# Beyond Lean Manufacturing: Combining Lean and the Theory of Constraints for Higher Performance

==========

***H. William Dettmer***

Senior Partner
Goal System International
Port Angeles, WA, USA 98362

## Summary

Lean manufacturing (LM) evokes images of efficiency and minimizing unnecessary costs, an attractive value for many companies. But other companies are already as lean as they can be. Does LM have practical limits? This paper demonstrates how the Theory of Constraints (TOC) can take the business performance of those organizations currently pursuing lean to the next level.

## Keywords

Lean thinking, lean production, lean manufacturing, *muda, kaizen,* 5S, *takt,* workload balancing, one-piece flow, Theory of Constraints, drum-buffer-rope, five focusing steps, Throughput, Inventory, Operating Expense, systems approach

## INTRODUCTION

*See my people, well here's my theory*
*Of what this country is moving toward.*
*Every worker a cog in motion,*
*Well, that's the notion of Henry Ford!*

*One man tightens, and one man ratchets,*
*And one man reaches to pull one cord.*
*Car keeps moving in one direction,*
*A genuflection to Henry Ford.*

*Hallelujah! Praise the maker*
*Of the Model T*
*(Speed up the belt, speed up the belt, Sam!)*

*Hallelujah! Hell, I'll take her!*
*Sure amazin' how far some fellas can see!*
*(Speed up the belt, speed up the belt, Sam!*
*Speed up the belt, speed up the belt, Sam!)*

*Mass production will sweep the nation,*
*A simple notion the world's reward.*
*Even people who ain't too clever*
*Can learn to tighten a nut forever,*
*Attach one pedal or pull one lever*
*For Henry Ford!*

*Grab your goggles and climb aboard!*
"Henry Ford"
From *Ragtime*[‡]

## MASS PRODUCTION

Every so often, commercial business undergoes a paradigm shift: a change in the nature of the game so fundamental that the old rules of operation and competition no longer apply. Henry Ford wrought such a change in the first decade of the 20th century by creating a manufacturing concept he called *mass production.*[‡‡] The essence of mass production is captured in Lynn Ahrens' lyrics, above, written for the stage play *Ragtime.*

Until the early 1900s, all complex machinery, including automobiles, was hand-built by skilled craftsmen. These craftsmen were expert in all aspects of manufacturing, and craft-based companies were able to give their full attention to individual consumers. **(30: p.25)** But this benefit to the consumer came hand-in-hand with a major drawback for the craft-based company: production costs were high and didn't drop with volume increases. In the case of cars, high costs meant that only the rich could afford to buy them. This, in turn, meant that the potential market for craft-built cars—or other craft-type products—was extremely limited.

Henry Ford's mass production concept brought down the costs of production so dramatically that it made mass consumption possible. **(30: p.100)** It represented a quantum improvement in process quality as well. No two craft-built cars were exactly alike. Craftsmen filed, trimmed, and shimmed components to get them to fit together. Sequential fitting of parts produced "dimensional creep," resulting in cars that looked identical but whose parts could not be interchanged. **(30: p.22)**

Ford, on the other hand, required complete and consistent interchangeability, which made attaching parts to one another a fairly simple process. It was *this* characteristic of mass production that made the moving assembly line possible. **(30: p.23)** ("Speed up the belt, speed up the belt, Sam!") To achieve interchangeability, Ford insisted on standardized gauging for all parts and pre-hardening metals before fabrication. **(30: p.23)**

There were two other noteworthy characteristics of Henry Ford's revolutionary concept

---

[‡] Stage play based on the novel *Ragtime*, by E.L. Doctorow. Lyrics to "Henry Ford" by Lynn Ahrens.

[‡‡] Henry Ford didn't actually use the term "mass production" until 1926. Prior to that time, many referred to his concept as "Fordism."

worth mentioning. The first concerns the work force. In craft-based industries workers were highly experienced, skilled, multi-talented decision makers. They truly "owned" their processes. They decided what to do and how to do it. Ford's assemblers, on the other hand, had only one simple task each: put two nuts on two bolts, or attach one wheel to each car. **(30: p.31)** ("One man tightens, and one man ratchets, and one man reaches to pull one cord.")

