percent. For fifteen years, it was "good enough."

Lean Distribution for Pull

When Toyota began to assemble cars in the United States at the Fremont, California, joint venture with General Motors (NUMMI) in 1984, it began to develop a network of suppliers for bulky and "commodity" items—tires, batteries, and seats. Then, when Toyota opened its mammoth Georgetown, Kentucky, plant in 1988, it needed a comprehensive network of suppliers for a wide variety of parts.

These same parts were needed for routine service and body-shop work at Toyota dealers; so in 1986, Toyota had opened a receiving warehouse for American-made service parts in Toledo, Ohio. This Parts Redistribution Center, or PRC, was where Shahid Khan shipped his bumpers once he started producing for Toyota.

A major mission for this facility was to reduce shipping costs per part by consolidating the less than truckload shipments of parts received from suppliers into fully loaded trucks for onward shipment to each PDC. However, this focus on low freight cost per part created a classic batch-and-queue operation in which a month's worth of parts were queued up at each supplier before shipment to the PRC. Upon arrival, the parts were queued again for quality inspection and then delayed one more time in a staging area awaiting a full truckload before shipment to each PDC.

As the yen strengthened at the end of the 1980s, and American competitors like Ford began to implement some aspects of the Toyota Production System, Toyota executives started to wonder how they could sustain their competitive advantage. In addition, Toyota's four-year replacement cycle for every model, its steadily expanding range of models on offer in the United States,[3] and the tendency of Americans to drive their cars longer and longer,[4] meant rapid growth in "active" part numbers which Toyota needed to stock as replacement parts to keep its customers happy. This seemed to require larger and larger inventories of parts and ever-growing distribution costs.

As Toyota executives pondered this situation, it occurred to them that they had never applied any of Toyota's lean thinking to their North American warehousing and distribution system. As they thought about this, it quickly became apparent that startling advantages could be gained if

The Toyota warehouses at that time were run in the familiar batch-and-queue mode we described in the Introduction and again in Chapter 3. Supervisors directed hourly workers to haul large carts or forklift loads of incoming parts from the receiving area down endless aisles for binning. The supervisors tried to ensure that the "stockers" were working hard when they were out of sight by giving each worker the same number of "lines" to stock during each shift. A "line" was a specific part number—for example, Bob Scott's deluxe chrome bumper carries Toyota corporate part number 00228-35911-13—with a varying quantity of that part, perhaps only one but sometimes hundreds.

Each "line" could therefore involve a very different amount of work. Putting one hundred spark plugs on a low shelf was a lot easier and could be completed much faster than hoisting one heavy bumper into an upper bin, yet both counted as one stocking line. Because each supervisor gave each stocker the same number of lines to complete during the shift, there were endless claims of favoritism or punitive assignments. "You're giving me all the heavy bumpers because I refused to go on the night shift when you were shorthanded," etcetera. What was more, it was practically impossible for supervisors to determine the cause when stockers failed to complete their runs in the allotted time. Was it because bins were too full to hold more parts or because of a faulty forklift, or was it simply a matter of unsupervised workers relaxing on the job? Without accurate identification of causes it was hard to implement remedies and improve practices.

The same organization and logic regulated "picking" parts for the weekly shipments to the dealers. In addition, there was an expediting system in place for the "hot list" VOR parts which were needed the next day by dealers. Unfortunately, the VOR orders often caused chaos among the pickers and slowed down the routine picks for weekly dealer orders, and it's easy to see why. A picker would be told at the last minute to run all the way across the warehouse to get a single part to meet the air freight pickup deadline. If this need had been anticipated, the pick could have been part of a complete circuit of the warehouse for many parts and would have been much more efficient.

But perhaps the worst features of the warehouse system in the late 1980s were the size of the bins, the inefficient use of storage space, and the size of the batches ordered as replenishments. Both the bins and reorder quantities were massive, involving hundreds or thousands of parts of a given type and number. This inevitably meant months of spare parts on hand and large facilities to hold all of them. Large facilities, in turn, were time-consuming for stockers and pickers to work their way around.

As Toyota executives thought about this situation, the solution to the stock-the size of the storage bins, and reduce the lot size for reorders. Instead of ordering from suppliers on a weekly or monthly basis, why not order *daily* and order *just the amount shipped to the dealers that day?* This was much more practical for domestic parts obtained from suppliers, like Bumper Works, who had mastered lean techniques and could respond to requests for small amounts. Fortunately, Toyota was rapidly transferring production of its parts from Japan to North America, and many suppliers were starting down the path pioneered by Bumper Works.

The other half of the problem, the picking, could be solved with an equally dramatic rethinking of relations with the dealers. Instead of asking dealers to order large batches weekly and then make special requests each night for missing parts, why not have the dealers order *daily* and order *just the amount sold to customers that day?*

Toyota knew that its dealers would strenuously object, unless the company offered to pay the freight for the daily shipments. Yet, a bit of analysis showed that if Toyota shipped parts from its eleven PDCs to the dealers in each of its eleven sales regions every night, the extra costs of the trucks would be offset by the simplification of the picking process, savings on inventory carrying costs, and the elimination of express delivery charges. In addition, day-to-day consistency in orders, with no sudden waves, would allow consolidation of some truck routes.

There was one last problem to solve. This was the crisis at the dealer when a customer like Bob Scott came in with a request for a part not normally carried in the dealer's parts inventory. Of course, the part could be supplied overnight by the new system, as it always had been, but the customer would be unhappy. Customers want their cars fixed *right now!*

Toyota realized that if dealers ordered every part *daily* to replace the exact number sold that day, dealer inventories of parts could be reduced dramatically. As dealers reduced their average stock of each part number, they could afford to increase the range of part numbers on hand. Instead of having hundreds of the most common parts requested and none of those requested less frequently, dealers could have a small number of each part across a very wide range. In this way they would be more likely to have a low-volume item like a bumper for an older vehicle when a customer like Bob Scott asked for one.

From Theory into Practice

The logic just described for introducing a pull system in warehousing that responds faithfully to actual customer demand was understood by Toyota's North American executives by the late 1980s. Getting it fully in place.

however, has required years, even in a supremely lean organization like Toyota, and the final steps needed are just now being taken. The translation of lean concepts into the warehouse has required considerable getting used to, for managers as well as employees, and Toyota has had to convince its employees that the new way of thinking would not cause anyone to lose his or her job.

The first step along the path, beginning in 1989, was to reduce bin sizes and to relocate parts by size and by frequency of demand. Trying to stock or pick a truck fender along with a spark plug on the same run was causing lost parts and the use of grossly oversized equipment, so it was important to segregate parts into small, medium, and large categories with their own sections of the warehouse. As this was done, those parts most frequently demanded were moved closest to the start of the sorting and picking runs and the length of the aisles was reduced markedly. The consequences of these steps for the layout of a typical PDC are shown in Figures 4.1 and 4.2. Note that a typical picking route was much shorter after downsizing and reorganization of bin locations. However, it's also important to note that because the batch size of replenishment orders was not changed, the total amount of a given part on hand remained the same. The extra stocks were stored in the "Reserves" area of the warehouse and moved to the "Active" bins as required.

The next step, beginning at the end of 1990, was to introduce the con-

FIGURE 4.1: TOYOTA PDC BEFORE LEAN THINKING

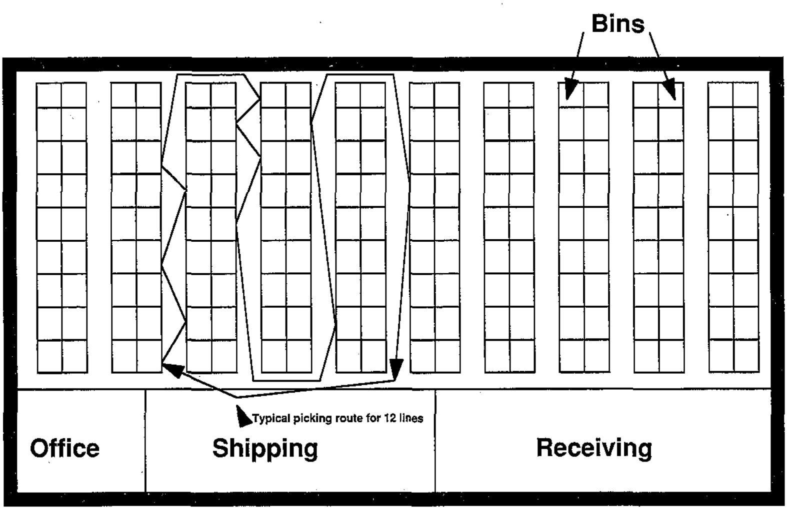
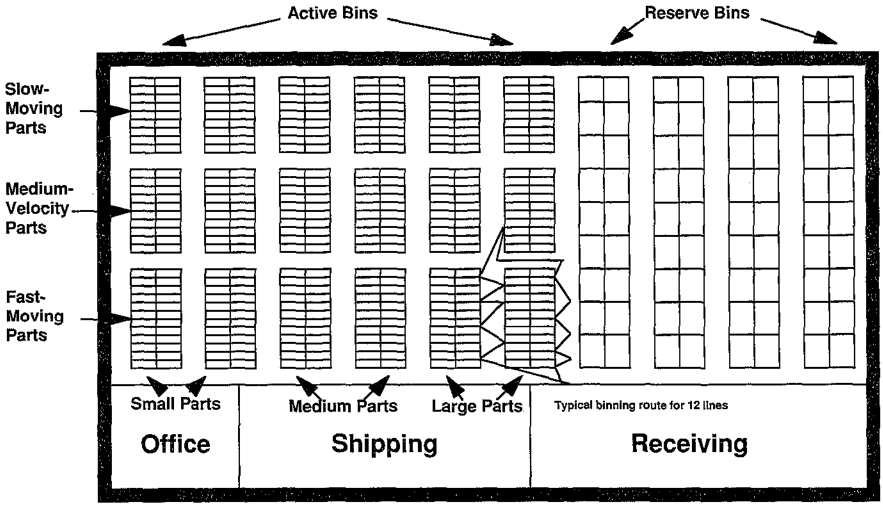


FIGURE 4.2: TOYOTA PDC AFTER DOWNSIZING



cepts of standard work and visual control by dividing the workday into twelve-minute cycles. An interval of this length was found to be the best compromise between walking distance and cart size in making a round of the bins to load or unload a cart. During each cycle an "associate," as hourly workers were now called, was expected to pick or bin a different number of lines, depending on the size of the part. For example, in a twelve-minute picking cycle an associate might pick thirty lines of small parts or twenty lines of medium parts or twelve lines of large parts.

A progress control board was constructed between the receiving dock and the shipping dock to show everyone the number of cycles to be completed and the time available. Each associate was given a stack of magnetic markers of a given color and asked to place a marker in the appropriate square on the progress control board each time a cycle was completed. This made it possible for everyone on the team to see exactly how the work was proceeding, in a striking example of visual control in a warehouse where everyone works out of contact with everyone else. The progress control board eliminated the need for "team leaders," as the supervisors were now called, to "supervise" their teams. Instead, everyone could look at the board, observe that one worker was falling behind, and provide that worker with a bit of help once other tasks were finished.

Visual control along with the use of exact work cycles also made it possible to address the causes of disruptions in work flow. The right side of the progress control board provided a blank area beside each cycle for associates to write in the reason that a cycle could not be completed on time. These reasons, when summarized, became the raw materials for directing work team *kaizen* activities when these were introduced in 1992.

One of the first *kaizen* activities was for the teams to build new work carts, using scrap materials and parts from local building supply stores, so the carts were right-sized for each type of picking or binning task. The carts were also designed to hold just the right number of parts—for example, with thirty part holding cubicles for routes for small parts—to provide another form of visual control.

At the same time the precise picking cycles were being introduced, Toyota's master computer back in Torrance was being reprogrammed to group orders from dealers by bin location in each PDC so that a set of picking labels in precise bin order was printed out at the beginning of each shift at each PDC. The picking labels were divided into twelve-minute cycles—based on the size of the parts and the knowledge of the team leader about current conditions in the PDC—and placed in pigeonholes in a dispatch box. The pickers obtained their jobs of exactly twelve minutes duration from the dispatch box, always taking labels from the next available slot so there could be no possibility of favoritism in work assignments. In this way, each associate was given five assignments per hour and the work could proceed in a smooth flow from the shelves to the shipping dock. Posting start times above the slots and visually controlling completion times also eliminated another traditional warehousing problem of working ahead to "beat the system." This practice invariably led to quality problems as associates in their haste picked the wrong part or put parts in the wrong bins.

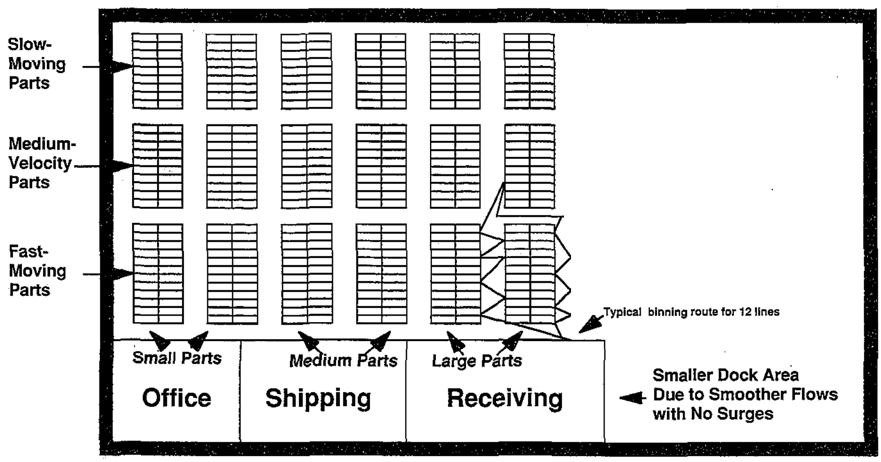
After six years of work Toyota was ready in August 1995 to transition from weekly to daily orders from its dealers and to do this without the need for an additional headcount at the PDCs. Indeed, at the end of 1995, the twenty-two pickers at the Toyota PDC near Boston were picking 5,300 lines per day while the hundred pickers at the Chrysler parts warehouse across the road were picking 9,500 lines per day using traditional methods, a productivity difference of 2.5 to 1.

When the new Toyota Daily Ordering System (TDOS) is combined with the relocation of the PRC for Japanese-sourced parts from Japan to Ontario, California, in October 1996 and the replenishment time to the PDCs from the PRCs is reduced from forty to seven days, it will be possible to dramatically reduce the stocks in the PDCs by eliminating the reserve stocks, as shown in Figure 4.3. The ability to get parts resupplied very quickly from the next level of the system, and therefore the ability to reorder in small amounts, is always the secret to reducing total inventories in a complex production and supply stream.

Technology for Lean Distribution

It is important to note that the Toyota PDCs are dramatically boosting productivity and reducing space requirements without resort to any spending for new technology. Indeed, the company has recently conducted its own test of the most appropriate technology for lean distribution by automating the Chicago PDC while converting the other ten PDCs in accord with the methods just described.

FIGURE 4.3: TOYOTA PDC AFTER DOWNSIZING, TDOS, AND RAPID REPLENISHMENT FROM PRCs



The Chicago experiment was undertaken at the end of the 1980s when Toyota back in Japan was obsessed with a shortage of workers during the Bubble Economy and was pressing ahead with much higher levels of assembly automation at its new Tahara plant near Toyota City. It seemed appropriate to try a high level of warehouse automation as well and the objective in Chicago was to completely automate the actual stocking and picking of parts.

By 1994, after much effort and enormous cost, the Chicago PDC was fully automated but productivity per employee lagged behind the other PDCs implementing standard work, visual control, and efficient bin size and location. While some direct effort was saved in Chicago, the amount of technical support needed to maintain the complex system offset the gains in direct labor and the capital costs made the whole approach uneconomic.

We'll have more to say in Chapter 10 about "appropriate" technology for a lean system and how to select it.

Level Scheduling Needs Level Selling

As Toyota thought more about installing a pull system in service parts production and distribution, another benefit emerged. If inventories and handling costs for service and crash parts could be slashed dramatically as the North American suppliers and warehouses implemented lean techniques and if production of more parts could be transferred from high-yen Japan to North America, it should be possible to offer the highest-quality and lowest-cost service and crash parts to Toyota dealers. If this were possible, special promotions to temporarily lower prices and boost sales—the bane of every distribution and production system in every industry—could be eliminated. Toyota dealers would always have the best deal for their customers.

In 1994, Toyota and its dealers together spent $32 million in the United States on direct mail, print, and broadcast advertising for "specials," offers by dealers to Toyota owners to perform anything from oil changes to complete maintenance programs at far below the "normal" price. They made these offers because the cost of "genuine" Toyota parts and dealer service was at best equal to—but often much higher—than the customer's best alternative, the independent garage or mass merchandiser. So promotions were conducted to bring in more service customers for limited periods, partly to support customer retention, partly in hopes owners could be enticed into looking at new Toyotas while at the dealer to service their current model.

The problem with promotions was very simple. They required the production of large amounts of parts in advance, yet it was never possible to predict how many would actually be needed. When not all of the parts made were actually needed, dealers shipped them back to the PDC and the PDC temporarily stopped ordering from suppliers until the excess inventory was consumed. Here we see one of the mechanisms of the familiar "pogo stick" phenomenon of "chaotic" orders coming into production facilities when the end market itself is actually quite stable, a tendency we'll examine further in a moment.

The net result was a temporary increase in Toyota orders to suppliers to a level far above long-term average demand (in order to build stocks for the promotion), followed by a dramatic drop in orders to far below long-term average demand. This was costly in both directions, requiring overtime in parts plants during the upswing and causing excess capacity during the downswing. It also created costs in the distribution channel to ship excess parts back from the dealers and for the excess stocking and picking costs of running the same parts through the warehouse system twice. The solution was to concentrate on "level selling" by keeping prices constant and making replacement parts at the exact rate parts were being sold.[5]

As Toyota executives thought about applying pull to the entire value stream, from the dealer service bay all the way back to the bumper chromer and similar "second-tier" suppliers, the more advantages they could see. But they knew it would be very hard to persuade the dealers to go along. They come from generations of batch-and-queue thinking.

The Bad Old Days of Car Service

Whenever we drive by a car dealer our first thought is always the same: "Look at all that *muda*, the vast lot of cars already made which no one wants." Similarly, when we see the large banner out in front offering "rebates" off list prices and "specials" on service and parts, we wonder, "Why did the dealer order cars and service parts which aren't needed, and why did the factory build cars and parts in advance of customer pull?"

The answer lies partly in the unresponsiveness of mass-production car makers. Chrysler in the United States is currently trying to reduce the wait for a specially ordered car from sixty-eight to sixteen days, yet for a generation, already, Toyota's lean production system has been able to build and deliver cars to order in Japan in about a week. Out of fear of losing sales to "impulse purchasers," mass producers create vast seas of cars on dealer lots, one of practically every specification, so no buyer need walk away unsatisfied. (Converting all factories to flow systems can deal with this problem, as we have already shown.)

But the answer also lies in the mentality of retailers and customers across the world. Dealers love to "deal" and the public loves a "sale." (One of us took a trip to France some years ago and discovered that the only phrase of high school French our wife could remember was "on sale"!) Changing the way retailers and consumers think about the process of ordering goods and making transactions may be difficult, but as we will see, it is essential to doing things a better way.

Pulling from the Service Bay

Most readers, we hope, have never been into the parts storage area of a car

Department at Bob Sloane's Toyota near Philadelphia in 1994, we found a rabbit warren of rickety shelves, meandering aisles, and dim lighting in two separate buildings. Clearly, the physical flow of parts was an orphan activity compared with the income-producing service bays for car repairs and the showroom where cars are sold.

When we first visited Sloane Toyota, the dealership had about a three-month supply of the average part, creating an inventory of about $580,000 in service and crash parts. When a car was driven into Sloane for repairs, it was taken to a service bay where the technician evaluated the problem and determined what parts would be needed. The technician then went to the parts window, requested the necessary parts, and waited for the counter person to go and get them somewhere in the labyrinth of bins and aisles.

Because Sloane received most parts in weekly batches, the workload of the parts stockers who took the parts from the receiving area and put them in the proper bin was very erratic. It generally took three days to get all of the parts from the receiving area into the bins, with the result that the counter person would often find empty bins when the computer showed the parts were in stock. Indeed they were, but they were "missing in action" somewhere between the receiving area and the proper bin. This knowledge touched off "treasure hunts," which are the distribution equivalent of the expediting always necessary in batch-and-queue production operations. Good counter staff would generally find their part, but the whole exercise was inherently wasteful. (The highly skilled technician, meanwhile, was standing idly at the window all the time the counter person was treasure-hunting.)

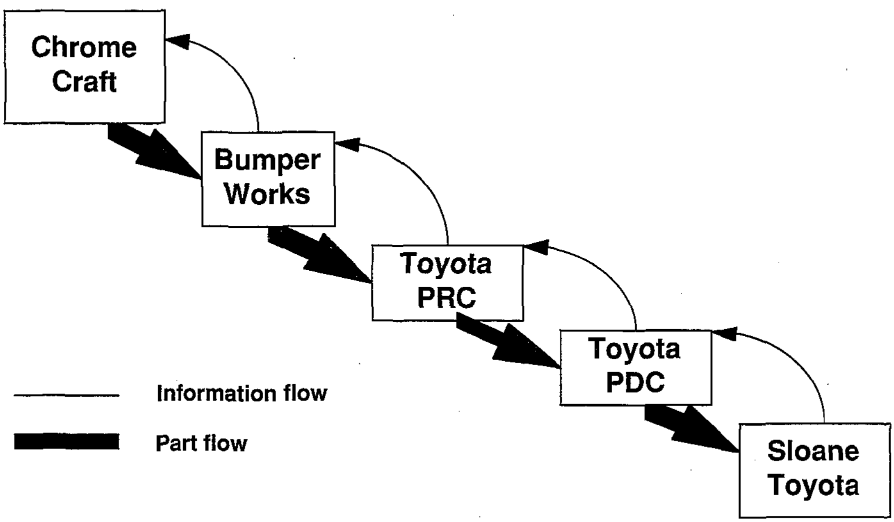
In 1995, when Sloane Toyota joined Toyota's campaign to introduce pull in the whole parts distribution and manufacturing system, it reorganized its parts storage area just the way Toyota reorganized its PDCs. By cutting the size of the bins dramatically, generally by three quarters, and reorganizing all parts storage into one building, Sloane found it possible to increase part numbers on hand by 25 percent (including Bob Scott's bumper) while cutting its storage area in half and reducing its parts inventory from $580,000 to $290,000. While freeing up $290,000 in cash from inventory Sloane was able to add four new revenue-producing service bays, created with practically no capital investment, in the empty second parts warehouse.