Ford's assembly line workers didn't have to think. They didn't have to order parts or tools, repair equipment, inspect for quality, or even understand what those beside them were doing. In fact, many of them couldn't even speak the same language as their supervisors. All they had to do was keep their heads down and continuously repeat the same mind-numbing task over and over. So, while mass production made mass consumption possible, it made factory work barren. **(30: p.100)** Is it any wonder that workers didn't care about looking for product quality deficiencies, or about fixing them when they did see them? The ultimate effect of passing poor quality through was a large rework effort at the end of the assembly line—often consuming 20 percent of plant floor space and as much as 25 percent of total labor effort. **(30: p.57)**

The second significant characteristic of mass production worth noting is its physical infrastructure. It uses a lot of expensive machinery. Historically, managers have justified and amortized the costs of production facilities by building huge volumes of standardized products in long production runs. To succeed, this approach demands stability. It's intolerant of disruptions, whatever the cause. Consequently, mass-producers add many buffers—extra supplies, extra workers, and extra space—to assure smooth production. **(30: p.13)** But all of these buffers add tremendously to fixed and inventory costs, which are typically allocated to units of product produced, meaning that even longer production runs are needed to amortize them.

We would be remiss in failing to mention the contributions of Alfred P. Sloan. Henry Ford is undoubtedly the "father of mass production," but Sloan at General Motors matured it—made it the complete system we ascribe to the term today. In doing so, Sloan succeeded where Henry Ford did not: he devised the organization and management system needed to establish control over the total system of factories, engineering operations, and marketing systems that successful mass production demanded. **(30: p.40)** Moreover, Sloan did so across General Motors' five-model product range,[‡] rather than just a single product line (Ford's Model T).

## THE PARADIGM SHIFT

Fewer than 20 years were required for mass production to supplant craft industry as the preferred approach to production. Over the next 70 years, mass production entrenched itself in manufacturing and is still the philosophy of choice for the largest percentage of industrial companies around the world. But to paraphrase Marx's commentary on capitalism, mass production contains the seeds of its own destruction. In a world where changes in market demands and technology advances come quickly, mass production exhibits all the maneuverability of the Exxon Valdez. Originally, Henry Ford was looking for a simple, efficient solution to the challenge of producing automobiles in high volumes while making them

[‡] Chevrolet, Pontiac, Oldsmobile, Buick, and Cadillac

affordable. What evolved 30 to 50 years after Model T mass production was a costly, inflexible behemoth. From 1955 on, mass production was a paradigm ripe for replacement, and that replacement has been brewing for the past 20 years. The only question was: what would replace it?

In the last half of the 20th century, two management philosophies emerged with the potential to provide more of mass production's benefits—as well as new ones— without its inherent drawbacks. These philosophies are the Toyota Production System (TPS) and the Theory of Constraints (TOC).

## Toyota Production System

That the Toyota Production System is better known than the Theory of Constraints is more a matter of birthdays than merit. Both TPS and TOC are capable of delivering results far superior to those of traditional mass production. But Eiji Toyoda and Taiichi Ohno were creating the Toyota Production System when Eliyahu M. Goldratt, the father of the Theory of Constraints, was only three years old. By the time Goldratt formulated constraint theory, the TPS had reached maturity in Japan. And it occupied stage center in the U.S. by the mid-1980s, owing to the shellacking Japanese products were administering to American products in their own domestic markets. Mass production had evolved into such a ponderous, wasteful way of doing business that Toyoda and Ohno didn't have to look hard to find a place to start. They began eliminating waste from the production process with a vengeance, starting with waste that was obvious and moving later to waste that was hidden. Their philosophy posed two questions, "Where is the waste in the manufacturing system?" and "What's the best way to get rid of it?"

However, in America the TPS hasn't always been embraced as a unified philosophy. Instead, many U.S. companies have accepted its components in a piecemeal fashion, adopting some of its aspects and methods while ignoring or rejecting others. Since the Toyota Production System was not generally known by that name in America, other terms such as statistical process control, concurrent engineering, cause-effect analysis, five why's, team work, supplier/supply chain management, horizontal integration, and just-in-time gained wider recognition instead. And the collection of these (and other) tools came to be generally known as "total quality management" or "continuous process improvement." More recently, terms such as six sigma have joined the lexicon. Whatever you choose to call it, as a unified philosophy, it originated as the Toyota Production System, and all of these components were elements of Toyoda's and Ohno's success.

The term "lean production," first used in *The Machine That Changed the World* **(30)**, was coined by a member of the International Motor Vehicle Program (IMVP) team at the Massachusetts Institute of Technology (MIT). It was a collective synonym for the Toyota Production System. The IMVP team completed a five-year international research study culminating in the book that introduced the term "lean" to the industrial world. The study compared the mass production system created by Henry Ford, extended exponentially by Sloan at General Motors, and practiced by virtually every major industry in the world up to that time (except Toyota), with the production system invented by Toyoda and Ohno—which the IMVP team dubbed "lean production." So, in essence, lean production *is* the Toyota Production System. What has come to be known as "lean manufacturing," as we'll see later, is not quite the

same thing.

## Theory of Constraints

In the early 1980s, about the time that America was waking up to the impact of the TPS on manufacturing businesses, Eliyahu M. Goldratt was simultaneously examining manufacturing processes from a completely different perspective. Goldratt, a physicist by education, naturally saw manufacturing as an integral part of a larger system, as did Toyoda and Ohno. But Goldratt arrived at the same end (higher-performing business systems) by asking a different set of questions.

Considering his background, it shouldn't be surprising that Goldratt should see systems—manufacturing or otherwise—in terms of physics. Archimedes once said, "Give me a lever long enough and a place to stand, and I will move the earth." Goldratt believed that each system contained *leverage points*—critical places where force could be applied and do the most good. ("Force," in this case being an illustrative, not a literal term.) In other words, changes at these leverage points would deliver a positive (or negative) effect on overall system performance considerably out of proportion to the magnitude of the change effort. Goldratt referred to these leverage points as *constraints*, because inaction at these locations prevented the system from realizing better performance in relation to its goal.

Initially, Goldratt began with internal resource constraints in manufacturing systems. Much as Toyoda's and Ohno's early efforts expanded into a whole-system methodology, Goldratt's vision expanded to encompass system elements beyond manufacturing processes alone: supply chains, distribution, sales and marketing, and product development (engineering).

Goldratt simply began with observable indications that something is wrong in a manufacturing company (losing money or inadequate profits; mountains of unsold inventory; mediocre sales; marginal customer satisfaction, etc.). He considered these to be undesirable effects—the unavoidable outcomes of deeper root causes—connected by networks of cause-and-effect. The complexity of the cause-and-effect network was proportional to the complexity of the interdependent relationships within the manufacturing system. Goldratt hypothesized that these deeper root causes prevented better business performance (e.g., more sales, more profit). Viewing a business as a whole "flow" system, he proposed that, much like a garden hose with a crimp in it, some kind of constraint served to "bottleneck" the flow of work through the business to the customer.

If one accepts this concept of a system, then even if the hose has more than one crimp or bottleneck in it, one probably constricts flow more than all the rest. For Goldratt, therefore, system improvement is an exercise in getting significantly more out of the system, rather than merely doing the same or somewhat more with less. And to obtain more, Goldratt prescribed a repeating cycle of finding and breaking the constraints of the system sequentially, beginning with the most restrictive one first. Goldratt wrote about this concept in two books, *The Goal* (1986, 1992) and *The Haystack Syndrome* (1990). These books have become classics in the business genre, with *The Goal* having sold over 3 million copies in 13 languages. The first 100 pages *The Haystack Syndrome* contain the most succinct explanation of constraint theory and how it applies to manufacturing operations. We'll see the specifics of the Theory of Constraints a little later.

Both methodologies—lean production and the Theory of Constraints—have a high degree of congruence. And where they don't coincide, each offers some strengths and benefits not provided by the other. More, each fills gaps or overcomes deficiencies inherent in the other. The essential difference between the two is that lean production seeks to eliminate waste in all parts the system (the value stream) simultaneously, principally through teamwork. Constraint theory suggests that, while waste elimination might be necessary, it might not be the best thing to do first to achieve an immediate quantum increase in business performance. Rather, TOC seeks first to identify the system constraints (leverage points) in the sequence that will produce the biggest, quickest benefit, and select the appropriate strategy to apply pressure to those points immediately.[‡]

## Combining TPS (Lean) and TOC

The overarching premise of this paper is that a hybrid of the two philosophies is potentially more robust—more productive, easier to implement—than either one alone. One word of caution, however: melding the two is as sensitive as transplant surgery. It craves careful grafting, and you can't selectively decide to adopt some aspects but not others. The balance of this paper will compare lean production with constraint theory and propose how to integrate the two.

## THE SYSTEMS APPROACH

The Toyota Production System—lean production—is much more than lean manufacturing alone. The whole lean concept is composed of five interlocking functions: product design and engineering, interaction with the customer (through marketing/sales, market/product research, etc.), the supply chain, manufacturing, and distribution. These are all interdependent parts of the same system, and ignoring any of them risks failing to achieve the level of success enjoyed by Toyota.