Sloane Toyota found that the number of cars which could receive "same-day service" went up substantially (reducing the number of cars in its overnight "loaner" fleet) even as its inventory generated cash and the number of parts the average picker could gather in a given period of time more than doubled. Most important, customers were happier because their cars were more likely to get fixed right away and the total cost of service had fallen dramatically. Indeed, Bob Scott was able to get his truck's bumper replaced

Pulling from Service Bay to Raw Materials

We can see the full magnitude of what is happening by "pulling" together all the pieces of the service value stream. By the end of 1996, when Toyota's new pull system will be in place throughout North America, the request of the customer arriving in a Toyota dealer service bay will become the trigger for pulling parts through four replenishment loops going all the way back to steel blanks, as shown in Figure 4.4.

FIGURE 4.4: PULL THROUGH FOUR LOOPS



Toyota dealers and parts suppliers will still rely on Toyota's computerized macroforecast for capacity planning to answer questions about the size of manufacturing plants and the number of warehouses that would be needed in the future. However, day-to-day part replenishment will now be handled in a radically different way: Each time a customer requests a part at the service bay, a series of replenishment loops will result eventually in more parts being made by the supplier in a situation which might be called "sell one; buy one" or "ship one; make one."

To see what this means, let's follow the bumper example all the way through the value stream. Before lean techniques were applied to any aspect of the system—that is, prior to 1989—the elapsed time from the arrival of steel blanks at Bumper Works until the bumper made from those blanks was

Bumper Works, two weeks at Chrome Craft, a few days at the Toledo PRC, six months at the PDC, and three months in Bob Sloane's parts inventory. (Lead time of this magnitude was the norm, not the exception, for the entire automotive parts industry in North America.)

By the end of 1995, the elapsed time had fallen to four months: forty-eight hours in Bumper Works and Chrome Craft, a few days in the Toledo warehouse, two months in the PDC, and one and a half months in Bob Sloane's inventory. And by the end of 1996, elapsed time should fall further to about 2.5 months as both the PDC and Bob Sloane shrink their inventories in response to falling resupply times. At the same time, the percentage of vehicles fixed the same day is increasing substantially, and costs—inventory, warehouse space, and direct labor—are falling dramatically.

Note that practically no capital equipment has been required. The tool modifications to permit quick changeovers and the specialized stocking carts in the factories and warehouses were created by production workers as part of *kaizen* activities, and the elaborate MRP systems formerly regulating activities inside the Bumper Works and Chrome Craft plants are no longer needed.

Just the Beginning

The savings we describe are just the beginning. Sloane Toyota, Toyota Motor Sales, Bumper Works, and Chrome Craft are now working on the value stream for service and crash parts as a lean enterprise under Toyota's leadership and are deeply committed to the concept of *perfection,* which we will discuss in the next chapter. They all expect to steadily reduce the elapsed time and cost of service parts. (Superlative quality is taken as a given, but quality will improve as well, as a natural complement to flow and pull.) One approach will be to extend the smooth-flowing value stream all the way to raw materials by helping the steel maker and steel fabricator overcome their current batch-and-queue thinking. At the other end of the stream, with encouragement and help from the dealer, customers may be able to schedule many of their service requirements in advance so the need for parts can be precisely predicted.

The parent Toyota company began to pursue this latter approach in Japan shortly after the 1982 merger of Toyota Motor Sales and the Toyota Motor Company, which formed the current-day Toyota Motor Corporation. Between 1982 and 1990, Toyota reorganized its service and crash parts business in a manner identical to the new North American pattern, except that it took two additional steps. It created Local Distribution Centers (LDCs) in each metropolitan area (jointly owned with the dealers) and took practically in Japan now carry only a three-day supply of forty commodity parts like windshield wiper blades. It then encouraged dealers to work intensively with every customer to preschedule maintenance so that parts needs could be precisely predicted in advance.

Because the Local Distribution Centers are only a short drive from each dealer, a "milk run" parts delivery vehicle can circulate from the LDC to every dealer every two hours, very much the way parts are sent from suppliers to lean assembly plants. And because the LDCs are large enough to stock a few of every active part, practically every car can be repaired the same day with no need for express freight from the Parts Distribution Center at the next level up the system.[6]

When the customer first schedules service for a given day, a preliminary order for the necessary parts is prepared. Then, the day before the scheduled visit, when the dealer calls the customer to be sure that the repairs will be conducted the next day, firm orders for parts are placed with the LDC for delivery on the next milk run. Finally, on the morning of the service visit, the dealer technicians examine the car to see if any additional parts will be needed and place orders for these extra parts, to be supplied in two to four hours from the LDC.

While some features of this system may work only in regions with a very high population density—for example, Japan and many areas of Western Europe—the additional gain in parts system efficiency and level of service for the customer is striking, as shown in Table 4.1.

Service parts warehouses are, of course, Type-One *muda*, being necessary to run service systems at the present time but not actually creating any value. However, as stock levels fall and replenishment orders grow smaller and more frequent, PDCs will look less and less like warehouses and more

TABLE 4.1: PARTS DISTRIBUTION EFFICIENCY AND LEVEL OF SERVICE, TOYOTA U.S.A. and JAPAN

	U.S.A. 1994 Parts/Days		U.S.A. 1996 Parts/Days		JAPAN 1990 Parts/Days	
Parts Distribution Center	50,000	120	65,000	30	60,000	18
Local Distribution Center	—	—	—	—	15,000	9
Dealer	4,000	90	6,000	21	40	3
Stock Level Index		100		33		19
Service Rate	98% in 7 days		98% in 1 day		98% in 2 hours	

Note: Toyota U.S.A. has eleven regional PDCs, serving 1,400 dealers; Toyota Japan has thirty-three regional PDCs, serving 273 DCs who in turn serve 4,700 dealers. (In the U.S.A., Toyota dealers also act as local wholesalers.) Each has on average the above days' worth of that number of parts in stock. The Stock Level Index is the total sum of the days times the part numbers in each system, with U.S.A.

like cross-docking points. Many parts on their way to a dealer will simply be moved from the incoming container to a roll-cage containing the dealer's order, without ever being binned. Instead of a series of deep lakes with little flow, the PDCs will gradually become wide spots in the channel where tributaries come together and parts are speeded to the required destination.

Perhaps at some completely lean point in the distant future it will be possible to use stereolithography and other emerging technologies to actually make parts at the dealership one by one as they are needed. However, the improvements instituted by Toyota in Japan and the United States in the past few years are available to any service business in any industry right now and constitute a remarkable leap compared with most current practice.

Is Chaos Real?

The introduction of pull in the Toyota service value stream, even to the partial extent achieved to date, raises profound questions going far beyond this particular value stream. Specifically, what happens to the "chaos" that observers have detected in many product markets when customers can pull value practically instantly from raw materials into reality? And what happens to the macroeconomy when lead times and inventories largely disappear?

Since James Gleick published his fascinating book *Chaos*[7] in 1987, it has become fashionable for business writers to talk about chaotic markets and the need for organizations to be able to instantly respond. Much of the writing on reconfigurable "virtual" corporations (whatever those are) and chaos management stems from this new perception of reality. Indeed, to apply to business MIT meteorologist Edward Lorenz's original metaphor for a chaotic system—the world's weather where the nonlinear nature of forces potentially makes it possible for a butterfly in Beijing to affect the weather a few days later in New York—managers today seem to be living in fear of butterflies.

In our view, this new way of thinking is appropriate for purely physical phenomena like the weather but miscomprehends the nature of customer-producer relations. Indeed, in looking at the great bulk of the world's industrial economy, the most striking feature of this decade is the relative stagnation and predictability of most product markets. In activities ranging from motor vehicles to aircraft to industrial machinery to personal computers to home building, the trajectory of product technology is quite predictable. What's more, the end-use demand of customers is inherently quite stable and largely for replacement. We believe that the volatility—the perceived marketplace chaos—in these industrial activities is in fact self-induced, the inevitable consequence of the long lead times and large tively flat demand and promotional activities—like specials on auto service—which producers employ in response.[8]

One solution—as recently proposed by Peter Senge[9]—is the creation of learning organizations which can reflect upon these phenomena and respond to them. One might think of a learning organization as a sort of intellectual MRP to take the kinks out of production and consumption.

We have a radically different proposal: Get rid of lead times and inventories so that demand is instantly reflected in new supply rather than the current situation of misjudged supply perennially searching for demand and creating chaos in the process. We are confident that the pattern of demand will suddenly be seen for what it is: remarkably stable except for a few new products—like multimedia—whose value and final form are being determined in real time.

Do We Really *Need* a Business Cycle?

If we can get rid of lead times and inventories to give people what they want when they want it, we believe that demand will stabilize for another reason: the damping effect on the traditional business cycle.

Conventional wisdom among economists is that about half of the downswing of economic activity in business cycles is due to consumers and producers working off the inventories built up toward the top of the cycle. Similarly, about half the upswing is due to building up new inventories in expectation of higher upstream prices ("buy raw materials now to get a bargain before prices go up") and to the expectation of greater downstream sales that require plenty of product in the distribution channel to supply, but which never quite materialize).[10] And no amount of government fine-tuning and countercyclical intervention has been able to damp the amplitude or frequency of the cycle during the fifty years since World War II.[11]

Unfortunately, our hypothesis that largely eliminating inventories will greatly damp the cycle can't be tested just yet, despite several decades of lean thinking in Japan and a decade of awareness about JIT in the United States and Europe. When one looks at the data on inventories, the amount associated with any given level of economic activity (normalizing for the business cycle) hasn't budged in America, Europe, or Japan. The reason, we believe, is that most applications of JIT, even in Japan, have involved Just-in-Time *supply*, not Just-in-Time *production*, and batch sizes have not been reduced by much. Thus, nothing has happened except to push inventories of the same magnitude one step back up the value stream toward raw materials, and one of the great prizes of the lean leap is still waiting to be claimed.

Pulling Value in Pursuit of Perfection

1.
We hope you can now see the need to precisely specify value and to identify
2. every step in the value stream for specific products, then to introduce flow, 3.
4. and next to let the ultimate customer pull value from its source. However, much of the potential of lean thinking is lost unless you take the final principle to heart. We'll end Part I of this book with some thoughts on
5. *perfection.*

CHAPTER 5

Perfection

The Incremental Path

When Joe Day, the president of Freudenberg-NOK General Partnership (FNGP) of Plymouth, Michigan, began in 1992 to introduce lean thinking in the North American alliance between the world's largest seal and gasket makers,[1] he noticed something very curious. No matter how many times his employees improved a given activity to make it leaner, they could always find more ways to remove *muda* by eliminating effort, time, space, and errors. What's more, the activity became progressively more flexible and responsive to customer pull.

For example, when Freudenberg-NOK set out to reorganize the manufacture of vibration dampers in its Ligonier, Indiana, facility, an initial *kaizen* event achieved a 56 percent increase in labor productivity and a 13 percent reduction in the amount of factory space needed. However, in revisiting this activity in five additional three-day *kaizen* events over the next three years, it was gradually possible to boost productivity by 991 percent while reducing the amount of space needed by 48 percent, as shown in Table 5.1. What's more, additional improvements are possible and planned for the future.

This seems to defy all logic. After all, there *are* diminishing returns to any type of effort, aren't there? *Kaizen* activities are not free, and perfection — meaning the complete elimination of *muda*—is surely impossible. So, shouldn't managers eventually stop efforts to improve the process and simply manage it in a steady state, avoiding variances from "normal" performance?

As we have reviewed data similar to those in Table 5.1 with senior managers in many firms around the world, we have found two prevalent reactions. One is that steady-state management—management of variances—really is the cost-effective approach once an activity has been "fixed." The other was summarized by a senior manager of an English firm, which had done nothing to fix its product development, scheduling, and production systems but

TABLE 5.1: REPEAT *KAIZENS* ON SAME PART NUMBER, FNGP LIGONIER, INDIANA, FACTORY, 1992–94

	FEBRUARY 1992*	APRIL 1992	MAY 1992	NOVEMBER 1992	JANUARY 1993	JANUARY 1994	AUGUST 1995
Number of associates	21	18	15	12	6	3	3
Pieces made per associate	55	86	112	140	225	450	600
Space utilized (square feet)	2,300	2,000	1,850	1,662	1,360	1,200	1,200

* Baseline performance before start of lean initiative on this three-shift operation with seven associates per shift.

NOTE: During this period OSHA reportable accidents and Workers' Compensation costs both declined by more than 92 percent. Total capital spending over this period was less than $1,000, for a right-sized, in-line painting system permitting single-piece flow.

time! Why didn't they conduct a thorough planning exercise to identify the perfect process at the outset so they wouldn't waste three years before finally getting it 'right'?"

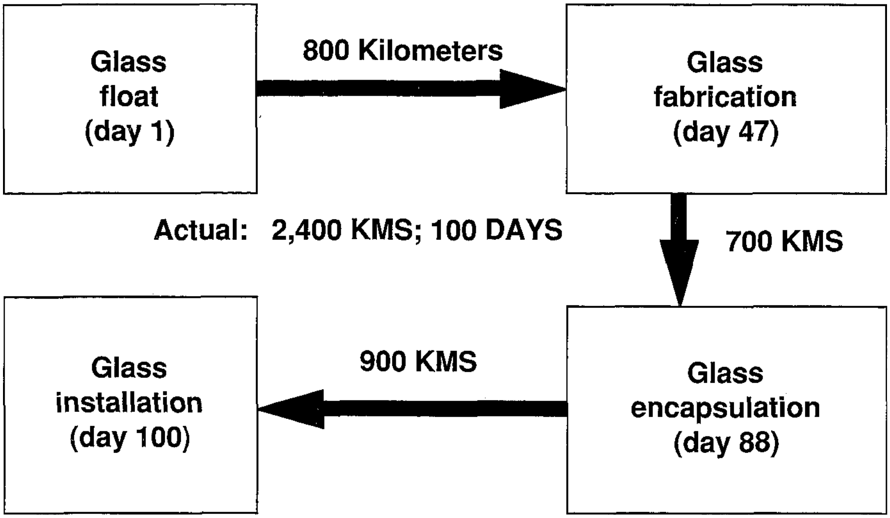
Both reactions show how traditional management fails to grasp the concept of *perfection* through endless steps, which is a fundamental principle of lean thinking. Because FNGP is one of the most relentless pursuers of perfection we have found, their approach makes an excellent illustration of what perfection means in practice and how to pursue it.

The Radical Path

There is an alternative, radical path to perfection, a total value stream *kaikaku* involving all the firms from start to finish. Glassmaking for the automotive industry provides an interesting example. Currently in North America, Japan, and Europe, manufacture of the fixed glass for cars and trucks (excluding the glass mounted in doors which moves up and down) involves very similar steps no matter which companies perform them. (These are shown in Figure 5.1.)

The first step is the glass float, a vast device in which silica is melted and floated on a reservoir of liquid tin. Sheets of glass are pulled off the float, cut into rectangular shapes, and carefully cooled. Because of the size of the typical float and the problem of getting batch-to-batch consistency, large batches are produced and stored for considerable periods before shipment

FIGURE 5.1: AUTO GLASS TODAY



The glass fabricator cuts the glass to net shapes (discarding about 25 percent in the process). The net shapes are then heated to just below the melting point and positioned in dies of the desired shape, where they are "drooped" (without any pressure) or "pressed" (using an upper die to stamp them into shape) into the final geometry needed to precisely fit the frame of the car. Again, the complexity of changing the dies and the problems of achieving batch-to-batch consistency have caused glass fabricators to manufacture enormous batches of a given part number and to store them before shipment to the glass encapsulator.

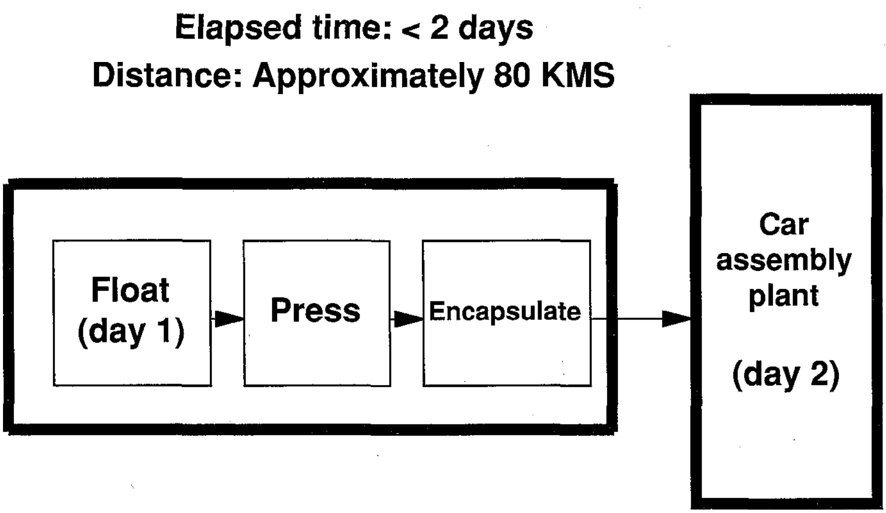
The encapsulator takes the glass from its own incoming storage and inserts each piece in a molding machine which injects some form of rubber or plastic (most commonly polyvinyl chloride) into a channel around the perimeter of the glass to create a waterproof seal and an expansion joint for attaching the glass to the steel auto body.

After some additional storage at the encapsulator, the glass is shipped to the auto assembly plant, where it is installed in the car.

Clearly, there would be substantial gains from incrementally improving each step in this process. For example, pull systems like those described in the last chapter could be introduced for each replenishment loop and tool changes could be speeded up, particularly by the glass presser, to make smaller batches. However, there would still be enormous amounts of *muda* due to the distant location of the four plants involved and the large amounts causing high levels of scrap would still be difficult to address because of the long time lags between the pressing, encapsulation, and installation steps, where problems with the previous step are most likely to be discovered.

A radical leap toward perfection in this process would involve right-sizing the glass float for the amount of product needed by a specific customer, dramatically reducing batch sizes in the pressing step and conducting it at the end of the float to save the energy required to reheat the glass, then conducting the encapsulation step in continuous flow at the next workstation from the pressing step, and finally locating this whole activity across the road from the auto assembly plant so the pull of the plant could be answered instantly (as shown in Figure 5.2).

FIGURE 5.2: AUTO GLASS AFTER RADICAL REALIGNMENT



No one has pursued this approach because, like most truly radical rethinks of a value stream, a number of firms (four in this case) would need to cooperate in changing their methods by forming a lean enterprise for this product (which might best be defined as all of the fixed glass needed for a specific auto assembly plant). However, if a lean enterprise were formed to rethink the whole value stream, additional radical reconfigurations would no doubt follow as the enterprise asked: What is the real value here for the customer and how do we create it? At a minimum, it would be necessary to rethink the proper location for product design (the auto company, the glass presser, the glass encapsulator, or some alliance of all three?) and the flow

Continuous Radical *and* Incremental Improvement

In fact, every enterprise needs both approaches to pursue perfection. Every step in a value stream can be improved in isolation to good effect. And there is rarely any ground for concern about investing to improve an activity which will soon be replaced altogether. To repeat the lesson from Chapter 3: If you are spending significant amounts of capital to improve specific activities, you are usually pursuing perfection the wrong way. Going further, most value streams can be radically improved as a whole if the right mechanisms for analysis can be put in place.

However, to effectively pursue both radical and incremental improvement, two final lean techniques are needed. First, in order to form a view in their minds of what perfection would be, value stream managers need to apply the four lean principles of value specification, value stream identification, flow, and pull. (Remember, you want to compete against perfection, not just your current competitors, so you need to be able to gauge the gap from current reality to perfection.) Then, value stream managers need to decide which forms of *muda* to attack first, by means of *policy deployment* (often called *hoshin kanri* in Japan, where these ideas originated).

The Picture of Perfection

At every step we've noted the need for managers to learn to see: to see the value stream, to see the flow of value, to see value being pulled by the customer. The final form of seeing is to bring perfection into clear view so the objective of improvement is visible and real to the whole enterprise.

We've just presented an example for glassmaking: a radical rethink of the whole value stream so that all value-creating steps are conducted immediately adjacent to the customer and exactly when needed. Toyota certainly had a picture of perfection—derived from its mastery of lean principles—when it set out in 1982 to rethink its Japanese service parts business, and then in 1989 when it began to apply the same concepts in North America. And Tesco needs a vision of perfection for the value and value stream of its beverage lines, as described in Chapter 2.

Paradoxically, no picture of perfection can be perfect. If the value stream for automotive glass could be reconfigured as we suggest, it would then be time (immediately!) to imagine a new perfection which goes even further. Perfection is like infinity. Trying to envision it (and to get there) is actually impossible, but *the effort to do so provides inspiration and direction essential to making progress along the path.* We'll return to this theme in Part III.

One of the most important things to envision is the type of product designs and operating technologies needed to take the next steps along the path. As we have seen repeatedly in the preceding chapters, one of the greatest impediments to rapid progress is the inappropriateness of most existing processing technology—and many product designs as well—to the needs of the lean enterprise. A clear sense of direction—the knowledge that products must be manufactured more flexibly in smaller volumes in continuous flow—provides critical guidance to technologists in the functions developing generic designs and tools.

In addition to forming a picture of perfection with the appropriate technologies, managers need to set a stringent timetable for steps along the path. As we will see in the examples in Part II, the greatest difference between those organizations that have done a lot and those that have accomplished little or nothing is that the high achievers set specific timetables to accomplish seemingly impossible tasks and then routinely met or exceeded them. The low achievers, by contrast, asked what would be reasonable for their current organization and disconnected value streams to accomplish, and generally defeated themselves before they ever set out.