Yet this is what typically happens in adopting complex methodologies. People like shortcuts! Witness the mixed results of Total Quality Management (TQM) in the 1980s and 1990s. Various sources suggest a success rate of no more than 25-30 percent among companies that have attempted TQM.[‡‡] There is much emphasis in America today on applying the principles of lean thinking primarily on manufacturing, leaving the engineering, external supply chain, customer interaction, and distribution functions until later. This probably happens because a) manufacturing is fairly well defined in scope, b) it's easier to reach out and touch (control), and c) it usually promises immediate returns—a dollar saved is believed to go straight to the bottom line. Applying lean principles to engineering, marketing/sales, or other functions is perhaps a little less distinct, maybe not so easy to do, and takes longer to show bottom line

---

[‡] Opinions on the number of such constraints differ. Goldratt contends that there is only one at any given time. Others allow that there might be more than one. But everyone agrees that there are *very* few—maybe no more than just one at any given time.

[‡‡] Success, of course is in the eye of the beholder. Some might consider marginal improvements to be success. But most companies that embraced TQM did so with the expectation of replicating the "Japanese miracle" in their own organizations.

results. So efforts in these areas are sometimes deferred until later (and often ignored altogether).

At the same time, non-manufacturing functions—especially product development/ engineering and the interface with customers—are just as critical to successful lean production as manufacturing. Nonetheless, some companies seem to be content with a partial effort, and they're usually disillusioned when they don't achieve the expected results.

It's beyond the scope of this paper to tell the whole story of either lean or TOC. There's no shortage of other books (see the References) that will do this more thoroughly anyway. So for now, we'll confine our discussion to the key characteristics, principles, and prescriptions of both lean and TOC. The real "meat" of the discussion will be a comparison of the two and recommendations on how to integrate them for better, more immediate system-wide effect.

## LEAN PRODUCTION

Toyoda and Ohno built their philosophy from the bottom up. They talked and thought mostly about specific methods applied to discrete activities, such as engineering offices, sales groups, purchasing departments, and factories. **(29: p.10)** They even wrote books describing these details. Except for Ohno's memoirs, the integration of all these "bits and pieces" was largely left implicit, for the practitioners to figure out. Consequently, many managers feel as if they're drowning in techniques as they try to implement bits of a lean system without understanding the "big picture."

Toyoda's and Ohno's entire focus was on the elimination of the waste so prevalent in traditional mass production operations. They used the Japanese word, *muda,* which they defined as any human activity that absorbs resources but creates no value. *Muda* typically includes: **(29: p.15)**

- Mistakes requiring rectification
- Producing items no one really wants (causing inventories and remaindered goods to pile up)
- Unneeded processing steps
- Movement of employees and goods without purpose
- People downstream waiting because upstream activities haven't delivered on time, and
- Goods or services that don't meet the needs of the customer.

Womack and Jones (1996) summarized lean thinking in five principles: ‡ **(29: p.10)**

1. *Precisely specify value* by specific product
2. Identify the *value stream* for each product
3. Make value flow without interruptions
4. Let the customers *pull value* from the producer, and

‡ Specific actions to remove waste tend to happen in principles 3 and 5. Numbers 1 and 2 provide for waste identification.

5. Pursue *perfection.*

*Value* is defined by the customer. It's expressed in the characteristics of the product or service (or both) that the customer finds attractive. At a very basic level these may be no more than *reliability, maintainability,* and *availability*.‡ At a higher level, *value* could mean more "bells and whistles," multiple functionality, or attractive styling. The definition of *value* establishes product design objectives.

The value stream for each product might be considered the process steps required to bring a product through three critical management tasks: **(29: p.19)** (Figure 1)

- *Problem solving.* Concept through detailed design and engineering to production.
- *Information management.* Order-taking, through detailed scheduling, to delivery.
- *Physical transformation.* Raw materials to a finished product in the hands of the customer.

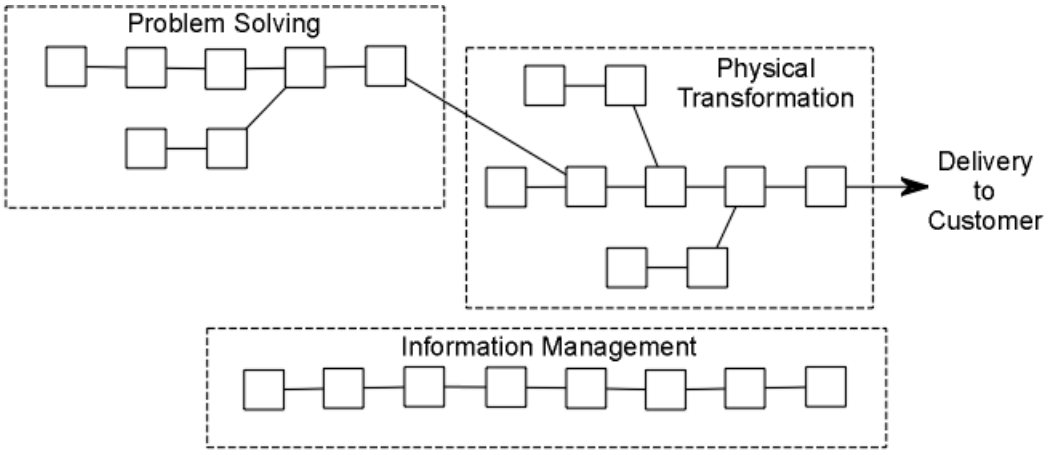



**Figure 1. The Lean Production Value Stream**

Obviously, many steps in the value stream create value. That is, they contribute directly or indirectly to creating the product characteristics important to the eventual customer. Other steps create no value, by this definition. They constitute *muda.* Some of these steps are unavoidable, even though they create no value. These are classified as Type 1 *muda.* **(29: p.20)** On the other hand, many steps that create no value are immediately avoidable. These are Type 2 *muda,* and they elicit most of the immediate attention of managers attempting to implement lean production.

Making value *flow* requires speed and consistency. The lean alternative redefines the

‡ *Reliability* implies that a product doesn't fail often. *Maintainability* means it's easy to fix when it does break, and *availability* means repair can be done very quickly, limiting out-of-commission time.

work of functions, departments, and firms. **(29: p.24)** The objective is to make work valued by the customer move through the system quickly and smoothly (i.e., without the starts-and-stops inherent in batch-and-queue operations).

*Pull* is a manufacturing philosophy based on synchronizing production objectives and rates with actual customer demand, rather than on forecasts or arbitrary finished inventory levels. **(29: p.24)** Ideally, *pull* approaches make-to-order. And effective *pull* system can achieve dramatic savings in both work-in-process and finished inventories. To succeed, however, a *pull* philosophy depends on exceptionally fast, smooth *flow.*

The final lean principle is "pursue *perfection.*" This implies that the first four principles are repeated in a continuous, ever-tightening cycle. Lean thinking maintains that there is no end to the process of reducing effort, time space, cost, and mistakes, while offering products that continually approach exactly what customers want. **(29: p.25)** "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." **(29: p.25)** [Emphasis added]

As an example of the pursuit of *perfection* in eliminating *muda,* Womack and Roos (1996) cite Pratt and Whitney (PW). **(28: p.26)** PW created a locally-designed U-shaped cell to replace a completely automated turbine blade grinding system. The new cell was installed quickly and at one-quarter the capital cost of the automated system it replaced. The new cell cut production costs by half, reduced processing times by 99 percent, and shortened changeover time from hours to seconds. PW can make exactly what the customer wants upon receiving the order. Womack and Roos contend that this application of lean thinking would pay for itself within a year, even if PW obtains only scrap value for the junked automated system. **(28: p.26)**

The implementation of the whole lean production philosophy is neither quick nor simple. Its inventors, Toyoda and Ohno, took 20 years to do it at Toyota. Even though the "template" for others to do it has been established, it's no easy matter to fundamentally change the way product design and engineering is done. Reconfiguring from a vertically integrated supply chain to a horizontal one isn't done quickly or easily, either. Even re-engineering the manufacturing operation is a major undertaking. And integrating all the facets may be the most challenging task of all. Not may companies—especially in North America—have the patience, determination, or persistence to follow through completely with a quest of this magnitude. So it's no wonder that companies gravitate almost exclusively toward the lean manufacturing component alone, especially since the large investment in facilities, equipment, and manpower represents fertile ground for recovering the costs of *muda.*