Focusing Energy to Banish *Muda*

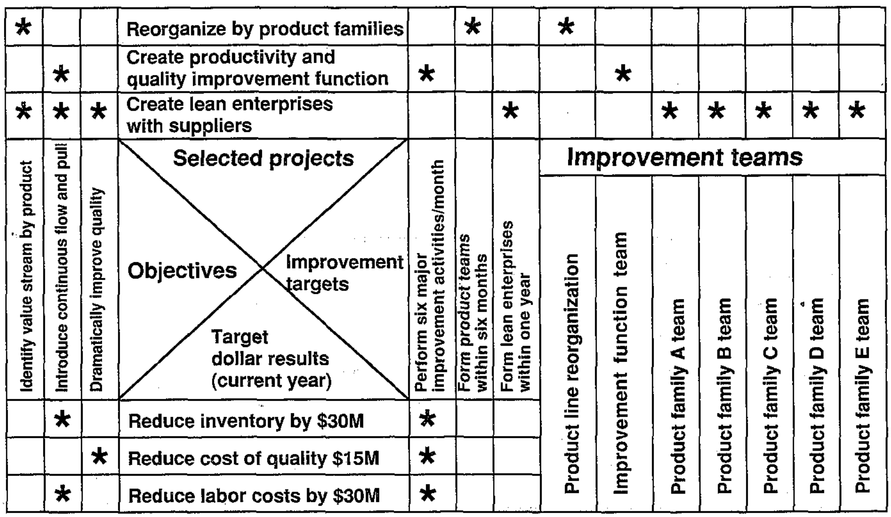
Firms which never start down the path because of a lack of vision obviously fail. Sadly, we've watched other firms set off full of vision, energy, and high hopes, but make very little progress because they went tearing off after perfection in a thousand directions and never had the resources to get very far along any path. What's needed instead is to form a vision, select the two or three most important steps to get you there, and defer the other steps until later. It's not that these will never be tackled, only that the general principle of doing one thing at a time and working on it continuously until completion applies to improvement activities with the same force as it applies to design, order-taking, and production activities.

What's critically needed is the last lean technique of *policy deployment*. The idea is for top management to agree on a few simple goals for transitioning from mass to lean, to select a few projects to achieve these goals, to designate the people and resources for getting the projects done, and, finally, to establish numerical improvement targets to be achieved by a given point in time.

For example, a firm might adopt the goal of converting the entire organization to continuous flow with all internal order management by means of a pull system. The projects required to do this might consist of: (1) reorganizing by product families, with product teams taking on many of the jobs of the traditional functions, (2) creating a "lean function" to assemble the expertise to assist the product teams in the conversion, and (3) commencing a systematic set of improvement activities to convert batches and rework into continuous flow. The targets would set numerical improvement goals and time frames for the projects—for example: Convert to dedicated product teams within six months, conduct improvement activities on six major activities each month and at least once on every activity within the first year, reduce the total amount of inventories on hand by 25 percent in the first year, reduce the number of defects escaping to customers by 50 percent in the first year, and reduce the amount of effort required to produce a given amount of each product by 20 percent in the first year.

Most organizations trying to do this find it easiest to construct an annual policy deployment matrix, as shown in Figure 5.3, which summarizes the goals, the projects for that year, and the targets for these projects so everyone in the entire organization can see them. In doing this, it's essential to openly discuss the amount of resources available in relation to the targets so that everyone agrees as the process begins that it is actually doable.

FIGURE 5.3: LEAN POLICY DEPLOYMENT MATRIX



It's also important to note that the process is top-down in the first step of setting goals but top-down/bottom-up in subsequent steps. For example, once the specific projects are agreed on, it's essential to consult with the project teams about the amount of resources and time available to ensure that the projects are realistic. The teams are collectively responsible for getting the job done and must have both the authority and resources from the outset.

As the concept of making a dramatic transition begins to take hold, we often observe that everyone in an organization wants to get involved and that the number of projects tends to multiply. This is exhilarating but is actually the danger signal that too much is being taken on. The most successful firms we've found have learned how to "deselect" projects,[2] despite the enthusiasm of parts of the organization, in order to bring the number of projects into line with the available resources. This is the critical final step before launching the lean crusade.

Smashing Inertia to Get Started

We've now reviewed the basic lean principles, the five powerful ideas in the lean tool kit needed to convert firms and value streams from a meandering morass of *muda* to fast-flowing value, defined and then pulled by the customer. However, there's a final and very serious paradox inherent in introducing thinking in real organizations to pursue perfection.

The techniques themselves and the philosophy are inherently egalitarian and open. Transparency in everything is a key principle. Policy deployment operates as an open process to align people and resources with improvement tasks. And massive and continuing amounts of problem solving are conducted by teams of employees who historically have not even talked to each other, much less treated each other as equals.

Yet the catalytic force moving firms and value streams out of the world of inward-looking batch-and-queue is generally applied by an outsider who breaks all the traditional rules, often in a moment of profound crisis. We call this individual the *change agent.*

In fact, there is no way to reconcile this paradox, no way to square the circle. The change agent is typically something of a tyrant—what one of our most thoughtful research subjects calls a "Conan the Barbarian"—hell-bent on imposing a profoundly egalitarian system in profoundly inegalitarian organizations.

Yet there are tyrants and there are tyrants. Those who succeed in creating lean systems over the long term are clearly understood by the participants in the firm and along the value stream to be promoting a set of ideas which have enormous potential for benefiting everyone. Those who fail (like many of the failed leaders of reengineering campaigns) are either identified as narrow technocrats with no concern for the very real human issues inherent in the transition, or they are dismissed by the organization as self-promoters who are simply seeking to advance their own position by riding the wave of the next "program." Both quickly fall victim to organizational lassitude, if not to active sabotage.

Because lean systems can only flourish if everyone along the value stream believes the new system being created treats everyone fairly and goes the extra mile to deal with human dilemmas, only beneficent despots can succeed. We hope that many readers of this book will take up the mantle of the change agent. And we are equally hopeful that self-promoters and cold-blooded technocrats will look elsewhere.

For those of you with the right spirit and a willingness to invest five years in gaining the full benefits, the examples in Part II are designed to show you how to succeed.

PART II

FROM THINKING TO ACTION: THE LEAN LEAP

Even once you begin to see the importance of the five lean principles, it's often hard to imagine how to install them in your own organization without a clear example of successful practice to follow, a template for action. This needs to be specific enough to show the real nuts and bolts, but broad enough to keep the big picture in view. What's more, the example needs to share enough of the characteristics of your situation that extrapolation is possible with confidence about the results.

We've therefore provided a series of examples selected from two dimensions—size and complexity, and nationality. We will begin with three American examples which progress from a small, family-owned firm with a simple product range and only a limited past to overcome, to a massive, publicly traded organization with highly complex product and process technologies, a complex supply and distribution chain, a culturally diverse, unionized workforce, and a long history to overcome of conflictual relations with its employees, customers, and suppliers.

Then we switch our focus to the three great national industrial systems by comparing the installation of lean principles in a leading German firm and in two Japanese firms of broadly varying degrees of complexity.

Your own organization is probably different from any of these in some important ways and some customization will be required. However, the examples are sufficiently broad and the results so startling that no manager can any longer claim that lean principles cannot be applied to their situation.

CHAPTER 6

The Simple Case

Pat Lancaster of Louisville, Kentucky, is a heroic American type, the standalone inventor-industrialist often found at the heart of capitalist lore. He grew up tinkering in the family workshop, convinced from an early age that he could be an inventor. After college, he tried the family business of selling packaging materials to industrial firms and then life in the product development group of a large chemical company. "But it just wasn't satisfying. From my earliest memories I wanted to be an independent inventor, manufacturer, and entrepreneur." When he was twenty-nine (in 1972), he had his big idea, a new way for manufacturers to wrap their products for shipment. He and his brother invested $300 in simple metalworking tools to build their first machine, rented a small warehouse, and went to work under the corporate name of Lantech, a contraction of Lancaster Technologies.

Lancaster's big idea was for a device to "stretch-wrap" pallets of goods (for example, the cases of cola we examined in Chapter 2) with plastic film so they could be shipped easily from plant to plant within a manufacturing system and then onward, as finished products, to the wholesaler and retailer. Traditional "shrink-wrapping" was then in wide use by manufacturers and distributors who laid plastic bags loosely around large pallet loads of goods that were then run through an oven to shrink the plastic and give a tight fit.

Stretch-wrapping, by contrast, pulled the plastic wrap tightly around the pallet load as it rotated on a turntable. As the plastic was stretched taut, it rebounded slightly to give a snug fit while eliminating the energy, equipment, effort, and time required for heat treating. In addition, stretching the wrap practically halved the amount of plastic required to secure a pallet load for shipment.

Lancaster's next invention was the key complement to his fundamental insight that the plastic should be stretched rather than shrunk. He discovered that a complex set of precision rollers (collectively termed the roll carriage) could exert a smooth force on the plastic to stretch it dramatically before it was wound around the pallet. Eventually, he found ways to decrease the amount of plastic needed to hold a pallet load together by a factor of 7.5 compared with shrink-wrapping.

When Lancaster obtained patents for his concepts at the beginning of the 1970s, they were so general and broad that he could easily fend off competitors for years. All he needed was a market. This was supplied by the world energy crisis of 1973, which unfolded just as he completed his first, hand-made stretch-wrapping machine. As energy prices zoomed, the amount of process energy and plastic (made from natural gas) which his new technique could save created an overwhelming advantage for stretch-wrapping in the contest with traditional shrink-wrapping.

Suddenly he had a real business and needed to think about how to make his product in volume. He had created his initial design and his first machine in a continuous flow of activities, so Lantech, like most start-up businesses, was born lean. However, it didn't seem plausible to run an established business this way.

In reconstructing his thought process during the transition from start-up to established firm, Lancaster recollects that "I had no production experience—remember, I was an inventor—so I decided I should get myself an experienced operations manager. What's more, I knew I would need to engineer a variety of configurations of my basic concept for different wrapping tasks, so I got an engineering manager. Finally, I had a complex product which needed explanation to the customer, so I got a sales manager. I knew instinctively about the division of labor and returns to scale, so it seemed natural that my operations, sales, and engineering managers should organize my rapidly growing firm into a series of departments, each with a specialized task, and each operating in batch mode."

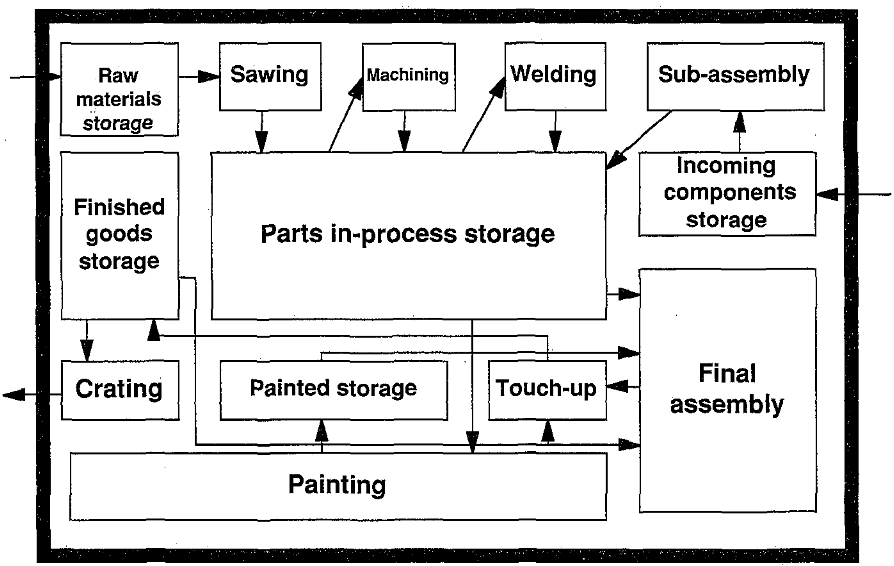
The operations manager created a series of departments in the manufacturing plant, one for each of the basic steps in building a Lantech stretch-wrapper. The Sawing Department used metal saws to fashion frame members from steel beams. The Machining Department drilled and punched holes in the steel to create attachment points for component systems. The Welding Department welded the frame members together to form the completed frame for the machine. The Painting Department applied a corrosion-inhibiting base coat and a cosmetic finish coat to the completed frame. Component systems—notably the roll carriage, the turntable, and the control module—were assembled in the Sub-Assembly Department from parts purchased from suppliers. These were attached to the frame in the Final Assembly Department.

Final Assembly was not the end of the line for products making their way from department to department and storage area to storage area. Because it was thought to be efficient, Lantech built its four basic types of machines in batches. Ten or fifteen machines of a type would be fabricated and assembled at a go. The nature of the product, however, meant that individual customers usually bought only one. Therefore, it was necessary to store many machines in a finished goods area for some time before they could be matched up with customers.

When it was time for shipment, it was often necessary to remove grime and to paint over nicks caused by moving machines from department to department. This meant a journey to a Touch-Up Department. Often the machine had to be sent back to Final Assembly as well to change its mix of optional features in order to accommodate changing customer desires. Finally, the machine was sent to the Crating Department for actual shipment.

The progression of a stretch-wrapper through Lantech is shown in Figure 6.1, often called a "Spaghetti Chart" by firms who have mastered lean thinking.

FIGURE 6.1: PHYSICAL PRODUCTION AT LANTECH



Physical production of the machine was not the only process to manage. The real complexity in volume production began to emerge as Lantech tried to move the orders gathered by the sales force (a group of about fifty independent firms distributing industrial machinery) through the office and into the plant.

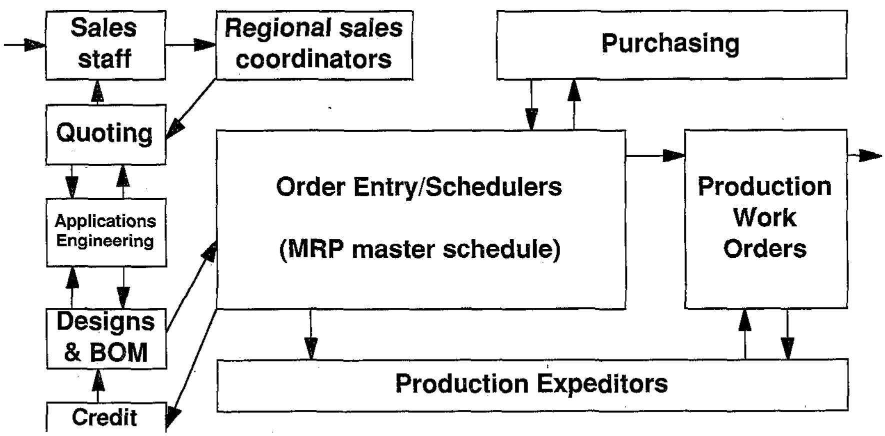
Because the machines were often customized and cost from $10,000 to

Instead the sales force contacted Lantech for authorization before quoting a price on any machine with special features. The proposal was sent to the Engineering Applications Department within Sales for cost analysis. After analysis, the "right number" was sent back to the sales force. Then, once the offer was accepted (with the distributor negotiating a final price with the customer which included the distributor's margin), the order was sent back to Lantech for production scheduling.

Upon arriving back at Lantech, the order proceeded from the Order Entry Department to the Credit Checking Department to the Engineering Applications Department (for its second visit). There, a Bill of Materials (BOM) was generated for the order. This was the precise list of every part which would be needed to manufacture a specific machine. Because every department had a waiting list of orders, there were usually delays. Typically, an order took twelve to fourteen working days to travel from the Entry Department to the Scheduling Department, while the actual processing time—what we will call "continuous flow time"—was less than two days.

The order with the BOM was then taken to the Scheduling Department inside Production Operations to work into the master schedule. Because it became apparent immediately that the flow of production through the plant would be very erratic, a separate Order Management Department was created in Sales to maintain liaison between the independent sales force and the plant on just where the machine was in the production process and to initiate expediting (using a technique we will examine in a moment) if the customer was getting restless. Information progressed through the system as shown in Figure 6.2.

FIGURE 6.2: LANTECH ORDER FLOW



The master schedule itself resided in the Scheduling Department inside Production Operations in the form of a computerized Material Requirements Planning system. The MRP melded a long-term forecast for orders with actual orders as they were received to create a daily production schedule assigning tasks to each department in the plant. Each morning, workers in each department—sawing, machining, welding, paint, sub-assemblies, final assembly, touch-up, and crating—would pick up a printout with their tasks for the day. At the end of each day, each department would report its progress back to the computerized Scheduling Department.

This system was fine in plan, but always a mess in practice because of the conflict between changing customer desires and the logic driving the production system. In order to gain scale economies, Pat Lancaster and his operations manager decided from the beginning that each department should do its work in batches: ten frames welded for the E model, then twenty frames welded for the T, then twenty-five welded for the V. This minimized the time Lantech's machinery was idle during the changeover to a new part. In addition, running long batches was thought to improve quality by minimizing opportunities to misset machines and by keeping operators focused on the operation itself rather than changeovers.

Separate departments for each production step, batches of parts run through the departments, and waiting time at the entrance to each department inherently meant long lead times. Typically, it took sixteen weeks to turn the incoming steel for the frame into a completed machine on the shipping dock. Most of this time was spent waiting as batches of parts were built in each department and then sent to storage to await the next fabrication step in the next department. The actual amount of time needed to complete the physical transformation of raw materials into a stretch-wrapper—the "continuous flow time"—was only three days.

Long lead times meant in turn that the sales force distributing Lantech's machines to the end user tried to figure out how to beat the system. A favorite approach was to order machines on speculation and then, as a real customer was found, to alter the options requested (or even the base model) very late in the production process. This tactic created the need either to rework the machine initially ordered or to slip the delivery date and build a properly configured machine from scratch.

Soon the factory was being pulled in opposite directions by two conflicting planning systems—the master schedule worked out by the Scheduling Department based mostly on sales forecasts, and the ever-changing demands from the Sales Group intent on pleasing actual customers.

These latter demands were met by a team of expediters moving through the plant with a "hot list." These were orders which were either long overdue for shipment or in which the sale would be lost if the product was not reconfigured to the new specification. The expediters visited departments in sequence and ordered the workforce to make just one item of a batch—a "partial"—so they could take that part immediately to the next department and move it to the head of the line in that department. In an extreme situation, when Pat Lancaster agreed that an order absolutely had to be expedited all the way through the company, it was possible to get a machine built in less than four weeks. However, when this was done, the schedule of every other machine in the plant slipped, creating the need for more expediting.

This system of order-taking and production sounds chaotic—and it was. But it was and is the standard method in most of the industrial world for making products when there is considerable product variety, long lead times, and a complex production process. To make matters worse, the production and sales technique of batch-and-queue soon had an exact analog in product development in Lantech's departmentalized engineering process.

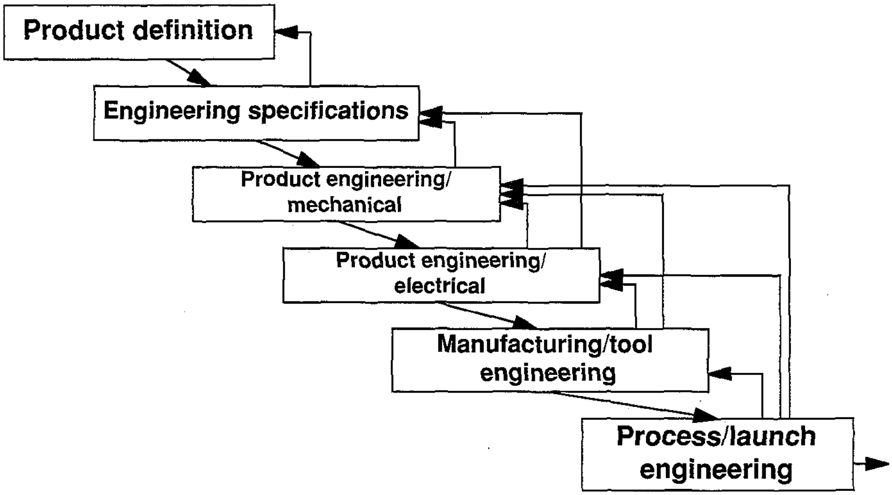
To create a new design, it was necessary for the marketing staff, engineers skilled in several specialties, the purchasing staff, and operations planners to work together. The marketing group determined what the customer wanted. ("A machine able to wrap forty four-thousand-pound pallet loads per hour in a fifteen-by-fifteen-foot work area at a cost of fifty cents per pallet.") The chief engineer then translated these desires into engineering specifications. ("A turntable able to support a four-thousand-pound pallet load, a turntable motor of *x* horsepower capable of *y* rotation speed, a control system able to direct the wrapping procedure automatically, etcetera.")

Next, a mechanical engineer designed the moving mechanical parts, notably the roll carriage and the turntable. Another mechanical engineer then designed the frame and an electrical engineer designed the control system to meet the engineering specification. The manufacturing engineer then designed the fabrication tools. Once the product design and tools were finalized, an industrial engineer from the Production Department figured out how to get the product to progress by steps through the plant.

The Engineering Department was initially quite small, with only a half dozen engineers, but even then the communication barriers between the one-person "departments" were substantial as the design was moved from marketing group to chief engineer to mechanical engineer to electrical engineer to industrial engineer. A considerable amount of rework and backtracking was required to get from the initial concept to a complete, production-ready design. (The prime cause of the backtracking was that the design didn't fit the needs of the next specialist in the line—"there's not enough room for my control panel," etcetera—and was sent back for modification. A frequently employed alternative to sending the design back was to secretly redesign it.) As Lantech grew and more engineers were added, these communication problems got worse.

What was more, each engineer typically had a stack of projects on her or his desk, so that expediters soon appeared in engineering as well as in the plant to get "rush" projects through the system. In practice, it typically took a year to introduce a minor improvement in a family of machines and three or four years to introduce a new family suited to a different task, such as wrapping small bundles. The "continuous flow time," by contrast, was only a few weeks for minor improvements and six months for a new family of machines. The progression of a design through the design and engineering system is shown in Figure 6.3.

FIGURE 6.3: LANTECH PRODUCT DEVELOPMENT SYSTEM



The three major activities undertaken in Pat Lancaster's new company—development of new designs, management of information on what to make, and physical production of the machines—were all conducted in a classic batch-and-queue manner. And they were conducted with great success.

Looking back, Pat Lancaster summarized his dream of becoming a highly successful inventor, manufacturer, and entrepreneur. "After 1973, we were selling a top-priced product which had major performance advantages over competitor products due to my patent position. Over the next fifteen years Lantech grew to 266 employees and $43 million in sales. We could and did deliver late because of conflicting demands for efficiency versus speed within the production process. We offered so-so quality in terms of manufacturing defects in machines delivered to customers. We took more than a year to develop 'new' machines which differed only in very minor ways from previous models. But we were way ahead of the competition and we made tons of money. For fifteen years my dream came true."

Then, on June 26, 1989, Lantech lost a patent infringement suit against a competitor offering lower-priced clones of Lantech machines. (The suit concerned a new generation of patents Lantech had obtained in the mid-1980s as follow-ons to its original patents obtained in the early 1970s.) This threw open the market to every packaging machinery firm. "By the end of 1989, clones with roughly comparable performance started to appear everywhere and the bottom fell out of my pricing. I was still turning a small profit but I knew worse was coming as soon as the business cycle turned down. In my heart I knew that Lantech was 'walking dead.' "

Pat Lancaster is by nature a highly dynamic individual. So he had plenty of ideas on what to do. In fact, he tried many of the remedies popular in the American business community at that time. His first approach was to reorganize the firm into profit centers for "standard products" and "specials" (those requiring extensive customization). This was to increase accountability and to move the highly customized products out of the path of easier-to-make "mass-production" machines. Then, as sales flattened, he considered laying off employees and shrinking Lantech—what we now call "downsizing." However, Lancaster was convinced that no firm had ever been saved by cost cutting and retrenchment alone.

He needed a new way to think about his business and sought it in the Total Quality Management (TQM) movement. After a visit to Milliken, the South Carolina textile giant, he came back to Louisville with plans for putting the voice of the customer first and foremost. The old "good enough" standard for delivered defects and customer service was quickly replaced with talk about perfection. TQM

Over the next few years this focus was supplemented with a process of "value-driven culture change" to create an empowered organization, build trust, and knock down departmental barriers. The original senior management team, which had been composed of hierarchical personalities accustomed to a top-down, command-control style, was replaced by a new group of managers willing to work in a team-based organization. (Lancaster is the only senior manager remaining from the 1970s.) In addition, extensive training was conducted in team processes, team leadership, and individual interaction.

These programs were an essential start, but they lacked a direct link to Lantech's core activities. As Bob Underwood, a longtime production worker, put it in retrospect: "We learned to respect each other and wanted to work together in teams, but we were all revved up with nowhere to go." The factory was still a mess. Product development was still too slow. The sales force was still playing games to beat the lead-time problem.

Max-Flex

The third approach to the crisis was a new production method called "Max-Flex." The idea was to dramatically reduce lead times by building inventories of major components—machine frames, roll carriages, turntables, control modules—far in advance and then mixing and matching the components to build complete machines to customer specification very quickly once orders were confirmed. The objective was to overcome Lantech's pricing disadvantage by promising more rapid delivery of machines with customer-specified features.

On one level the performance of the new Max-Flex concept was impressive—lead times fell from sixteen weeks to four. But the costs were enormous. Engineering change orders were frequent in Lantech's business now that it had become highly competitive. These changes were both to add product features to keep up with the competition and to rectify defects discovered in service. Therefore, it was often necessary to work backwards, "retrofitting" changes into the mountain of components built in advance. Obviously, the cost of carrying this mountain of "just in case" components was substantial, and Lantech began looking for a new warehouse to store components as storage space in its plant was exhausted. But most exasperating, despite Lantech's best efforts at planning production, cases quickly arose where one critical component needed to complete a machine was lacking. (Taiichi Ohno noted long ago that the more inventory you have, the less likely you are to have the one part you actually need.) The solution was a new team of expediters to move the missing component through the production system.

Yet a fourth approach to the crisis was better technology. A new scheduling system, based on the next generation of MRP, was installed in 1990. It permitted every worker to have direct access to the status of every machine in production and to input their own data as they moved a part or a whole machine ahead. This permitted every worker to get work orders from a terminal at his workstation and, in theory, to feel full "control" over his activities. (As Pat Lancaster noted: "It seemed to be a wonderful marriage of technology and democracy. Everyone could look into the computer to see what was going on all over the plant and get their work orders immediately. Our slogan was 'Data to the people.' ")

The new system required a new computer, a new Management Information System Department with four people on the day shift and three more on the night shift to keep all of the data current, and direct inputting of every work task by workers on the plant floor as they completed it. As Jose

Zabaneh, Lantech's manufacturing director, noted, "Pretty soon workers were fully in 'control,' yet the system was wildly inaccurate because many items simply never got entered and there was no means of catching errors. The old MRP system was slow but 99 percent accurate. Our new 'democratic' MRP system was a complete catastrophe; instead of information we had given *muda* to the people." To compound the situation, the magnitude of inputs and changes was causing the computer to run very slowly. Lantech's information technology consultant recommended that the best solution would be a much more powerful and expensive computer.

By the end of 1991, Lantech's orders began to fall for the first time, despite price reductions, and the factory was finding it nearly impossible to respond to continuous shifts in demand. As Pat Lancaster summarized the situation later, "We began losing money for the first time and our fundamental ideas on how to run the business were in a meltdown." Then he discovered lean thinking.

The Lean Revolution

Ron Hicks does not look like a revolutionary. He looks like an accountant (although he was trained as an industrial engineer) and talks in dispassionate tones. But he brought a revolution when he came to work at Lantech as vice president of operations in March of 1992.

He had learned how to be a revolutionary while working at the Danaher Corporation, a collection of fifteen manufacturing companies collected by Steve and Mitchell Rales in the 1980s. Quite improbably, these two youthful entrepreneurs from Washington, D.C., had become acquainted with the lean concepts pioneered by Taiichi Ohno, and the firm had convinced some of Ohno's Japanese disciples to establish operations in the United States in 1987 to support Danaher's conversion effort. They grasped that lean thinking could revolutionize their firms, which had initially been bought because they were attractively priced, as part of their effort to diversify out of their core real estate business. One of these firms was Hennessy Industries of Nashville, Tennessee, a manufacturer of automotive repair tools and garage lifts. Ron Hicks was working there as vice president of operations.

Ron Hicks remembers the day in 1989 when "the light went on." "I went to visit the Jacobs Brake Company in Bloomfield, Connecticut, another Danaher company, and discovered they had followed Ohno's advice by completely eliminating their traditional production departments. They had installed work cells in which all of their machines were realigned into the actual processing sequence needed to make specific product families of truck engine components. Each part was then manufactured in a continuous flow with absolutely no buffer stocks between steps using a concept they called 'single-piece flow.'

"What really amazed me was that on the day of my visit they were conducting an improvement exercise and had decided that the work flow for a particular item would be much smoother if they moved a massive machine from one position to another. They decided to do it early in the morning, got the moving team together almost instantly, moved the machine, and were back in production in a few hours.

"In my fourteen years as an operations manager at the General Electric Company, where I worked before moving to Hennessy, it would have taken an act of Congress to move such a large machine. But these guys just did it, and it worked. I suddenly realized I was living in a different world."

By March 1992, when Hicks received a phone call from Pat Lancaster, he had transformed himself from a "concrete head" into a lean thinker and was ready for a new challenge. Lancaster had screened hundreds of applicants in his search for a new operations vice president and was sure that Ron Hicks had the ability to transform a manufacturing operation. The question was exactly how and how fast.

In the newly empowered spirit of Lantech, Ron was invited to Louisville and interviewed by those he would manage. His simple proposal came as a revelation: Lantech would immediately form teams to rethink the value stream and flow of value for every product in the plant, and then every step in order-taking and product development. Lantech would line up the essential activities required to design, order, and manufacture a stretch-wrapping machine and perform them in sequence, one machine, one design, one order at a time. Batches, queues, backflows, and waste—*muda* of all sorts—would be banished. The *value stream*—the irreducible minimum set of activities needed to design, order, and make a stretch-wrapper—would flow smoothly, continuously, and rapidly.

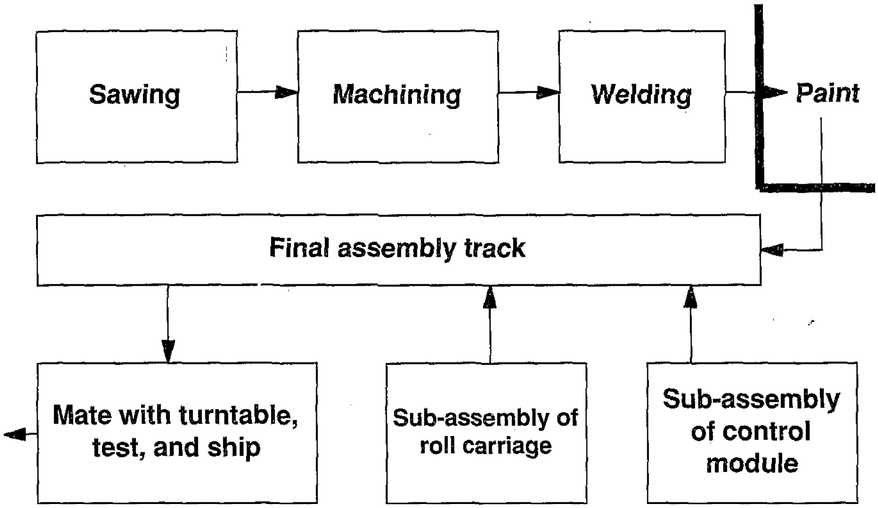
Ron Hicks was hired and immediately went to work with a simple plan: Disaggregate the four basic types of machines flowing through Lantech's departmentalized, batch-and-queue production system; eliminate all of the production departments; create a production cell—four in total—for each type of machine; and then line up all of the activities required to make each machine within a cell and perform them in a continuous flow. This was the *kaikaku* phase in the Lantech plant, the time to completely tear things apart and recombine them in a totally different way.

Zellen pro Produkt

The T/V model, which was soon replaced by the new Q model, was the acid test. A team of Lantech's best workers was selected to rethink the flow and quickly, in only one week, devise and put into production the plan—shown in Figure 6.4.

The sawing operation was located immediately adjacent to the machining

FIGURE 6.4: FLOW OF Q LINE



operation, which was only a few steps from the welding operation. Although it was still necessary for all four models to share a massive, centralized paint booth, continuous flow picked up again with sub-assembly and final assembly. Testing and crating were placed at the end of the line and conducted by the work team. Even though only eight machines were made each day—one per hour—an imperceptibly moving track was installed in final assembly as a pacing device.

Each morning the saw operator would start production of a new machine on the hour. A kit of all the frame parts required for the machine was prepared by the saw operator by the end of the hour and rolled about three feet to the machining station. From there it would proceed about four feet to welding. Fourteen hours later—about half of this due to the curing time required by the paint booth—a completed machine was ready for shipping.

To make this simple system work, Lantech had to change a generation of industrial thinking about how to do work and how to work together. Because all of the jobs were directly linked, with no buffers, it was necessary that everyone think about *standard work*, which is to say the best way to get the job done in the amount of time available and how to get the job done right the first time, every time. (By design, either the whole cell is working or nothing is working.) Every step of every job was soon charted by the work team and posted for everyone to see.

Similarly, because in this new system machines are only made when ordered—remember that production lead time has fallen from sixteen *weeks* to fourteen *hours* so there is no need to build machines ahead of time on speculation in order to permit rapid deliveries—it was important to introduce the concept of *takt* time. This is the number of machines to be made each day to meet the orders in hand divided into the number of hours in the day. (With production at eight machines per eight-hour day, *takt* time is one hour.) The important point about *takt* time was that when orders did not require the full utilization of equipment and workers, *takt* time was increased. The machinery was slowed down and each of the multiskilled workers in the Q cell performed several of the jobs in the cell while excess workers were put on other tasks around Lantech. This reversed the age-old tendency to work ahead and build inventories if no orders were immediately on hand.

takt-time

Two other concepts were needed as well. Lantech had to *right-size* many of its tools and devise a number of new tools so that smaller saws and machining tools could be fitted in the work cells. (As it turned out, the excess workers freed up by rethinking production flow were able to make most of the tools needed.) Finally, Lantech had to learn how to perform *quick changeovers* on all of its tools so it could make all of the parts for each machine and a variety of product options for successive machines with very little downtime.

right-size tools

quick changeover

When the new cell concept was proposed, many of the production workers were baffled or dismayed. As Bob Underwood, one of the most skilled workers on the floor, noted: "We were used to a system in which each of us had a set of hard-earned skills—welding, machining, and, in my case, the ability to adjust nonconforming parts so they would fit. We were used to doing our own work as we saw fit at our own pace in our own department. As long as we met our daily production quota we were left alone. What's more, the real kick in the work was 'fire fighting,' in which the Lantech Volunteer Fire Department went into crisis mode to get an emergency order through the system or eliminate a sudden production bottleneck. I was one of the best fire fighters at Lantech and I loved it."

Ron Hicks was proposing a new system of standard work and *takt* time that sounded like oversight by the industrial engineer, which every skilled tradesman hates. (The difference, of course, was that the work team would standardize its own work.) What was more, he was proposing making complete machines one at a time. Finally, he claimed that if the work was standardized by the work team, the machines were realigned to permit single-piece flow, and *takt* time was adhered to with no working ahead, there would be no more fires to fight. As Underwood remembers, "It didn't sound like much fun and I thought it would never work."

When the conversion week was completed and the new cell was ready to go . . . it didn't work. All kinds of problems, long submerged in Lantech's massive inventories and closely held work practices, suddenly emerged. Some steps had not been included in the standard work charts. Poor tool maintenance—easily tolerated in the old batch system—repeatedly stopped the whole cell. The supply of components to the cell was not dependable. The widespread feeling was that Ron Hicks was pushing a novel concept that would never work at Lantech.

At this point Jose Zabaneh, the production manager, played the key role: "I was so fed up with our failures and so taken with the logic of the new system that I threw my heart into it. I called a meeting of the workforce and announced that I would stay all night and all weekend to work hands-on on fixes to the problems we were encountering with the new cell, but that I would not spend one second discussing the possibility of going back to the old batch-and-queue system."

Pat Lancaster gave unfaltering support to the new system, Ron Hicks (along with his consultant, Anand Sharma, who had advised him earlier on the conversion of Hennessey) had the technical skills to work the bugs out, and Jose Zabaneh was "our spark plug." Gradually, it all began to come together.

(We'll see that these three attributes—taking the long view, technical virtuosity, and a passionate will to succeed—are essential for any organization making the lean transition. Sometimes they are possessed by a single individual, sometimes, as at Lantech, they are shared by a group of leaders. However initially distributed, they are all necessary and eventually they must be shared by the whole organization.)

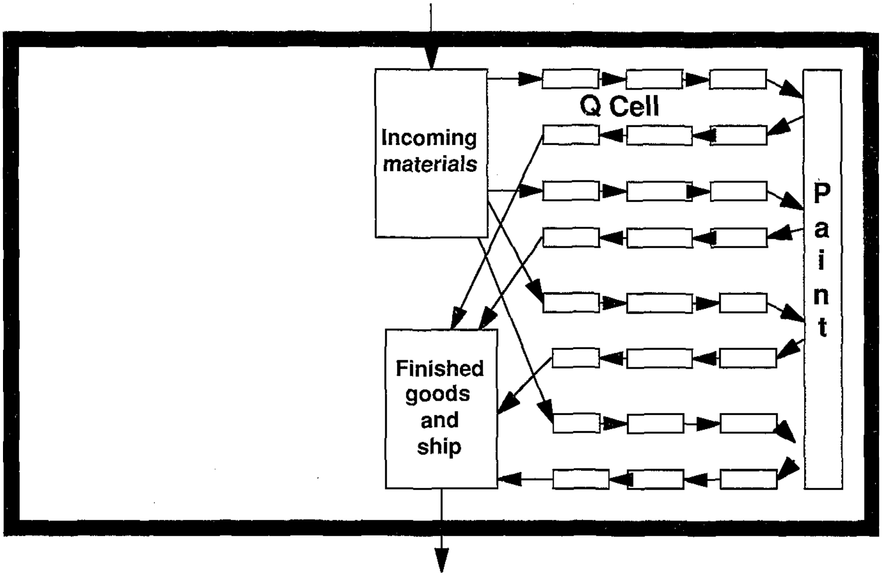
By the fall of 1992, the whole Lantech production system had been converted from batches to single-piece flow, including the cell for the largest Lantech machine—the $50,000 H model—made at the rate of only one per week. The plant now looked as shown in Figure 6.5. results:

The *consequences for performance* were truly staggering. Although Lantech's headcount stayed constant at three hundred, the number of machines shipped doubled between 1991 and 1995. (The sales growth was due to a general recovery in the market, aggressive pricing by Lantech to capture share, and a host of new products, to be described in a moment.) The plant, which had been bulging at the seams with inventory, now had 30 percent excess space despite the doubling of output. The number of defects reported by customers fell from 8 per machine in 1991 to .8 per machine in 1995. Production throughput time, as we have noted, fell from sixteen weeks to fourteen hours. The percentage of machines shipped on the date agreed with the customer went from 20 to 90 percent.

production ↑
excess space
defects ↓

To speed this remarkable transition, Pat Lancaster made two promises to his workforce. These seemed almost quixotic in 1992, given the financial

FIGURE 6.5: NEW LANTECH PRODUCTION FLOW



condition of the firm, but have proved critical to success. First, he promised that no one would be let go because of the lean conversion. Instead, a *kaizen* team was created from freed-up workers who were deployed to plan the improvement of other activities. Bob Underwood, the original skeptic and chief "fireman," was made head of this team. After every improvement, the best (not the worst) workers in the revamped process are transferred to the *kaizen* team, making clear that this is a promotion, not a punishment. The steady growth in output of the newly competitive Lantech has meant that within a short period these workers have been needed again for production work.

At the same time, Lancaster reviewed Lantech's wage policy and adjusted the base wage upward from about $7.00 to about $8.50 per hour. As Ron Hicks noted, "We had been running unskilled employees through like McDonald's, with a sharp premium for our small core of skilled workers. It quickly became apparent that all workers in the new Lantech would be skilled workers, but with a very different type of skills. So we had to pay all of them a better wage. As a result, turnover quickly fell to just about zero." (Note that because each machine is now being made with one half of the formerly needed hours of human effort, a 25 percent wage increase is easily affordable.)

• • •

As the lean revolution gained momentum in the plant, it was time to turn to the office and in particular the order-taking process. As Pat Lancaster put it, "We wanted the goodness of the plant to suck the badness out of the office. If we could make a machine in fourteen hours, how could we live with an order-taking process which required three weeks?" In one notable case, Lantech made and delivered a machine in four days—long before the credit check could be completed—only to discover that the buyer was insolvent.

The technique employed to transform the office was exactly the same. Lantech set up a *kaizen* team to collectively rethink the process. It included all the workers involved in a specific process, the firm's technical experts—including production workers from the plant *kaizen* team and one outside consultant (Sharma). The group mapped the entire value flow and looked for wasted time and effort. As each process was rethought and turned from batch-and-queue into flow, the best of the workforce was assigned to the *kaizen* team to lay the groundwork for the next process review. No one was laid off and the move to the *kaizen* team was clearly an acknowledgment of superior performance.

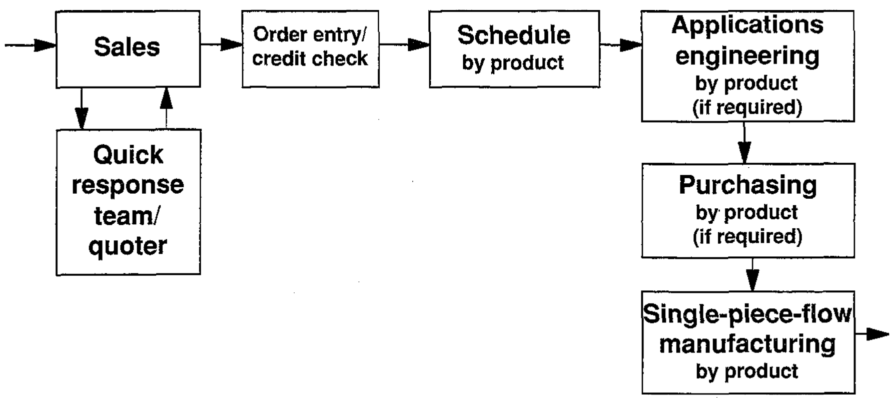
When these techniques were applied to the entire order-taking and plant-scheduling system, the results were truly astounding. Because Lantech now understood its costs much better, it was possible to publish fixed prices on all but truly custom-built machines and to eliminate the haggling step between Lantech and the distributor. The order itself, once at Lantech, could be inserted into the production schedule in only two days.

Perhaps most remarkably, most of the computerized scheduling system was no longer needed. MRP was retained for long-term materials ordering from suppliers, but day-to-day scheduling is now run off a large white board in the sales office. The production day is divided into slots by *takt* times and orders are written on the board as they are confirmed. At the times we have visited Lantech, the white board slots have been filled anywhere from three days to two weeks ahead of the current date and no machine will be made except in response to a confirmed order.

A large white board easily visible to everyone in the firm has proved to be a remarkable spur to the sales force, particularly as the amount of filled space gets smaller and the amount of empty space gets larger. This is an excellent example of yet another lean technique, *visual control*, in which the status of an activity is displayed so every employee can see it and take appropriate action.

The final step in this process is to copy down each evening the roster of machines to be made the next day and to take this list to the four production cells. For each machine, the cell is given the name of the actual customer and the promised delivery date, typically two days from the start of the build sequence for high-volume machines and ten days for the lowest-volume, large machine. The former Management Information System Department with its seven full-time workers has been eliminated because the parts within the plant are pulled along to the next workstation automatically. Information flows that had been automated have now been completely eliminated because product and information have been combined into one. The full results, as shown in Figure 6.6, can be contrasted with the labyrinthine order process shown in Figure 6.2.

FIGURE 6.6: NEW LANTECH ORDER FLOW



The main transitional problem has been that the distributors and buyers of industrial equipment are unaccustomed to getting rapid and on-schedule deliveries. Orders have often been "guesstimated" on the presumption that many weeks were available to firm up the precise specification, notify the manufacturer of changes, and plan for the machine's installation. In one notable case, Lantech made and delivered a machine within one week of the order, just as promised, to find the customer quite upset: "You've sent us our machine before we've given any thought to how to use it. We thought we were placing an order simply to guarantee ourselves a place in the production line, that we would have time to respecify the options, and that you would deliver late as usual. Now, you've gone and made it already!"

The final step in transforming Lantech has been to rethink the product development process. Pat Lancaster knew from the early days of the plant conversion that he would need to grow his business dramatically in order to keep everyone busy, as he promised, while productivity zoomed. This meant turning strategic thinking on its head: "I didn't have time to find a brandnew business to go into and I didn't have the money to buy out any of my major competitors. Instead, I needed to revitalize and expand my product range so I could sell more in an established market I knew well. At the same time, I knew that a total redesign of my products to make manufacturability a key consideration would slash my costs even further and dramatically improve quality and flexibility for the customer."

He also knew that his batch-and-queue product development system would take years to come up with market-expanding products if not given the same treatment as the plant and the office. He wanted to put new product designs into single-piece flow, just like orders and machines. "We needed the design to move continuously from the initial concept to the launch of production. This meant no stopping due to the bureaucratic needs of our organization, no backflows to correct mistakes, and no hitches during production ramp-up."

Lantech had experimented with development teams in the late 1980s and early 1990s but without much success. A few "bet the company" projects were pushed through by a designated "dictator" who was effectively a new type of expediter (slowing down all other projects to get his project through). Otherwise, weak "team leaders" tried to coordinate the activities of the numerous technical specialists needed to develop a complete product, each with their own priority list. In no case was the team leader—dictator or weak coordinator—responsible for the end results of the project: Did the product please the customer and make money for Lantech during its production life? No one was really in charge and not much had changed despite the new "team" terminology.

In 1993, Lantech went to a new system of dedicated teams led by a Directly Responsible Individual (DRI) clearly charged with the success of the product during its lifetime. The corporate annual planning process identified the major projects to be developed that year and ranked them. A team of dedicated specialists was designated for the two top-ranking projects. This consisted of marketing, mechanical engineering, electrical engineering, manufacturing engineering, purchasing, and production (including hourly workers from the plant *kaizen* team who would actually build the machine once launched). These teams were co-located and told to work nonstop on the designated project and to do nothing else until it was done. The welter of minor projects which formerly cluttered up the Engineering Department were simply dropped (or "deselected" in Lantech-speak). As the engineering director noted, "We never would have finished them anyway!"

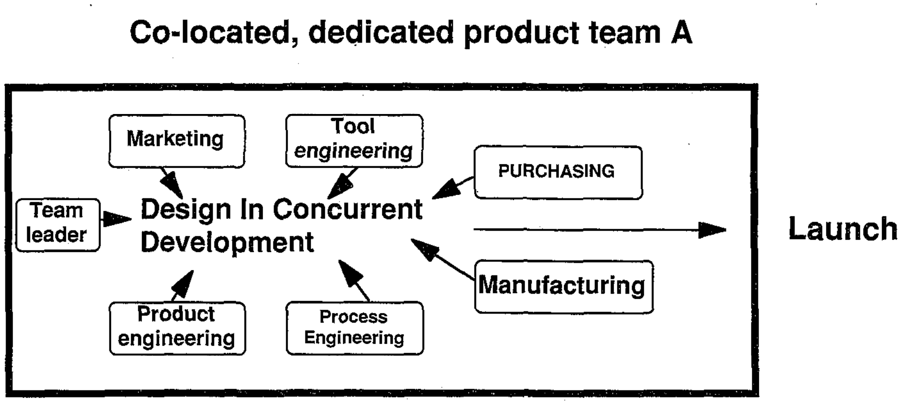
A *kaizen* of Lantech's prototyping process showed that if all of the needed skills were available, a working prototype for the top-ranked project could be put together in a week, a process which formerly would have taken three months. And the presence of the actual production staff on the team quickly identified manufacturing problems which the mechanical and electrical engineers had never imagined.

The major objections to dedicated teams—that work flow is uneven, so some team members will be underutilized some of the time and teams will be in conflict for scarce skills needed at specific points in development—were overcome in two ways. First, it developed that team members actually had much broader skills than they had ever been asked to use. (After all, they had been reengineering each other's designs in secret for years!) They could quickly develop additional, narrow skills to address specific problems. Mechanical engineers could actually help manufacturing engineers with their work and the reverse. This meant that the uneven work flow problem could be largely corrected within the team.

Second, it turned out that a bit of careful scheduling could identify conflicts in requirements for skilled personnel far ahead of time. Moving a few specialists from one team to the other and back, as needed, could solve the problem.

Under the new Lantech product development system, the progression of the design looks as shown in Figure 6.7, in contrast with the maze shown in Figure 6.3:

FIGURE 6.7: NEW LANTECH PRODUCT DESIGN FLOW



The first product to come through the new system showed its dramatic potential. The new S series, launched in mid-1994, was developed in one year (compared with four years for its predecessor) with about half the effort previously thought to be required. (Remember: There were no delays for lack of personnel or queue time, no backtracking, and no secret rework.) Then, the launch was much smoother than in the past and the number of defects reported by customers was a tiny fraction of the rate experienced with previous new products.

The Bottom Line

The conversion of Lantech from classic batch-and-queue to lean thinking has produced a dramatic box score of performance improvements (see Table 6.1).

TABLE 6.1: THE LEAN TRANSFORMATION AT LANTECH

	BATCH-AND-QUEUE/1991	FLOW/1995
Development time for a New Product Family	3–4 years	1 year
Employee hours per machine	160	80
Manufacturing space per machine	100 square feet	55 square feet
Delivered defects per machine	8	.8
Dollar value of in-process and finished goods inventory*	$2.6 million	$1.9 million
Production throughput time	16 weeks	14 hours–5 days
Product delivery lead time†	4–20 weeks	1–4 weeks

* Note that sales doubled during this period. If Lantech's traditional sales-to-inventory ratio had held constant, $5.2 million in inventory would have been needed to support 1995 sales volume.

† Product delivery lead time is the period customers must wait before their product can be delivered. In 1991, most of this time was in-process time in the production system. In 1995 most of this time was wait time for a production slot as Lantech's sales zoomed.

However, the result by which any business in a market economy must be measured is the ability to make enough profit to renew itself. If the transition at Lantech cost a fortune in new investment or disrupted the firm's ability to satisfy customers it would be an interesting technical exercise rather than a revolution in business practice.

In fact, the amount of investment required was substantially zero. The tools were moved and reconfigured, for the most part, by workers freed up from inefficient production tasks. The reconfiguration of the office and the development process were performed much the same way. Fewer computers, less space, and less expensive tooling were needed at every step. And the effect on customers was dramatic: Lantech's share of the stretch-wrapping market zoomed from 38 percent in 1991 to 50 percent in 1994. As a result, the large operating losses of 1991 were turned into solid profits by 1993 and an industry-leading financial performance by 1994.

Work as "Flow"

As noted in Chapter 3, the rethinking of work in accord with lean principles produces the potential for greatly expanded experiences of psychological "flow." Workers in the Lantech manufacturing cells can now see the entire work flow from raw materials to completed machine. *Takt* time, standard work, and visual control (including posted work charts for all tasks) give an immediate sense of how the work is proceeding. Multiskilling and job rotation make full use of each worker's skills and the frequent repetition of *kaizen* events (as described in Chapter 5 on perfection) gives an opportunity to participate actively in work design. The constant elimination of *muda* and the movement of workers out of work cells as more efficient methods are discovered mean that the work is a constant challenge. Finally, there are few interruptions in the form of line stoppages and sudden demands to shift to a completely different task to deal with a crisis.

The situation in the office is very similar. Visual control in order-taking makes it clear to everyone where Lantech stands and the new order entry system in which one employee can perform the whole task makes it possible to get immediate results. The *kaizen* process in the office melds thinking and doing, planning and acting, just as it does in the plant.

Finally, the rethinking of development work gives a true sense of feedback as everyone involved in a project works in the same space and projects move rapidly to completion. (Formerly, the majority of Lantech's development activities were *never* completed because market conditions changed before the cumbersome development process could be concluded. We have found this same phenomenon in a wide range of firms over many years.) Employees respond positively to gaining new skills and being encouraged to use all the skills they've always had. The lack of interruptions and conflict over which task to work on next has come as a great relief.

As Bob Underwood characterizes the situation now compared to the recent past, "We were living in darkness and now we have come into the light."

Yet it would be inaccurate to characterize Lantech as some sort of paradise. Indeed, coming out into the light can be painful to your eyes. The reorganization of work tasks into a continuous flow seems to have produced widespread psychological satisfaction in daily tasks, but it is also producing the need for constant change. "We just get something working smoothly when it's time to improve it again" is a common refrain, and it's clear that each change, at least subliminally, carries risks: "Will Lantech really honor its commitment to retain excess workers? Will my contribution to improvement activities be recognized and rewarded?" Perhaps most important, many employees ask, "What will change mean for my career? Am I going anywhere or just flying a holding pattern while Lantech prospers?"

These are all important questions which firms must face once they make the initial leap to lean thinking. We will return to them in Part III on the challenge of building a lean enterprise.

The Last Step

One last step in the conversion of Lantech from a batch-and-queue to a flow organization remains to be discussed. In April 1995, Pat Lancaster promoted himself to the new position of chairman (at age fifty-two) and stepped down from day-to-day operations, turning the CEO job over to his son, Jim. Now he is starting a new creative process by thinking again about the value of his products to the customer.

As it happened, the lean transformation at Lantech was easy in one important respect because customers were quite satisfied with current-generation stretch-wrapping equipment in terms of its performance, price, and service support. That is, its *value* to them was not in question, and Lantech could safely skip the first step of lean thinking described in Chapter 1.

However, in a supremely ironic twist, Lantech has revitalized itself by banishing batches and their associated *muda* from the design and production of a product whose sole use is to wrap. . . . batches! Stretch-wrapping machinery exists for the purpose of quickly and efficiently packaging large pallet loads of goods which are shipped from firm to firm within complex production and distribution chains.

Pat Lancaster has therefore embarked on a new strategic exercise to think through the nature of packaging his customers will need in the emerging world of small-lot production, single-piece flow, and right-located facilities. Lantech needs to be ready with the right-sized, right-tasked process machinery likely to be needed in the future in order to provide the desired value for the customer.

Beyond the Simple Case

Lantech is a striking example of what happens when a small American firm makes the value stream flow smoothly as pulled by the customer in pursuit of perfection. What's more, there is absolutely no magic involved. Any small firm can follow the conversion steps just described.

However, Lantech is a simple case. Pat Lancaster is a patient investor, not beholden to the impatient stock market. He had the authority as a change agent to "make it happen." Lantech has only one plant, and it is still possible for senior management to know everyone's name. The product range is relatively simple, really just four variants of one basic concept. The workforce is relatively young and has never shown an interest in joining a union to square off against management.

While the world is full of small firms like Lantech (which can make excellent investments for an individual or small group with the skills and energy to make the lean conversion), the majority of industrial activity in almost all countries is accounted for by much larger firms with much more complexity. What does it take to carry through a lean revolution in a larger and more traditional company?

CHAPTER 7

A Harder Case

Art Byrne of West Hartford, Connecticut, presides daily over his own United Nations. Within the main plant of the Wiremold Company, of which he is president and CEO, are representatives of twenty-four nationalities. A substantial fraction of the workforce is foreign born and 30 percent list a language other than English as their original tongue.

Wiremold's polyglot workforce produces a set of objects which Art Byrne describes as "splendidly mundane": wire management systems that route complex combinations of power, voice, and data wiring through buildings, and power protection devices such as surge protectors and line conditioners, which protect sensitive electronic equipment from voltage fluctuations.

Wiremold employees use simple production machinery—plastic injection molding machines, stamping presses, and rolling mills—to make products for mature and highly competitive markets. The workforce is organized by the International Brotherhood of Electrical Workers, one of the most traditional unions in the United States. The main plant was built in the 1920s and has been expanded over the years by hodgepodge additions of one small annex after another, making continuous flow and transparency difficult to achieve.

In short, Wiremold is the typical instance of "smokestack" America: a "low-tech" product made with "low-tech" tools by a unionized, immigrant, aging workforce with limited skills, working in an ancient facility; the type of firm which has had great difficulty in world competition in the past twenty years.

When Art Byrne arrived in September 1991, Wiremold was in a profound crisis, with declining sales, deteriorating production assets, and practically no profits. Four years later, the company has more than doubled its sales with the same workforce, increased wages, upgraded its physical plant, entered into a permanent growth trajectory, and become outstandingly profitable. How this happened is an object lesson in leaning American industry.

"We Nearly JITd Ourselves to Death"

In the late 1970s, family-owned Wiremold, which had been a successful manufacturer of wire raceways since 1900, switched from family to professional management and, in the words of Orrie Fiume, its longtime vice president for finance, "asked what we wanted to be when we grew up." The wire raceway business seemed to have practically no growth potential, so Wiremold decided to enter the surge protector business. These are the ubiquitous devices, generally found on the floor under your desk, which protect your personal computer from the electric company.

The easiest route seemed to be through an acquisition, and after some searching, Brooks Electronics of North Philadelphia, Pennsylvania, was acquired in 1988. Brooks brought with it not only an established market position but also a close acquaintance with W. Edwards Deming. President Gary Brooks had embraced Deming's Total Quality Management (TQM) in the early 1980s, struck up an acquaintance with Deming, and taken not only his entire management but half of his total workforce to Deming's weeklong seminars.

When Brooks was acquired, TQM was embraced by Wiremold as well, and the management of Wiremold was soon enrolled in the Deming seminars. As Orrie Fiume notes, "Deming's Fourteen Points were a perfect fit with our values and we all loved the principles. There was only one problem: Deming teaches what he called the 'Theory of Management,' or what I call a philosophy of change. But like a lot of good management theories, it was critically short on implementation details."

By 1989, Wiremold was ready to try harder at implementation, and sent its vice president for operations to visit Japanese factories. He came back praising the concept of Just-in-Time (JIT) and immediately set about pulling down inventories and reducing lot sizes. What he could not do, because no one knew how, was introduce flow and pull by reducing changeover times for Wiremold's tools and building to a level schedule.

As Orrie Fiume remembers, "Our customer service went *completely* to hell! We soon discovered that our MRP had years earlier had a 50 percent extra margin added to the safety stock calculation. We also discovered that our reliance on enormous batches and mountainous inventories meant not only that we could tolerate slow tool changes but that we could skimp on tool maintenance. If a tool was installed in a machine and found to be defective, there was plenty of time to send it out for maintenance and get it back before we actually ran out of parts. Our tools had deteriorated to a shocking extent without the management ever realizing what was happening."

Between 1989 and 1991, Wiremold slid steadily from record profits toward breakeven. Some of the problem was lost sales when Wiremold couldn't deliver, but total sales went down only a few percent. The real problem was costs, as Wiremold paid express freight, added a whole customer service staff to explain why deliveries would be late, and paid to fix its tools. As Fiume notes wryly, "We nearly JITd ourselves to death by doing it the wrong way."

By 1991, Wiremold's longtime president was ready to retire, creating the opportunity to find a chief executive who could actually implement a lean system. As Fiume recalls, "You might think we would have simply gone backward to large batches and massive inventories, but something had permanently switched over in our minds as a result of exposure to Deming and the rudiments of lean thinking. We gave no thought to going back to the old way, but instead set out to find someone who could implement the new way."

The Change Agent

For Art Byrne, the "light went on" in 1982 when he was general manager of a small business unit, the High Intensity and Quartz Lamp Division, within the vast General Electric Corporation. One of his manufacturing managers had gone on a study trip to Toyota and had come back with amazing stories about inventory reductions due to JIT. Byrne began to read the available literature, then took his own trip, and was soon ready to give JIT a try. In one of the first JIT applications in GE, Byrne and his colleagues were able to reduce in-process inventories in his business unit from forty days down to three. He remembers, "It seemed like a miracle."

Art Byrne's problem was not with JIT but with GE. "I hated the 'make-the-month' mentality where everything was evaluated on the basis of short-term financial performance, and I became convinced that I would never be allowed to take the more difficult next steps in creating a lean organization. I already knew that when you try to create continuous flow there is going to be a step backward for every two steps forward, and I doubted that GE's instant-results management culture could deal with it."

So Byrne left to become a group executive of the Chicago Pneumatic Tool Company, a manufacturer of small air-driven tools for industrial users. However, he had hardly arrived at Chicago Pneumatic in 1986 when it was taken over by the Danaher Corporation (which we heard about in Chapter 6), and Art Byrne was soon put in charge of eight Danaher companies.

The Knowledge

One of the Danaher firms in Byrne's portfolio was the Jacobs Equipment Company (commonly known as Jake Brake) of Bloomfield, Connecticut. The sales and marketing vice president of this firm was George Koenigsaecker,[1] a particularly eager advocate of lean ideas who had made numerous study trips to Japan, including to Toyota, and read every book and article he could find on lean production.

When he was promoted to president of Jake Brake at the end of 1987, Koenigsaecker and his new operations vice president, Bob Pentland,[2] began moving machinery from process villages, tearing out conveyors (which are really moving warehouses), and setting up their first cells to make truck engine components in single-piece flow. They began to get dramatic results, but neither Koenigsaecker nor Pentland felt they knew as much as they needed to know, and they were constantly looking for ways to learn more.

Early in 1988, Koenigsaecker noticed that a weeklong seminar and *kaizen* event on the Toyota Production System was being held at the Hartford Graduate Center and in the plant of a nearby firm. He, Pentland, and Byrne decided to attend. The organizer of the course was Masaaki Imai, then becoming well known for his book, *Kaizen*. The other instructors were Yoshiki Iwata, Akira Takenaka, and Chihiro Nakao of the Shingijutsu consulting group in Japan, whom none of the Danaher group had ever heard of.

After the Danaher delegation had listened to the first day of the Shingijutsu presentation on TPS and discovered that they had worked for years as pupils of Taiichi Ohno in his efforts to spread lean thinking through Toyota's supplier group and beyond, they thought they were on to something. Koenigsaecker approached the instructors about visiting Jake Brake.

As Bob Pentland remembers, "We had never met a Japanese-style teacher, or *sensei*, and we weren't prepared for being turned down cold. Iwata simply said 'no' and stalked away. However, George is a uniquely persistent person and kept approaching Iwata, first at lunch, then at the afternoon coffee break, then at the end of the day. Every time he posed the question through Iwata's translator, the answer was a brusque 'no.' The next day George was at it again, before class, at lunch, and during coffee breaks. Finally, at the end of the second day, Iwata and his colleagues agreed to go to dinner, probably so George would stop asking.

"The minute we sat down to dinner, I pulled out a layout of our plant with the new single-piece-flow cell [identical to the Lantech cells described in Chapter 6] which we had just created. I laid it on the table in front of Iwata, and asked him whether we were doing the right thing. There was a *whatever* I tell you to do?' George and I said, 'Of course.' Iwata responded, 'If this is true, roll up the drawing, let me eat my dinner in *peace*, and I will come to your plant this evening.' "

When they arrived at the plant around 10:00 P.M., the Japanese team took one look at the new cell and pronounced it all "no good." They explained that among other problems it was laid out backwards (the work should have been flowing counterclockwise) and it would be necessary to move all the machines immediately. Koenigsaecker and Pentland had done no preparation for the visit and they knew their union leaders would be upset about the abrupt changes (which they were), but it was also clear that this was the test: "Would we do immediately exactly what they told us?" So everyone pitched in to reconfigure the cell and by 2:00 A.M. it was running again, with results far better than before.

With this introduction to the "just-do-it" mind-set of the lean *sensei*, Koenigsaecker knew he had entered a new world. "My whole notion of how much improvement was possible in a given period of time was fundamentally and permanently altered. I also realized that these guys could be a gold mine for the Danaher group."

Koenigsaecker and Pentland assumed that they had passed the critical test and that arranging a consulting relationship would be easy. So they were dismayed when Iwata abruptly headed out of the plant once the cell was running, explaining that he had done what he could but that Jake Brake managers were hopeless "concrete heads" beyond his ability to help further.

Fortuitously, the *kaizen* event held during the rest of the week at another Hartford area firm ran into the entrenched resistance of the firm's management, which refused to do any of the things the *sensei* requested. By Friday, the Danaher delegation was ready to ask for help again. This time Iwata responded that Danaher managers seemed to have no idea how to operate their business but that compared with the other American managers he had just met, there was at least some hope. However, he and his colleagues also said they were too old to learn English and that America was too far away.

Art Byrne was determined not to give up and arranged to meet them in Japan a short time later. There, after asking for help a third time, he finally got an agreement for a one-week trial to see if Danaher was truly serious.

The first day of the trial was conducted at the Jacobs Chuck Company, another Danaher subsidiary, in Charleston, South Carolina, which manufactures drill chucks for small electric drills of the type most of us have in our home tool kit and for industrial models as well. Byrne and Jacobs president Dennis Claramunt thought they would start with a one-hour plant tour and go from there. However, after five minutes, Iwata, Takenaka, and Nakao announced they had seen enough. "Everything is no good," they announced through their translator. "Will you fix it now?"

assembly and the other with Takenaka and Nakao to work on machining the steel bodies for Jacobs's industrial drill chucks. Byrne and Claramunt followed Iwata but were soon interrupted by Jacobs's manufacturing engineers, who were upset that Takenaka and Nakao were demanding to move all of the heavy machinery used for machining the chucks during the lunch hour.

Claramunt told the engineers to let Takenaka and Nakao do whatever they wanted and then went to the machining area with Byrne after lunch to see what was happening. With their sleeves rolled up and pry bars in hand, Takenaka and Nakao were working furiously to move the massive machines out of their departments and into the proper sequence for single-piece flow while Jacobs engineers and the rest of the workforce stood with their mouths hanging open.

On one level it was pure theater; the Japanese visitors surely understood what an extraordinary scene they were causing. But on another level, they were prying Jacobs loose from their bureaucratic, departmentalized, batch-and-queue past. As Byrne remembers, "By moving those machines themselves in only a few minutes—when many hadn't been moved in years and Jacobs executives would never have dreamed of touching any machinery themselves—they demonstrated how to create flow and what a few determined individuals can do. Neither Dennis nor the rest of the Jacobs workforce was ever the same again. They all threw away their reservations and got to work."

So Danaher passed the test and Japanese advisers agreed to work intensively for Danaher as their exclusive North American client. "With our *sensei* on board and with the full backing of the Rales brothers as they began to grasp these ideas in mid-1989, we had the knowledge and the authority to push lean thinking faster and faster."

By 1991, Art Byrne had introduced lean thinking all the way across the eight companies in his group, with spectacular results. He was also instrumental in spreading lean thinking in the five other Danaher companies, led by John Cosentino, who became a true believer. The transmission device was Byrne's innovation of the "presidents' *kaizen*" in which the presidents of all of the Danaher companies and their operations vice presidents were required to participate hands-on every six weeks in a three-day *kaizen* event in a Danaher plant. They moved machines themselves and in many cases learned the realities of the shop floor and the ordering and scheduling system for the first time. (One of these companies was Hennessy Industries, where Ron Hicks, whom we met in the last chapter, made the transition from "concrete head" to lean thinker through his experiences in presidents' *kaizen*.)

However, Byrne was growing restless. Like most change agents, he wanted to run his own show, but the top jobs at family-controlled Danaher were unavailable. Wiremold, on the other side of Hartford, had heard about Byrne's work at Danaher, and a match was made.

The Leaning of Wiremold

When Art Byrne arrived at Wiremold in September 1991, he found about what he expected, a classic batch-and-queue system in production operations, order-taking, and product development. Products took four to six weeks to go from raw materials to finished goods. Orders took up to a week to process. New products, even when they were nothing more than a reshuffling of existing parts, required two and a half to three years to progress from concept to launch. As a result, only two or three new products were being launched each year. Thick departmental and functional walls were everywhere, damming up the flow of value and making it impossible to see.

Byrne quickly realized that by applying lean techniques he could run the company at its current sales volume with half the people and half the floor space. Given the financial situation, he had to take immediate action. So, as his first step, he tackled the excess-people problem.

Dealing Up Front with Extra People and Anchor-Draggers

In November 1991, Art Byrne announced that the crew was too large to keep the ship afloat and offered a generous early retirement package to the aging workforce in the plants and to the office staff. Although he believed that only half the workforce was needed, he set the headcount reduction goal at 30 percent, knowing that as soon as he got the product development system working right, sales growth would absorb the remaining excess people.

Almost all of the eligible hourly workers took the retirement offer, but only a small fraction of the office staff accepted. Art and Judy Seyler, his vice president for human resources, therefore conducted a "de-layering." They classified every job in management as either:

- value creating (defined as the ability of Wiremold to pass the costs of the job along to the customer),
- nonvalue creating (from the standpoint of the customer) but currently necessary to run the business (for example, the environmental expert helping the company meet government regulations, Type One *muda*), or
- nonvalue creating and unnecessary (Type Two *muda*)

They then classified each manager as either:

- able to create value,
- able to create value with some development of skills, or
- unable to create value, even with development (usually due to unwillingness to change their attitudes about the organization of work)

After years of creating lean organizations, Art had concluded that about 10 percent of existing management will not embrace the new system. "Lean thinking is profoundly corrosive of hierarchy and some people just don't seem to be able to make the adjustment. It's essential that these anchor-draggers find some other place to work—after all, there's still plenty of hierarchy out in the world—or the whole campaign will fail."

The people in the first two categories were therefore matched up with jobs in the first two categories to create a new organization structure (compare Figure 7.1 with 7.2) with a new roster of players. Employees not finding useful jobs were given a generous severance and within thirty days of Art's arrival the new structure and player roster was in place. Only one outsider was recruited, Frank Giannattasio, the new vice president for operations.

As Judy Seyler looks back on this event, she notes that it was terribly traumatic in a hierarchical, paternalistic organization in which no one had ever been asked to leave. "Even though the financial cost was very large,

FIGURE 7.1: OLD WIREMOLD ORGANIZATION

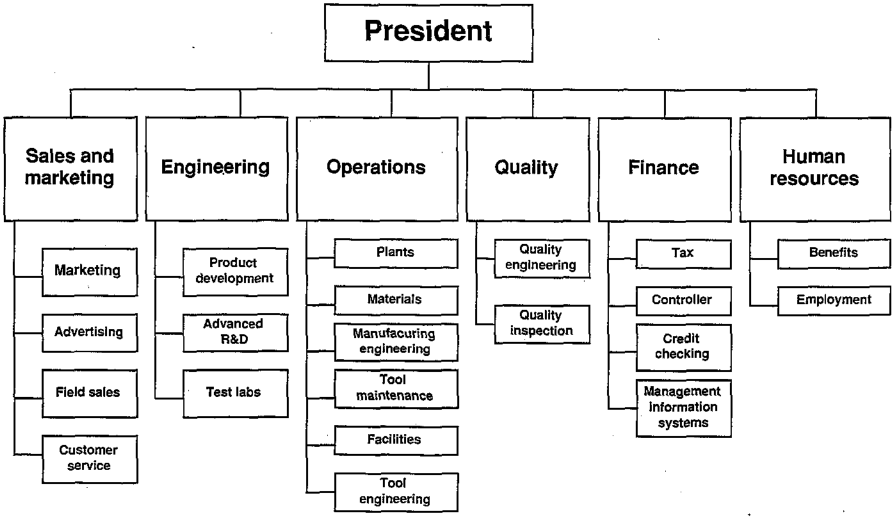
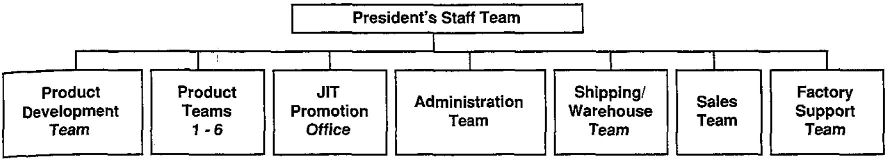


Figure 7.2: New Wiremold Organization



particularly given our lack of profits, Art was determined to be generous with people while making it clear that in the future everyone must create value by working together in a different way."

When the headcount reductions were completed, Art Byrne called a meeting of the entire workforce of the parent company and announced that no one would ever lose their job as a result of the improvement activities that would start immediately. "The bad part is over; now we will all learn how to continuously create more value so that we never have bad days again."

Byrne was in effect giving job guarantees to his union workforce without asking anything in return except that they be open-minded to change. "I'm certain that 99 percent of American companies wouldn't do this, but taking away the fear of job loss is at the very core of a lean conversion. Think of it logically from a human perspective rather than as some corporate bureaucrat. If I asked you to help me reduce the number of people needed to make a particular product from five down to two, and after you did, I followed up by laying off three people, one of whom was your cousin and another a good friend, what would you say to me when I asked you to help me do the same thing a month later for another product?"

Teaching People How to See

Based on his experience in "leaning" eight separate businesses in the Danaher group, Byrne had concluded that the single most effective action in converting an organization to lean practices is for the CEO to lead the initial improvement activities himself. "This is where most American companies fail right at the outset. CEOs want to delegate improvement activities, partly because they are timid about going out on the shop floor or to the engineering area or to the order-taking and scheduling departments to work hands-on making improvements. As a result, they never really learn anything about change at the level where value is really created. They continue to manage in their old by-the-numbers manner, which kills the improvement activities they thought they started. The fact is that big changes require leaps of faith in which the CEO must say 'Just do it,' even when 'it' seems contrary to common sense. If the CEO spends time in real operations learning just how bad things really are and begins to see the vast potential for improvements, he or she will make the right decision more often."

Because no one else in the company understood lean principles, Art Byrne led the initial training sessions himself. Using a manual he had written, he conducted two-day sessions on lean principles for 150 employees followed immediately by three-day *kaizen* exercises so employees could use the skills they had just learned. (This was very different from Wiremold's previous improvement activities, conducted as part of TQM, where improvement teams met weekly for an hour or two, mostly to plan improvements to be implemented weeks or months later.)

Byrne then gathered his managers and union head together and took them on "the walk of shame" through every part of the plant and through the engineering and sales departments. "There was *muda* everywhere and my managers were now able to see it. I told them that we were going to convert every process, including product development and order-taking, into continuous flow and that we were going to learn how to pull. I also told them that I was going to get them the best help in the world, from Iwata and Nakao, who were at the end of their exclusive agreement with Danaher and ready to work for Wiremold."

Attacking Every Value Stream Repeatedly

Soon hundreds of weeklong *kaizen* activities were under way (and continue to this day), involving practically every employee, as every value stream in Wiremold was repeatedly evaluated for ways to make it flow better and pull more smoothly. Wiremold's assumption is that every stream can always be improved in pursuit of perfection and that every stream must be improved in pursuit of perfection. Equally important, it is presumed that results can be achieved very quickly, the common expression being that "if you can't get a major improvement in three days you are doing something wrong." Once this mentality is reinforced by results—and employees begin to believe management's guarantee that no job will ever be lost due to improvement activities—improvement can become self-sustaining.

Re-creating the Production Organization to Channel the Value Stream

When Art Byrne de-layered the Wiremold organization (again, see Figure 7.2) he did far more than remove tangential jobs and frills that could no longer be afforded. He smashed the departmental barriers to focus everyone's efforts on the value stream by creating dedicated production teams for each of Wiremold's six product families. The purchasing, manufacturing, and scheduling (MRP) groups within the Operations Department, the Engineering Department, and the "process villages" (stamping, rolling, molding, painting, assembly, etcetera) in the plants were eliminated, with their personnel reassigned to product teams provided with all of the resources needed to produce a specific product family.

Let's take Tele-Power™ Poles as an example. (These are the steel or aluminum columns extending from floor to ceiling in open office settings, with power and communications outlets on every side of the column to permit a host of adjacent workstations to plug in. They are offered in an enormous variety of shapes, lengths, plug configurations, and colors.) Team leader Joe Condeco was given complete responsibility, and profit-and-loss accountability, for Wiremold's "pole" products from initial launch through their production life. More radically, the team leader, the product planners, the buyers, the factory engineers, the production supervisors, and the production associates were all co-located on the factory floor immediately adjacent to the realigned machinery producing the poles in single-piece-flow cells.

The team was given its own punch presses and rolling mills, as well as assembly equipment, so it could be self-sufficient. Before, the assembly activity was dependent on the Rolling Mill Department for their bases and covers. Despite large stocks on hand, they would often lack the right base or enough covers. When they asked the Rolling Mill Department for more of a missing item, the response would often be, "Sorry, but the master schedule generated by the MRP system calls for us to make other items now. You'll have to wait until next week or take your problem to a higher level." Now the Tele-Power™ team has all of the equipment it needs. *There can be no excuses.*

The new setting was initially a great shock to the "white collars" who had always worked in a remote office and seen themselves in a very different light from "factory workers." (Wiremold soon implemented a casual dress code based on Art Byrne's belief that "neckties cut off circulation to the brain and inhibit teamwork." This was another problem for office workers who somehow felt that their appearance rather than their skills and their contributions made them special.) Reassignment to product teams was also a shock to the process specialists working in the process villages like the Rolling Mill Department who had traditionally hoarded their tricks of the trade. However, everyone soon came to like it. For the first time, they could actually see value flowing!

Introducing a Lean Financial System and "Scoreboard"

To get the production teams to work in accordance with lean principles, Wiremold needed to junk its traditional system of standard cost "absorption" accounting, which allocated costs by labor and machine hours in accordance with mass-production thinking. Production managers knew from experience that they had to "absorb" allocated overhead by spreading it over as many machine and labor hours as possible. This system gave an overwhelming incentive to keep every worker and every machine busy—to "make the numbers"—by producing inventory, even if the inventory consisted of items no one would ever want.

As Orrie Fiume remembers, "Standard cost and variance analysis were declared dead as concepts immediately after Art's arrival. We looked at Activity Based Costing but knew it wasn't the answer. Its advocates will tell you it is based on cost drivers, but in reality it's just a different method of allocating overhead. There is still too much allocation of aggregated costs downward. We were determined to work from the bottom up."

The key to the new way of thinking was to organize production by product families, then let each product team do its own purchasing and buy all its own tools. A simple system could then be devised to assign real costs to each product line. Today, more than 90 percent of the costs involved in making a Tele-Power™ pole, for example, come from product-specific cost analysis. Only a small fraction of cost is an allocation outside the control of the team, specifically occupancy costs for whatever space the team is using in a plant. And even in this case, the team is charged only for the space it actually uses, so costs can be reduced by using less.

Some elements of the old standard cost accounting system are retained in the computer because the financial statement needs them—for example, the value of in-process inventories. However, these are deemphasized in evaluating the performance of the product teams, which are told to concentrate on the cost of manufacturing instead. Similarly, the financial implications of running down the inventories during the transition period were not shown to the product team leaders for fear they would do the wrong thing.[3]

In addition to a simple profit-and-loss calculation, Wiremold's production teams were given a new "scoreboard" consisting of some simple, quantitative performance indicators:

- productivity of the product team (expressed as sales per employee),
- customer service (expressed as percent of products delivered on time),
- inventory turns, and
- quality (expressed as the number of mistakes made by the team)

The team leaders and their teams can see these indicators at all times because they are prominently posted. In addition, the two primary ways to improve are obvious. First, smooth the flow of products through the system, with no backflows for reworking quality problems, no scrap, and no in-process inventories. Then, make only those products customers actually want, because productivity is measured as end-market sales (not additions to in-process inventories) per employee.

To keep everyone marching at the same pace, Wiremold equips the scoreboard with a set of expectations as well. Specifically, team leaders and their teams are expected to:

- reduce defects, as shown in the quality indicator, by 50 percent every year;
- improve productivity, expressed as sales per employee in constant dollars, by 20 percent every year;
- deliver 100 percent of products *exactly* on time;
- increase inventory turns to a minimum of twenty per year; and
- increase profit sharing to 20 percent of straight wages (as explained in a moment)

"Variance analysis" is still performed but not based on variances from standard costs. Instead, when the trend line starts to diverge from performance targets, the team collectively searches for the root cause of the variance rather than maneuvering to "make the numbers," as in the old days.

Running Down Inventories

Because Wiremold is privately held and the board of directors understood what was happening, the special financial problem of inventory reduction in a lean transition was not a major concern. For a publicly traded company, however, rapidly running down inventories can be a real problem, one worth a brief digression. As firms move from batch-and-queue to flow systems, enormous amounts of cash are suddenly made available from freed-up inventories. (This offers the firm a special strategic opportunity, as we will see in a moment.) The problem is that the removal of these inventories increases production costs, *as shown on the financial statement,* and can easily wipe out profits.

Let's take a simple example. Firms typically calculate their costs of production and profits in the following manner, as shown in the left-hand column of Table 7.1.

TABLE 7.1: CONSEQUENCES OF INVENTORY REDUCTIONS FOR PROFITABILITY

	MASS PRODUCTION METHODS	LEAN PRODUCTION METHODS
Beginning in-process inventory	$ 576,000	576,000
Direct materials purchased	924,000	637,000
Direct labor	958,000	958,000
Indirect manufacturing costs	465,000	465,000
Subtotal	2,923,000	2,636,000
Minus ending in-process inventory	− 576,000	− 100,000
Total costs of production	2,347,000	2,536,000
Total revenue from sales	2,500,000	2,500,000
Profit (loss)—pretax	153,000	(36,000)
Cash flow—pretax	153,000	440,000

Now, suppose that the new "lean" management takes in-process inventory down dramatically from $576,000 to $100,000 while holding everything else constant (except, of course, material purchasing, because products are being made largely from the inventories already on hand). Running the numbers again, as shown in the right-hand column, it's apparent that the new management, while trying to "do the right thing," has moved the company from a $153,000 profit to a $36,000 loss (even as cash flow has zoomed).

This phenomenon can be very bad for publicly traded companies unless the management actively explains the situation to stockholders in advance. The only alternative to education is a slash-and-burn campaign of headcount and cost reduction (in the direct labor and indirect manufacturing cost accounts) to restore short-term profits. This can, however, easily set back the introduction of lean thinking or even make it impossible if the traumatized workforce refuses to cooperate with lean initiatives.

Creating a "Lean" Function

To help the product teams continuously improve, Art Byrne created a new function, the JIT Promotion Office (JPO). The old Quality Department, some of the training activities formerly conducted by the Personnel Department, and several high-potential associates from different areas of the firm were grouped under the JPO. With the JPO, the task of working through the entirety of Wiremold, value stream by value stream, could be speeded up.

The product team leader and the JIT Promotion Office jointly evaluate the value stream for the product to determine what types of *kaikaku* and *kaizen* activities should be performed and when. A team leader from the product team and a facilitator from the JPO are then assigned to each improvement team (which might be a subset of the product team, the whole team, or some portion of the team plus outside experts with needed skills). Because the team leader will go back to her or his job on the product team once the *kaizen* is finished, the facilitator from the JPO shoulders the critical responsibility for seeing to the completion of the follow-up work invariably resulting from a weeklong improvement effort.

In addition to planning and facilitating improvement exercises, the JPO teaches every employee the principles of lean thinking (identifying the value stream, flow, pull, and the endless pursuit of perfection) plus lean techniques (standard work, *takt* time, visual control, pull scheduling, and single-piece flow in particular) and periodically reteaches them. As Frank Giannattasio notes, "This is an enormous but critical challenge. Your middle management, in particular, feels threatened by the lean transition and the removal of all those safety nets. When in doubt, they will take you right back to making batches and building inventories unless you reinforce the message through continued teaching, coupled with continuous hands-on improvement exercises."

Offering Ironclad Job Guarantees in Return for Flexibility

As we noted earlier, Art Byrne knew that if the value stream for every product was going to be unkinked continuously, people would continually be left by the side of the stream. Resistance to continuous improvement would be chronic unless he guaranteed that workers would not be out on the street, even if their specific job was eliminated. He also knew that the existing work rules in Wiremold's union contract—restricting stampers to stamping, painters to painting, molders to molding, and so on—would make it impossible to introduce flow and to continuously improve every activity. Finally, he knew that his workforce would have a very hard time differentiating layoffs due to weak demand from layoffs due to *kaizen*. Therefore, once the initial retirement offer was accepted, Art went immediately to his union and offered job guarantees for the remaining workers in return for their cooperation in working in a new way.

The union was suspicious at first. Wiremold's former director of labor relations had been an old-fashioned hard-liner and the union presumed that any management offer of job guarantees must contain fine print which somehow reversed its on-the-surface meaning. In the end, however, the union decided that Byrne would deliver on his promises.

Curiously, for reasons Art Byrne finds hard to understand, executives in many companies in the Hartford area were more skeptical about his job guarantee offer than his union. "People tell me all the time that I'm crazy to make an ironclad guarantee of jobs. They say, 'What if something goes wrong and your sales fall off?' But my view is that management has five lines of defense before showing people the door: (1) reduce overtime, (2) put the extra people on *kaizens* (to get a future payback), (3) in-source some components from marginal suppliers we plan to drop anyway (remembering that our equipment is now highly flexible), (4) cut the workweek across the board, and, most powerful, (5) develop new product lines to grow the business. Our employees are now all highly skilled in process improvements and only a concrete head would fire skilled people due to short-term business fluctuations."

Re-creating the Product Development System to Channel the Value Stream

The product development system Art Byrne found in the fall of 1991 was clearly not going to grow the business. Engineering vice president Steve Maynard remembers that thirty products were under development and all were making slow progress. "We had long queues between the stages in development, we had departments within engineering with batch production, and we had expediters. There were absolutely no priorities except that some projects at some times had 'the voice of the president' behind them and received expeditious treatment." The average project actually making it through the system took three years, but many stragglers were lost in action along the way.

Fortunately, Steve Maynard already knew what to do. He had learned, at a University of Hartford seminar in the fall of 1990, that Quality Function Deployment and dedicated development teams are an unbeatable combinaand Productivity, and MIT professor Don Clausing, one of the disseminators of the "House of Quality" concept,[4] took Steve through the steps needed to introduce the "voice of the customer" in a highly structured, continuous-flow development process.

However, back at Wiremold, the senior management was so busy with the ongoing TQM effort that there was no time for another program. They told Steve Maynard, "Wait until next year." Fortunately, by "next year" Art Byrne was on the scene. "When I first met Art, I said, 'What do you think about QFD and dedicated development teams?' He said, 'Do both *immediately*. And by the way, your new target for product development time is now three to six months, not three years.' We were off and running within a week."

Steve Maynard's first step in the fall of 1991 was to start formal in-house training in QFD, using a consultant for technical support.[5] All the senior executives attended this training, just as every manager, no matter how senior, no matter what their job, participated in shop-floor *kaizen* activities. Art Byrne's theory was that every manager in an organization must understand the basic activities of that organization, notably product development, production operations, and sales/scheduling, and that the only way to learn was intense exposure to systematic principles.

Next, Maynard and the senior management team asked an obvious but previously neglected question: What businesses are we really in? They reviewed the thirty ongoing development programs and "deselected" those—most of them, in fact—which did not support a specific business: tele-power, power and data management, plastic products, etcetera.[6] This shrunk the number of projects dramatically, and those which remained were prioritized. These projects were then placed in a product plan, showing their target dates for introduction.

For each program judged worthy of continuing, Maynard designated a three-person team consisting of a marketer, a designer/product engineer, and a production/tool engineer. The team was sent to talk directly with prospective customers in the building design and construction community to come up with a broad definition of the product through an initial QFD process. They asked the "value question" described in Chapter 1 and came back saying, for example, "What we really need is a Tele-Power™ Pole which can accommodate any height of ceiling, which can be ordered in a very wide range of colors, and which is unobtrusive."

Steve Maynard remembers the amazement among many Wiremold old-timers when these teams were formed. "They asked me, 'Why have we got a tool-design guy out in the field talking to customers? Doesn't the need for specialization and the division of labor require that tool designers mentalized, everything-in-its-place way of organizing work was truly striking."

Once the broad definition of the surviving products was determined, a truly multifunctional team was formed to develop a detailed product specification in engineering language. The team was co-located in a dedicated space in the Engineering Department and included the team leader from the appropriate product family (Tele-Power™ Poles in our example) plus the production planner, the production/tool engineer (a member of the original three-person, product-definition team), and a buyer. The team was told to achieve a target cost determined by estimating the market price and subtracting an acceptable margin.

When the precise specification of the product was accepted, detailed part and tool design was conducted by the team, again working to target cost. Toward the end of the process, the whole team moved its desks to the factory floor to go through process-at-a-glance and standard work exercises with the production team handling the product. (Remember that thinking about manufacturability has been present from the beginning. The production/tool engineer was on the original definition team.)

By mid-1992, Wiremold was ready with its first product under the new regime. It had taken only six months and tool costs were only 60 percent of what had been originally budgeted, based on past experience. Even as Wiremold's managers in the physical production and order-taking process were learning how to see, Wiremold's marketers, product designers, and engineers were learning how to hear the voice of the customer and how to make designs flow quickly and directly through the development process.[7]

Fixing the Order-Taking Process

The third key activity of any business is order-taking, scheduling, and delivery, and Art Byrne made no distinction between this "business process" and the firm's physical production. It was subjected to the exact same *kaikaku* and *kaizen* process at the same level of frequency as every production activity.

As in most batch production organizations, Wiremold's order entry and shipping was disconnected from physical production. A master schedule in the MRP system, based on market forecasts, was supposed to ensure that adequate stocks of finished goods were always on hand in a massive centralized warehouse, so that when an order was received it could be processed and then shipped from inventory.

The orders themselves were also processed in a batch mode through a central Customer Service Department. This department entered orders throughout the day into a computerized order-processing system. The orders were processed overnight in a batch and, if inventory was on hand, pick lists of what to ship were printed out the next morning in the Shipping Department. Over the next two or three days the Shipping Department at the warehouse would gather the goods and send them to Wiremold's distributors.

However, items on a customer order often were not available, despite large inventories, so very few orders were shipped complete. Instead "back-ordered" items were shipped over an extended period as they became available. Because of the MRP system and the large batches in each production run, it was not unusual for a single customer order to ship over many weeks or months. Also, because most orders had delayed items, a large Customer Service Department was needed to keep track of orders and to respond to customer questions about delayed items.

The end result of all this handing off of orders and the massive warehouse was that it took almost a week to process and ship an order when everything was in stock. Yet most orders called for items which were delayed for extended periods and the system had many potential sources of errors. The Customer Service Department found it very difficult to keep up with its dual role of making customers happy about delayed or incorrect shipments and spurring the rest of Wiremold to get the job done right.

After a series of *kaizen* teams went through the entire series of activities—from order-taking to shipping—it was possible to shorten the order-receipt-to-ship time from more than a week to less than a day. To achieve this, orders were sent to shipping four times during the day (rather than in one big batch overnight) and the central warehouse was closed, freeing up 70,000 square feet of space. Upon receipt of the orders, shipping circulated carts past the small finished stock racks at the end of each product team's production process.

As the shipper withdrew parts from the rack and pushed empty parts containers down a return chute, this became the signal—the only signal—for the product team to make more of a given part. (The MRP system which formerly kept track of the movements of individual parts within the Wiremold production system was gradually given the much smaller task of long-term capacity planning and ordering parts from suppliers not yet on pull systems.)

This new approach, which required many fewer people and resulted in fewer errors, could only be introduced over a period of about two years as Wiremold began to convert from batches to product teams with single-piece flow. Parts which had been produced in one-month batches were soon being produced every day, a feat which required that many machines be changed over twenty to thirty times per day rather than the former three to four times per week.

Although Wiremold's competitors in the electrical industry are now being forced to match its quick-shipping capability, they seem to be doing it the way so many American firms are achieving "just-in-time," by maintaining even larger inventories of finished units or by switching to a "Max-Flex" system such as we saw at Lantech, in which mountains of component parts are prepared in advance so that final assembly can be conducted in direct response to customer orders. Both approaches are inferior to a truly lean pull system from start to finish.

Linking Compensation to Profits

Wiremold had always paid base wages at slightly above the average level for the Hartford area. It then tried to reward its workers for good results through a profit-sharing plan funded with 15 percent of pretax profits, paid quarterly in the form of a check, and by contributing shares of company stock as the employer contribution to a savings plan. The problem was that, in the years just prior to Art Byrne's arrival, there had hardly been any profits and stock values had slumped. In addition, the old batch production system made it hard for employees to see any connection between their own efforts and the success of the firm.

Art Byrne resolved to keep the existing profit-sharing arrangement but to steadily increase profits ("by working smarter than our competitors") and to show everyone the financials so that the reasons for profitability would be clear. During the first years of "lean management," profits at Wiremold have increased from 1.2 percent of wages in 1990 to 7.8 percent in 1995, and Byrne is still strongly committed to increasing profit sharing to 20 percent of every employee's pay.

Improving Suppliers

After many internal improvements were made, it became increasingly apparent that many of Wiremold's problems were external. Purchased goods and raw materials accounted for a significant percentage of Wiremold's total costs, yet no effort had ever been made to improve supplier performance. Instead, Wiremold's traditional purchasing operation had concentrated on controlling supplier profit margins by ordering every part and type of material competitively from multiple sources.

Kaizen teams moved quickly to dramatically reduce the number of suppliers, from more than 320 in 1991 to 73 by the end of 1995. This was essential if Wiremold was going to be able to take time with each supplier to improve its performance. But then it was necessary to start with the most critical suppliers and teach them to see.

In April 1992, a Wiremold *kaizen* team paid a first visit to Ryerson, a giant steel fabricator much larger than Wiremold, with fabrication facilities spread all across North America. Ryerson supplies Wiremold with large rolls of steel which Wiremold stamps or bends to make the cases for many of its products. Ryerson had adopted state-of-the-art techniques to the extent that it had just begun to deliver to Wiremold every day, "just-in-time." However, at the back of Ryerson's plant, the Wiremold JIT team found just what they had expected: a neat row of steel coils, each a day's supply for Wiremold, fifty days in a row produced by Ryerson in one enormous batch. Just-in-time had been nothing more than an inventory shuffling exercise because Ryerson didn't know how to produce in small lots.

The Wiremold team therefore went to work on Ryerson's massive steel cutting machines, which took two shifts to change over from one cutting pattern to the next. This, of course, was the cause of the massive lot of coils laid out in the shipping area. In a short time it was possible to bring changeover times down from two shifts to about thirty minutes, and Ryerson began to meet Wiremold's needs each day for delivery during the day.

Even better, from both Ryerson and Wiremold's standpoint, Ryerson was soon able to produce for all of its other customers on a true "Just-in-Time" basis, driving down costs across the board. Wiremold, of course, expected something from Ryerson in return for its trouble, and negotiated a range of special services—like absorption of materials cost increases for extended periods and extra-short runs of steel for certain low-volume applications. As a result of Wiremold's proactive stance toward a key supplier, Wiremold, Ryerson, and all of Ryerson's other customers were much better off, a win-win-win achievement for lean thinking.

Devising a Growth Strategy

Art Byrne notes that "our production system and its needs are fundamental to our strategy." Because the application of lean thinking to batch-and-queue organizations liberates tremendous amounts of resources—people (including engineers and managers), space, tools, time (to get to market much faster), and *cash*—it is both possible and necessary to grow rapidly. It is possible to grow rapidly because the means are self-generated; it is necessary to grow rapidly to provide work to support the job guarantees which are the social basis of the system. In consequence, Wiremold has grown rapidly along three tracks.

One important means of growth for a lean organization is to rethink what can be done in continuous flow. We believe that many organizations try to do too much—in particular, to control suppliers of "key" technologies. But many organizations, like Wiremold before Art Byrne arrived, also do too little in the way of physical production because they imagine that scale economies require the purchase of many items from firms using enormous, high-volume machines in centralized plants to supply these items in massive batches to many customers.

Cord sets are a nice example. Wiremold products use enormous numbers of cord sets—the wire and plug ends used to connect surge protectors and other power conditioning devices to a power source. In the past, these were produced in large batches by cord set manufacturers supplying many firms like Wiremold across a range of industries. The problem was that Wiremold's production was constantly being jeopardized by the lack of the proper cord sets as sales trends changed. Wiremold would have brown when only white cords were wanted or have twelve-foot cords when the customer wanted fifteen-foot cords. Resolving these shortages often took two to four weeks due to the batch production methods of cord set suppliers.

When Byrne arrived at Wiremold he asked, "Why can't we produce cord sets, indeed at the same rate and in continuous flow with our end product?" And, as is usually the case, when Wiremold's tool engineers looked at the economics of cord set production, they found that the cost and time savings from using small, simple machines merged into the production sequence for the finished product not only overcame the problem of having the right cord set on hand as demand shifted, but also reduced the cost per cord set. So, Wiremold has begun to supply its cord set needs in-house. After all, Wiremold has plenty of excess space, plenty of extra people, and cash readily available to buy or make the necessary, simple machines.

Any would-be lean producer needs to look at this issue more generally, asking in each case, "What physical activities can we incorporate directly into a single-piece-flow production process?" Doing this also reduces the number of suppliers dramatically, making improvement of the remaining suppliers much easier.

Wiremold's second growth strategy has been to buy up small firms with allied product lines (and who use batch-and-queue methods) in order to increase the scope of Wiremold's product offerings. The first wave of inventory reductions at Wiremold (during the first two years of comprehensive *kaizen* activities) produced $11 million in cash. This money was used to buy five firms with complementary product lines generating $24 million in sales volume.

In essence, Wiremold was able to convert $11 million of *muda* (in the form of inventory), which would have cost about $1.1 million in carrying million in new sales volume, which at a 10 percent operating margin generates $2.4 million in income. The $3.5 million income swing is highly significant for a company the size of Wiremold (with about $250 million in annual sales). Equally important, because the product lines of the five firms were complements to existing lines, Wiremold's sales force suddenly had a much more complete range to offer customers, which helps increase the overall growth rate.

The fact that Wiremold freed up approximately 50 percent of the space in all of its operations (excepting the central warehouse, which was totally eliminated) greatly helped the acquisition campaign. While Art Byrne's philosophy is to retain and upgrade existing management, several of the companies purchased were available because the family management could no longer run them successfully and wanted out. This provided consolidation opportunities.

For example, two of the firms purchased were consolidated into Wiremold's Brooks Electronics operation in Philadelphia. Prior to the acquisitions, the three companies had operated independently, utilizing 114,000 square feet of space. Now, the combined operation has increased its total sales significantly, yet is located in Brooks's original 42,000 square feet of space. Inventory has been reduced by 67 percent, the number of employees needed to run the combined operation has been reduced by 30 percent, and the surplus buildings have been sold.

In effect, Art Byrne and Wiremold are running a lean vacuum cleaner across the world of batch-and-queue thinking in the wire management industry. Each time Wiremold's vacuum sucks up a batch-and-queue producer it spits out enough cash to buy the next batch-and-queue producer! Because of Wiremold's need to grow to utilize freed-up resources, this process can and must be repeated indefinitely. (As we will show in Chapter 11, the first firm to adopt lean thinking in any industry can and must perform this same feat.)

The third and final element of the Wiremold growth strategy is the rapid introduction of new products, utilizing the new product development system with its dedicated teams and Quality Function Deployment methods described earlier. For example, the new product line described in Chapter 1 has increased sales by 140 percent, both by creating a new niche in the market and by stealing sales from competitors unprepared to match Wiremold's pace of product introduction.

All three strategies are critically dependent on the lean techniques introduced in production, order-taking, and product development. Indeed, the rapid introduction of these techniques *is* Wiremold's fundamental strategy. Art Byrne remembers that in previous jobs he had often wanted to go faster interested in massive long-range "strategic" planning efforts which they believed should take precedence. "To my way of thinking, this is exactly backwards. Introducing lean techniques in every business activity should be the core of any company's strategy. These provide both the opportunity and the resources to generate and sustain profitable growth. Profitable growth is what the strategic planners of the world are always seeking, but find hard to achieve because their company's operations can't deliver on their strategies."

The Box Score After Five Years

As we will see in Chapter 11, three years is about the minimum time required to put the rudiments of a lean system fully in place and two more years may be required to teach enough employees to see so that the system becomes self-sustaining. Wiremold's performance over the five-year period from the end of 1990 to the end of 1995 is therefore a good test of the potential of lean thinking. The results are quite striking.

To begin with product development, time-to-market has been consistently reduced 75 percent, from twenty-four to thirty months down to six to nine months. Sixteen to eighteen new products are being introduced each year (compared with two to three in the period through 1991), but the engineering/design headcount has stayed the same.

Several new computer-aided design technologies might be assigned some of the credit for these gains, except that these techniques were adopted in 1990–91, *before* the time-to-market and productivity gains. As we have emphasized throughout this volume, advances in hard technologies can be useful and in many cases are very important, but they are unlikely to yield more than a fraction of their potential unless they are incorporated in an organization which can make full use of them. By placing product designs in single-piece flow with a dedicated, multiskilled, co-located team and no interruptions, Wiremold has eliminated backflows and rework in the development process while also reducing manufacturing costs and dramatically spurring sales with products which accurately address customer needs.

The rethinking of order-taking, scheduling, and shipping has produced the same results. The old batch system, which needed more than a week to receive, process, and ship a typical order, now needs less than a day. Past-due orders are now less than one tenth of their 1991 level and continue to fall as Wiremold refines its pull system through all six product teams. Order entry errors have been practically eliminated, and misrouted or unanswered queries in the much smaller Customer Service Department have fallen from 10 percent to less than 1 percent.

In physical production, the results are exactly as we would expect. The amount of plant space needed to produce a given volume of product has been cut by 50 percent and productivity has been increasing at a rate of 20 percent per year. The time for raw materials and components to travel from the receiving dock to the shipping dock in Wiremold's plants shrank from four to six weeks to one to two days. Inventory turns have increased from 3.4 in 1990 to 15.0 in 1995.

To make this possible, Wiremold has continued to reduce setup times on all of its machines and to convert all production activities for its product families to single-piece flow. For example, punch presses with large progressive dies that used to take two to three hours to change are now changed in one to five minutes; rolling mills which took eight to sixteen hours to change over in 1991 can now be changed in seven to thirty-five minutes; plastic injection molding machines that took two to four hours to change over in 1991 can now be done manually by one Wiremold employee in two to four minutes. As a result, machines that previously shifted from one product to the next two to four times per week, now change products twenty to thirty times a day.

By aggressively implementing single-piece flow, operations requiring five to eight operators in 1991 are conducted with one to three employees today. By utilizing single-piece flow, JIT, and Total Productive Maintenance in the largest and most complicated assembly operations, productivity has been increased by 160 percent over three years. Equally important, single-piece flow has been instrumental in reducing defects by 42 percent in 1993, another 48 percent in 1994, and another 43 percent in 1995, almost at Wiremold's target rate of 50 percent per year indefinitely. At the same time, standard work, *takt* time, and visual control have been slashing accidents and injuries, which are down by more than half compared with 1991.

Putting the improvements in product development, order-taking, and physical production together, we find that sales per employee more than doubled, from $90,000 in 1990 to $190,000 in 1995. However, this and the figures just cited are all relative to Wiremold's previous performance. The indicators which truly count in the marketplace are sales, profits, and market share. Happily, between 1990 and 1995, Wiremold's sales in its core wire management businesses—owned before the lean vacuum cleaner was turned on—more than doubled in an otherwise stagnant electrical equipment market and profits of the whole firm—including the new businesses—increased by a factor of six. What's more, the growth rate, including acquisitions of related businesses, is picking up, in line withWiremold's strategy of doubling its sales every three to five years for the foreseeable future.

All of these indicators are summarized in Figure 7.2, a "box score" for Wiremold under lean management.

TABLE 7.2: WIREMOLD UNDER LEAN MANAGEMENT

	1990	1995
Sales per employee ($000s)*	90	190
Throughput time to produce average product	4–6 weeks	1–2 days
Product development time	3 years	3–6 months
Suppliers	320	73
Inventory turns	3.4	15.0
Space required (index)	100	50
Sales (index)	100	250
Operating profit (index)	100	600
Profit sharing (% of straight wage)	1.2	7.8

* Note that Wiremold's degree of vertical integration in manufacturing has increased substantially as items such as cord sets and plug outlets have been brought in from suppliers. Thus, value created per employee has increased even more if adjusted for the portion of the value stream under Wiremold's direct management.

But What About Firms with More Severe Problems?

The Wiremold story is extraordinary. The firm has been transformed in a remarkably short time and now gives every prospect of growing rapidly into an industrial giant. What's more, we could repeat this story in dozens of medium-sized firms we have discovered across the United States during research for this book.

Wiremold was a greater challenge than Lantech, given the age and narrow skills of its workforce, the stagnation of its core market, and the entrenched us-versus-them mentality of the old management and union, but is it still a fair test of lean thinking? After all, Wiremold has only fourteen hundred employees, operates primarily in two neighboring countries (the United States and Canada), and has relatively simple product and process technologies. What about the aging industrial giants who present the most visible managerial challenges? What about the publicly traded, mass-production firm with tens of thousands of employees, global operations, complex technologies housed in deep technical functions, and a complex network of component systems suppliers? Can lean techniques produce the same results, and in the same time frame? We turn now to Pratt & Whitney, which is truly the acid test of lean thinking.

CHAPTER 8

The Acid Test

On June 1, 1991, Mark Coran drove across town, from the Hartford, Connecticut, headquarters of the United Technologies Corporation, to the East Hartford headquarters of Pratt & Whitney, UTC's largest subsidiary and the world's largest builder of aircraft engines. UTC chairman Bob Daniell had just given him a new assignment—one for which his background as UTC's corporate controller and star cost-cutter seemed ideal preparation.

The problem at Pratt appeared to be structural and substantial but not desperate. As the world's largest builder of military jet engines[1] (accounting for a third of its total business in the 1980s), Pratt was faced with the end of the cold war, a reality to be confirmed shortly with the collapse of the Russian countercoup in August 1991. It suddenly seemed likely that much of the military engine business was gone for good.

In the short term, the loss of military business was offset by an extraordinary boom in orders for commercial engines. As the world's market-share leader[2] in commercial aircraft engines, Pratt had ridden the wave and racked up a record operating profit of $1.01 billion in 1990 on a record $7 billion of military and commercial sales. However, anyone familiar with the roller-coaster demand cycle in the commercial engine business knew that sales at this level couldn't be sustained for long, and in fact, orders for spare parts had already started to fall. Therefore, Mark Coran's job, as the new executive vice president for operations at Pratt, was to prepare the manufacturing operations in a massive company with 51,000 employees for a perhaps 10 percent permanent reduction in the size of the business, and to do this before the commercial-order boom collapsed.

As it turned out, Mark Coran had no time to work with. June 1991 would prove to be the peak month of production volume in the history of Pratt & Whitney, with "shop hours" of work—the conventional Pratt measure of production activity—running at an annual rate of 11 million. Soon, commercial jet aircraft orders, which had reached a record high of 1,662 in

1989, started to drop steeply as the world recession set in, falling to a low of only 364 in 1993.

Much worse for Pratt's finances, the airlines were dipping into their inventories of jet engine parts to repair their fleets, rather than ordering new parts from Pratt. Orders for Pratt spares were sliding rapidly by the fall of 1991, and by 1992 were running at only 63 percent of the 1989 peak. This was a crushing blow because spare parts account for the great majority of the profits of every aircraft engine company, due to the industry practice of selling new engines at substantial discounts in order to capture market share and create a large user base for their highly profitable, captive spare-parts businesses.

To make matters worse, Pratt and its two global rivals—General Electric in the United States and Rolls-Royce in the United Kingdom—were locked into spending large sums right away—$3 billion in total among the three firms—on development of the next generation of jet engines. These are the 84,000 to 100,000 pounds of thrust "monster motors" needed for the Boeing 777 and possibly for the proposed 600-seat Airbus A3XX. (The first of these, the Pratt PW4084, entered airline service on the Boeing 777 in June 1995.)

Because of its four-year product development cycle for new engine designs and the eighteen-month production lead time to physically build an engine once ordered, Pratt was helpless to respond to a dramatically changed world. Capital spending on the PW4084 was locked in and many engines were already under construction for customers who suddenly no longer wanted them. What was more, the airlines were sending a clear signal that they now wanted low-cost rather than high-performance engines for the 1990s, designs that could not be ready for years.

The first half of 1991 had continued the record profits of 1990, but the turn in the market was breathtaking and Pratt was suddenly heading for a $1.3 billion swing in its operating results within a year, culminating in a $283 million loss for 1992. As Coran remembers, "Very suddenly, just when I arrived, everything that could go wrong went wrong. Rather than a simple cost-cutting exercise to deal with a 10 percent drop in volume, I realized that we needed to rethink the whole business."

Fortuitously, just at the time of the crisis, several key executives in the UTC group—including Coran, George David, the president of UTC's Commercial and Industrial Group, and Karl Krapek, the president of Carrier—had become familiar with lean principles, mainly from the accident of being located in Hartford where Art Byrne was working steadily to apply them. In addition, Coran had a major advantage. He had never had an operating job prior to arriving at Pratt and therefore had none of the biases of the traditional mass-production operating executive. He therefore re-

The attempt to do so represents the acid test. If Pratt can apply these principles quickly in a massive, publicly traded, high-tech organization with extraordinarily deep technical functions and life-or-death demands on product quality, *plus* all of Wiremold's problems, then literally any American firm can.

From the American System to Mass Production[3]

Pratt provides a wonderful example of the mass-to-lean conversion because the firm was centrally involved in creating the very mass-production system which eventually threatened its survival. What's more, it twice went through the progression from flexible start-up to stuck-in-the-mud mass producer that we saw at Lantech.

The original Pratt & Whitney Company was created before the American Civil War by Francis Pratt and Amos Whitney. These "Yankee mechanics" learned their trade as inside contractors at Samuel Colt's armory, opened in Hartford, Connecticut, in 1855. They produced the individual parts needed for Colt pistols and rifles, hiring their own workforce but using Colt's plant and tools.

Of central importance to our story, Pratt and Whitney also built many of the four hundred machine tools and the gauges Colt needed to achieve his goal of totally mechanized gun production in which parts were interchangeable and handwork for "fitting" was eliminated.[4] This approach became known as the American System, in comparison with the European System in which parts were individually handcrafted, with each part fitted to those already in place to create a completed product.

When Pratt and Whitney left Colt in 1860 to establish The Pratt & Whitney Company, they took with them a fundamental set of ideas about manufacturing practice which dominated the company until very recently. They believed that best practice called for the creation of special-purpose machines able to perform specific operations on specific parts, if possible at high speeds in high volumes. They further believed that machines performing similar types of tasks should be grouped together in departments and that simple logic called for setting up a machine to make a given part and then making a batch of them before setting up the machine for the next part. In other words, they built the precision machinery needed for the familiar world of batch-and-queue and, over time, organized their own factory in accord with these principles.

Over the next sixty-five years, Pratt & Whitney grew from a small workshop under the direct management of the two founders into a massive and highly successful organization. In its many departments focused on specific parts needed for lathes, grinders, millers, cutters, and borers for metalworking industries. The firm also pioneered extremely precise gauges to check the accuracy of parts and sold these along with their tools. Over the years, Pratt's machines became more complex and capable of more delicate and sophisticated tasks. In addition, advances in metallurgy made it possible to work prehardened metals so parts could be made to net shape without fear that subsequent hardening steps would interfere with interchangeability. However, the basic philosophy of production did not change.

The Rise of the Eagle[5]

In the summer of 1924, Frederick Rentschler resigned as president of the Wright Aeronautical Corporation in New Brunswick, New Jersey, because the bankers investing in the firm would not back his idea for a radial, air-cooled engine much larger than the revolutionary Wright Whirlwind just entering production.[6] He believed this large engine would swing the military away from liquid-cooled designs and make commercial aircraft economically viable for the first time.

With the encouragement of the U.S. Navy, Rentschler sought new financial backers and early in 1925 contacted Pratt & Whitney in Hartford, which was experiencing a slump in its business and found itself with excess plant space and tools. In addition, Rentschler noted that the Hartford area was full of Yankee mechanics skilled in operating the types of tools Pratt produced, precisely the tools needed for aircraft engine manufacture.[7]

Rentschler proposed to play a similar role at Pratt & Whitney to that Francis Pratt and Amos Whitney had played seventy years earlier at Colt's armory. He outlined a plan to set up a company within a company, using P&W's long-established name with its worldwide reputation for precision machinery. He proposed borrowing a million dollars from P&W's owners (in return for giving them 50 percent of the stock in the new Pratt & Whitney Aircraft Company)[8] and using Pratt's underutilized plant space and tools to make his new engine. An agreement along these lines was reached in July of 1925, and Rentschler was back in the aircraft engine business.

In 1925, aircraft engine design was still a trial-and-error process of building a prototype and testing it to failure, then strengthening the part that failed and testing the design again. Rentschler knew that the key to success was to attract the most experienced engineers in the industry and to quickly create a scaled-up version of the Wright Whirlwind that would work well on the first try. He soon convinced several senior engineers from Wright to join him at Pratt, and his new design team made spectacular progress.

total payroll of thirty including Rentschler) were able to design the new Wasp engine (with about two thousand parts), incorporate a key processing innovation to save weight,[9] build three prototypes, and have them ready for testing by potential buyers. When tested, the Wasp produced 50 percent more power (425 horsepower) than the Wright Whirlwind air-cooled engine and weighed only 650 pounds compared with the 1,650-pound Curtiss Liberty liquid-cooled engine producing the same horsepower. (The latter engine was the standard design used by the U.S. military at that time.)

Orders from both military and commercial customers poured in, and by 1929, Pratt & Whitney was the world leader in the tiny but rapidly growing aircraft engine business. The Pratt engine quickly established a reputation for reliability and was chosen for the next generation of commercial transports, beginning with the Ford Tri-motor. (The corporate logo—an American Eagle encircled by the words "Pratt & Whitney—Dependable Engines" —was affixed to every engine from the beginning and became familiar to airline passengers around the world.) In 1929, Rentschler was able to buy out the Pratt & Whitney machine tool company's interest and build a new headquarters and vast production facility in East Hartford.[10]

In the beginning, the three key activities of Pratt & Whitney—design of new products, order-taking, and production—could be accomplished effectively in an utterly simple organization. Indeed, the initial production run of two hundred Wasp engines for the U.S. Navy was designed and then built in one large room by a group of highly skilled machinists directly interacting with the tiny group of product engineers.

By the early 1930s, as production volumes grew from dozens of engines to hundreds, organizational differentiation like that undertaken by Lantech seemed to be required. Departments were created for each major activity—sales, engineering, prototype building and testing, quality control, purchasing, production, and service. Shops were created inside each department for specialized activities; for example, heat treatment, paint, and final assembly shops were established within the Production Department. As long as Pratt had only one product in development (the Hornet, which followed the Wasp and increased horsepower to 500) and only the Wasp in production, this system worked well without the need for cross-functional management.

However, by the mid-1930s, as Pratt expanded its product offerings to include the 300-horsepower Wasp Junior and the 800-horsepower Twin Wasp, and conducted experiments with a range of new engine configurations, something more was needed. A new position was created, the "project engineer" reporting to the heads of engineering and production. The project engineer was given the job of coordinating all of the activities involved in the design, production, and installation in the customer's airplane of a departments and shops.[11] The project engineer was only a coordinator with no dedicated employees or resources—in today's terminology, a "lightweight" program manager[12]—but a startling conceptual leap had been taken, going far beyond a purely functional organization and common management practices at that time. Indeed, the concept of a project engineer to oversee the entire value stream foreshadows the lean principles described in this book.

As Pratt grew in the 1930s, changes were required in the factory as well. Initially, all of Pratt's metal-cutting tools had been relatively small machines—lathes, drills, millers, borers, etcetera—which could be lined up in the actual sequence of work flow.[13] For example, in 1936, the Cylinder Shop in the East Hartford plant was organized as follows:

> ". . . the first shop . . . immediately following the raw material inspection and the Experimental Department is the Cylinder unit. On one side of the main aisle are produced all of the steel cylinder barrels. On the other side are produced all of the aluminum alloy heads and in addition, the barrels are assembled to the heads, together with valve seats, bushings, valve guides and other minor parts, so that when the cylinder is ready to leave the department, it . . . proceeds directly to the Finished Stores Department. . . . with spare parts requirements, there are approximately 50 separate active cylinder designs. The equipment has been laid out so that the machines are in a sequence and the raw material proceeds in a straight line. Naturally, not all machines are required for any given cylinder."[14]

Similar shops had been created for master links and rods, crankcases, crankshafts, pistons, rocker shafts and valve guides, and cams. These sound remarkably like the work cells for complete components we have encountered throughout this volume and it is clear that Pratt's operations managers at that time had at least a rudimentary notion of flow: ". . .the scheme of production is a relatively simple one. Raw material is received by rail or truck through the front of the shop [factory] and then flows through the various manufacturing departments to the Finished storeroom at the rear."[15]

However, it is also clear that continuous flow was strictly limited to assembly and those activities which could be performed with simple machines. Special shops were created for machining parts made from magnesium and hard steel alloys as well as for heat treating, painting, and polishing. Because most of the parts in each completed component needed at least some of these treatments, much material was moved back and forth from shop to shop.

In addition, an elaborate system of centralized storage areas, tool cribs, and inspection stations was in place. It was taken as a given that quality inspections must be done independently of the primary workers by technicians reporting to the head of that function, not to the head of production, and that production could be more tightly controlled by storing tools, fixtures, and parts-in-process in a central location. These decisions meant that every part and every worker moved to a central storage area between each major production stage and during setups for the next job.

Finally, the philosophy of the company was that many defects could only be detected by test running of completed engines. Therefore, a row of test cells went across the entire rear section of the factory. Each engine was run for eight to thirteen hours, then completely disassembled. The parts were inspected, replaced as necessary, and reassembled. The engine was run for five to twelve more hours and then, in the event no problems were found, it was shipped.[16] As we will see, this final safety net created an "assemble it, then tinker until we get it right" mentality which persisted at Pratt until 1994.

Even with a relatively simple plant layout and product line, it is clear that, by 1936, Pratt was having to work very hard to get products through the system. An organized system of "shortage lists" and "follow-up" (read "hot lists" and "expediting") was in place and the assistant general manager was eager to tell an audience of peers that "high-tech" help with these tasks was already in place:

> It might also be of interest to point out that all the shortage lists and schedule sheets are made on electrically operated Hollorith machines[17] from punch cards made in the store room office so that these lists are printed and supplied to the Schedule Dept. and Follow-up Dept. in a neatly printed segregated form and without any delay. This is a major factor in the efficient control of shop production.[18]

In short, Pratt & Whitney was for the second time moving down the path from a lean workshop to a massive mass producer. The major innovation during the second transition was that the growing emphasis on complex tools housed in specialized departments could be supported by automated information management to shepherd products from raw materials to finished goods.

What should have been the major organizational innovation, the project engineer, never worked as planned. By 1939, Chief Engineer L. S. Hobbs was writing to his superiors, "It has been fairly obvious from the time of our institution of the Project Engineer system that in reality the system has not functioned as such."[19] Instead, the project engineer was a lightweight manager within the product development organization and products moved through sales, scheduling, production, and installation as best they could, with expediting by the centralized information management system but with no individual or team fully responsible for their progress.

World War II as the Engine of Mass Production

When the flow of orders increased from hundreds to hundreds of thousands in World War II,[20] Pratt made the final leap to mass production in the factory. A shortage of skilled workers meant that the new machine tools for the war effort were designed for very specialized tasks with only modest skill requirements by the operator. The number of shops, each assigned a narrow task, grew dramatically as the division of labor continued. What's more, the volume of orders meant it was often feasible to dedicate a given machine to a given part, perhaps for years at a time, so the need to do frequent setups was reduced. Work-in-process, travel within the production system, rework in the test department at the end of production, and managerial complexity all increased but engine output increased even more, and the latter was the only important consideration during the war.

Not surprisingly, by the end of the war, the mentality of the workforce had changed. Rather then being highly skilled, semi-independent craftsmen, the new workforce was much more narrowly trained, assigned to largely interchangeable jobs, and under much tighter management control. A conventional union had little appeal to the initial generation of craftsmen at Pratt, but by 1945, a different mentality and a different shop-floor reality created an environment in which the International Association of Machinists easily won an election to unionize the workforce.[21] A maze of work rules and grievance procedures soon followed as a mirror image of the division of labor instituted by management.

The second important consequence of World War II was in product development, where the growing complexity of designs and the need to extract ever more power from the basic radial engine configuration created the need for very deep technical functions. The key disciplines were materials scientists to develop new materials, structural engineers to address weight and durability problems, aerodynamicists to tackle the problem of airflow and drag through and around the engine, and mechanical engineers able to design and link together the thousands of individual parts required for each engine. Each of these specialties gained its own department within the vast Pratt & Whitney Engineering Division.

By the end of the war, Pratt's Wasp Major engine had thirty-six cylinders turbocharged to yield 4,600 horsepower (compared with the 425 horsepower of Pratt's original nine-cylinder Wasp). Along with the turbo-compound engine being developed at the same time by the Curtiss-Wright Company (the merged successor firm to Wright Aeronautical and Curtiss), the Wasp Major was one of the most complex pieces of purely mechanical apparatus ever devised.[22]

The Jet-Propelled Eagle

During World War II, the U.S. government directed Pratt and Curtiss-Wright to stick to what they knew: designing and building reciprocating piston engines. Other American firms with no previous experience in aircraft engine building (General Electric, Westinghouse, and Allison) took the lead in jet engines, and by war's end Pratt was the clear world leader in a technology with no future. What was worse, it was nowhere with the technology that did have a future—the jet turbine.

In 1946, P&W took a tremendous but unavoidable gamble by abandoning research on piston engines. It attempted to leapfrog its new jet-age competitors with a two-shaft, axial-flow jet engine considerably larger and more complex than any previously envisioned. Curtiss-Wright, by contrast, continued to elaborate the piston engine with its turbo-compound version for the Douglas DC-7 and the Lockheed Super Constellation in the early 1950s. C-W exited the industry when jet aircraft quickly supplanted these final iterations of the piston-engine airplane.

Jet engines were based on different principles but required many of the same technical skills in Pratt's existing engineering functions. The materials scientists were now concerned with managing the extreme heat in the hot parts of the engine. The structural engineers were concerned with vibration in complex turbo-machinery. The aerodynamicists were concerned with airflow past the compressor and turbine blades. The mechanical engineers were still concerned with detailed design of the thousands of parts, now rotating rather than reciprocating, aggregating to a complete engine. The big difference was that the nature of the knowledge was now highly scientific and the amount of effort required was much greater.[23]

Pratt's technical functions became deeper and more silolike as the nature of the necessary knowledge became more arcane. The project engineer system within product development groaned as the walls between functions thickened, giving rise to the "Pratt Salute" of arms crossed and pointing in opposite directions to assign fault to other departments for all design and manufacturing problems.

The production system, for its part, was remarkably unaffected by the jet special-purpose devices such as electron-beam and fusion welders—were located together in shops inside departments to feed batches of parts to a bench-assembly operation creating the finished engine. Every engine was then extensively tested and "tuned" (reworked) before shipment. The common joke was that the average part traveled farther inside Pratt's plants during production than it did in airline service. But there seemed to be no better way.

Pratt's leap to jet propulsion in 1946 produced a technical and commercial triumph by 1952. The P&W J-57 engine powered the American eight-engine B-52 bomber first flown in that year. Slightly modified and renamed the JT3, this engine captured 100 percent of sales for the initial versions of the four-engine Boeing 707 and Douglas DC-8 by the end of the decade. P&W quickly followed up with an entirely new engine, the JT8D, to power the entire world fleet of Boeing three-engine 727s and two-engine Douglas DC-9s and the initial versions of the two-engine Boeing 737. When the American military awarded Pratt a contract in 1970 as the sole source of the F100 engine for the F15 and F16 fighter planes, the company totally dominated the global aircraft engine business. Indeed, at the end of the 1960s, Pratt held a staggering 95 percent share of the world's commercial jet engine market (outside the Russian bloc) and nearly a 50 percent share of American military orders.

In the process of reaching industry dominance, Pratt and its organization fine-tuned and hardened the standard features of a mass producer. Tasks were finely divided in physical production, with specialized machines making batches of parts with long lead times. During product development lightweight team leaders coordinated engineering efforts across thick functional walls.

In fact, this system was adequate if not perfect for its environment. For decades aircraft engines were ordered by regulated airlines—competing on service but not on price—and by the military—interested in wartime performance with purchase price only a secondary consideration. In addition, advances in materials science and aerodynamic analysis meant that each new generation of product could achieve substantial performance improvements. As long as Pratt's technical depth could produce products which performed better than competing products, the fact that they took unnecessary time to design and manufacture, cost more than they needed to, and sometimes failed to perform properly when first launched in service could all be overlooked.

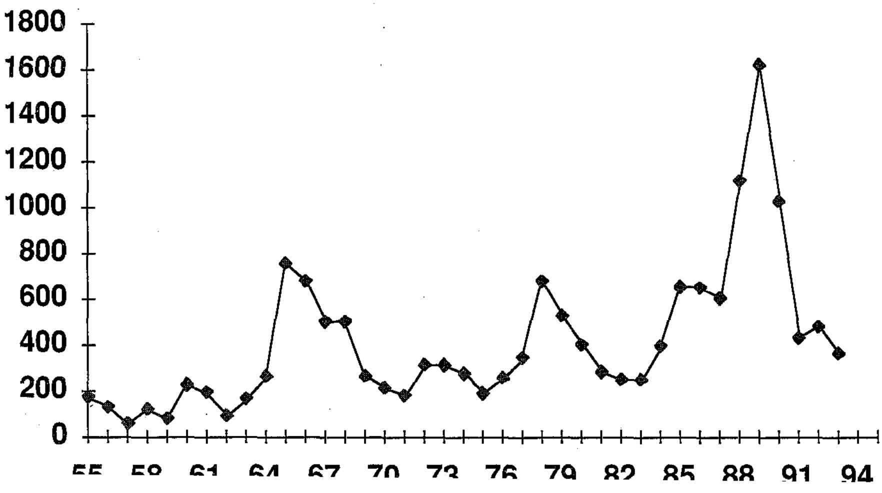
During this golden age the specification of new products at Pratt tended to work backwards. The senior engineers decided what technologies were ready for introduction in the next product generation and specified the duction cost and selling price as a sort of resultant. Once in production, costs were not rigorously tracked, but instead rolled up in the profit-and-loss statement in the president's office, by which point it was too late to do much about them.

By the 1980s, as airframe makers began to offer a choice of two or three engines (from Pratt, GE, and Rolls) for each wide-body aircraft type, the issue of production costs was confused further by the industry practice of progressively discounting prices for new engines, eventually to far below costs.[24] This was done in the hope that profits could be recovered from sales of spare parts, in particular turbine blades, where the engine makers had a monopoly. For example, the spares purchased by an airline during the operating life of a JT8D were likely to equal five times the initial purchase price of the engine. In this environment, the manufacturing side of the jet engine companies could easily get confused about the importance of costs —after all, the engines were being sold for prices far below any imaginable production cost.

The final feature of this mature mass-production system was its peculiar method of order-taking. The twenty-four-month lead times needed to physically produce an engine conspired with the three-year lead time needed to produce the complete airplane to create gigantic waves of orders for jet aircraft in the postwar era,[25] as shown in Figure 8.1.

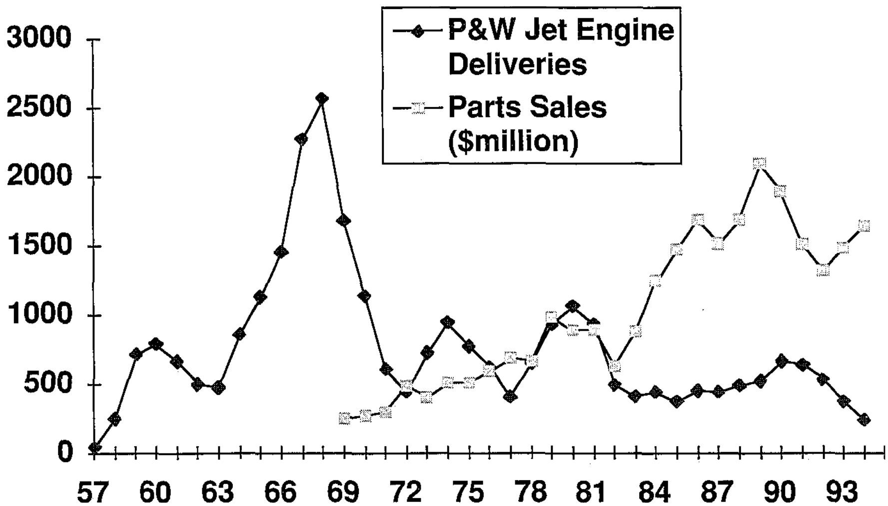
As the airline industry emerged from recessions, aircraft customers signed up for planes and engines they might not need in order to ensure themselves a place in the production queue, while sales departments often made special

FIGURE 8.1: COMMERCIAL JET ORDERS



deals for large orders even when sales were booming in order to hold market share and protect the spares base. These orders could suddenly evaporate when the economy slumped, but waves of military orders often offset slumps in civilian demand and spare-parts purchases often went up when new engine deliveries slumped after 1980, as shown in Figure 8.2.

FIGURE 8.2: JET ENGINE DELIVERIES & PARTS SALES



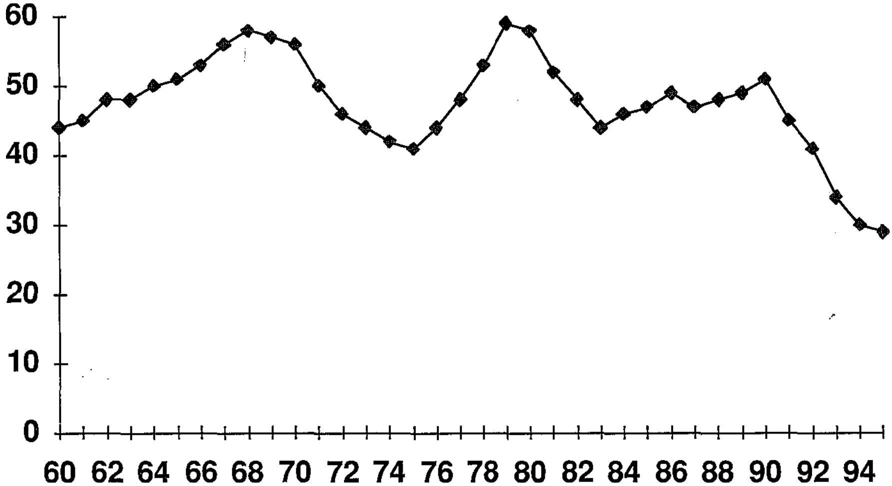
In consequence, employment at Pratt was more stable than orders until 1990, as shown in Figure 8.3. There were periodic layoffs, but these were likely to be short and it was easy for Pratt employees to think they would always have a job, particularly if they had a few years of seniority.

When the Eagle First Came to Earth

Big companies like IBM, General Motors, and Pratt usually receive (but ignore) a number of warnings that the world has changed before the roof finally caves in, and the collapse of both the military and civilian markets in 1991 was not Pratt's first wake-up call. That came in 1984, when Pratt so infuriated military customers with its failure to fix operational problems with the F100 engine that GE was brought in as a second source of supply and given roughly half the U.S. military's business for F16s.[26]

At the same time, the launch of Pratt's PW2037 engine for the Boeing 757 infuriated airline customers. The Pratt engine's fuel consumption was

Figure 8.3: Employment at Pratt & Whitney (000s)



superior to that of the competing Rolls-Royce RB211-535 and pricing was competitive, but the Pratt engine had a terrible record of mechanical problems, causing flight cancellations when introduced into service. As Fred Hetzer, the project engineer on the PW2037, remembers, "We were like the aging slugger in baseball who can still see the ball clearly but can't swing the bat fast enough to hit it. We knew about the problems in the PW2037 a year before they surfaced with commercial customers, and we worked day and night to fix them, but the organization was so sluggish and cross-functional communication was so difficult that we just couldn't get them fixed in time." As a result, Pratt had a superior engine ready first but wound up with only half the business in the forty-thousand-pound-thrust class.

Finally, Pratt badly misjudged the trend of demand in the jet engine market. Thinking that large, double-aisle aircraft were the primary growth market and reluctant to compete against its currently best-selling JT8D, Pratt failed to develop a replacement engine for the JT8D powering the 727s and 737s. When Boeing decided in the early 1980s to modernize the 737 by lengthening the fuselage to carry more passengers and updating its systems, Pratt did not have an engine with modern, high-bypass technology and lower specific fuel consumption. A consortium formed by GE in the United States and Snecma in France (CFM) did and ran off with most of the business for what became by far the world's best-selling airplane. When Airbus introduced the A320 to compete against the 737, 100-to-160 passenger, single-aisle jets became by far the largest aircraft market segment.[27]

Leaner but Not Lean; Necessary but Not Sufficient

Suddenly, in the mid-1980s, Pratt faced competition in all its major product categories and its market share began to slip across the board. In addition, total engine deliveries in the industry began to fall due to the shift from four- to two-engine designs. Pratt's management was not completely asleep and three innovations, which seemed earthshaking at the time, were introduced in response, one in production and two to bridge the chasm between product development and production.

The major innovation in the physical production system, introduced in 1984, was the "focused" factory with flow lines and business units organized by categories of parts. Pratt's factory structure emerging from three hot wars (World War II, Korea, and Vietnam) and one cold war was a hodgepodge of isolated shops working on parts with no relation to the part being made in the next shop. In one notable case, the distance traveled by a part within Pratt plants (not counting the distance traveled between plants) was measured and found to total eighteen miles.

In 1984, Pratt reorganized its facilities so each would take responsibility for a major category of engine parts. The massive North Haven plant would work primarily on turbine blades while the Southington plant would work primarily on rotors and discs and the Middletown plant would take on all final assembly work. Within each plant, activities were further reorganized so that many of the steps in the physical processing of each category of part[28] were grouped and lined up in a logical progression in a "flow line," insofar as tool designs would permit. Note that this is exactly the concept described in 1936 by Carlton Ward, the assistant general manager of production at Pratt.

Finally, each part category—for example, high-temperature turbine blades for the JT8D engine—was placed in a "business unit" whose leader knew the cost for his operations. The business unit head was fully in charge of getting the parts made at cost and on time, in accordance with the master schedule (now run off a massive computerized Material Requirements Planning system).

By the mid-1980s, Pratt's senior managers were aware that as the jet engine matured it was becoming sensible to apply similar design principles to "standard" design problems confronting each category of part. For example, why not specify the same grade of chromium for each high-temperature turbine blade, rather than fiddling endlessly with minor changes in the alloy mixture, which produced negligible improvements in performance? Yet it was apparent that Pratt's design engineers working on each category of part were doing the exact opposite. They were doing what comes naturally in