## Lean Manufacturing

Five primary elements are required to support the manufacturing component of lean production: **(9: p.4)** *manufacturing flow, organization, process control, metrics,* and *logistics.* On the manufacturing floor, work is divided into discrete cells based on natural groupings of related tasks. *Manufacturing flow* concerns the physical changes and design standards deployed as part of each work cell. *Organization* establishes people's roles and functions, and trains them in new ways of working and communicating. *Process control* includes efforts to monitor, control, stabilize, and improve discrete manufacturing process steps. *Metrics* involves

establishing visible, results-based performance measures, determining targets for improvement, and recognizing work teams for their process improvements. *Logistics* defines the operating rules and mechanisms for planning and controlling the flow of material. **(9: p.4)** Figure 2 indicates the basic tools and methods used to satisfy the requirements of each of these five lean manufacturing elements.

| MANUFACTURING FLOW | ORGANIZATION | PROCESS CONTROL |
|---|---|---|
| 1. Product/quantity assessment (product group)<br>2. Process mapping<br>3. Routing analysis (process, work, content, volume)<br>4. Takt calculations<br>5. Workload balancing<br>6. Kanban sizing<br>7. Cell layout<br>8. Standard work<br>9. One-piece flow | 1. Product-focused, multi-discplined team<br>2. Lean manager development<br>3. Touch labor cross-training skill matrix<br>4. Training (lean awareness, cell control, metrics, SPC, continuous improvement)<br>5. Communication plan<br>6. Roles and repsonsibilities | 1. Total productive maintenance<br>2. Poka-yoke<br>3. SMED<br>4. Graphical work instructions<br>5. Visual control<br>6. Continuous improvement<br>7. Line stop<br>8. SPC<br>9. 5S housekeeping |
| **METRICS**<br>1. On-time delivery<br>2. Process lead-time<br>3. Total cost<br>4. Quality yield<br>5. Inventory (turns)<br>6. Space utilization<br>7. Travel distance<br>8. Productivity | **LOGISTICS**<br>1. Forward plan<br>2. Mix-model manufacturing<br>3. Level loading<br>4. Workable work<br>5. Kanban pull signal<br>6. A, B, C parts handling<br>7. Service cell agreements<br>8. Customer/supplier alignment<br>9. Operational rules | *SOURCE: Feld, William M. Lean Manufacturing: Tools, Techniques, and How To Use Them. St. Lucie Press, 2001, p. 5* |

**Figure 2. Methods and Tools of Lean Manufacturing**

Let's look at the five elements in Figure 2 a little more closely. The methods listed under *Manufacturing Flow* and *Logistics* are very much unique to flow of work under the lean concept. Most of those under *Organization* and *Process Control* are probably familiar to most readers as techniques promoted in throughout the world under total quality management or continuous improvement (TQM/CI) programs since the 1980s. *Metrics* include indicators related to both workflow and quality. In essence, these methods support three basic objectives of lean manufacturing: make only quality stuff, do it fast, and do it efficiently. The presumption of lean thinking is that if you do these things, you save costs, produce faster, and are more flexible to respond to changes in market demand.

Space does not allow more detailed discussion of lean thinking or lean manufacturing here. For a better appreciation of what "lean" really is, readers are encouraged to examine in more detail four related books (listed in the bibliography): *The Machine That Changed the World* **(29)** and *Lean Thinking* **(28)**, by Womack and Jones; *Lean Manufacturing: Tools, Techniques, and How To Use Them* **(9)** by Feld, and *Just Another Car Factory?* **(22)** by Rinehart, Huxley, and Robertson.

## THEORY OF CONSTRAINTS

Like lean production, the Theory of Constraints (TOC) has evolved into a whole-system

philosophy. It, too, has a manufacturing component, but the methodology's focus remains at the system level. How do we define "system?" This could be the whole business, or it could be a strategic business unit (an independent division, for example). It could also be one particular plant in a division, though in this case some consideration of external dependencies is usually required.

Constraint theory is comprised of principles/concepts and tools. Much like the five principles of lean production, TOC principles/concepts provide the overall "navigational direction" to make sure company efforts stay pointed in the right direction. TOC tools evolved to satisfy four specific needs most companies must deal with regularly: *problem-solving, production* (whether manufacturing or service), *project management,* and *metrics.*

## Systems as Chains

Let's consider TOC concepts and principles first. Goldratt characterized systems as chains. **(13: p.53)** These, of course, aren't chains in the literal sense. They're chains of interdependency. And they don't necessarily have to be a single sequence of links, either. Figure 3 shows two possible chain configurations.

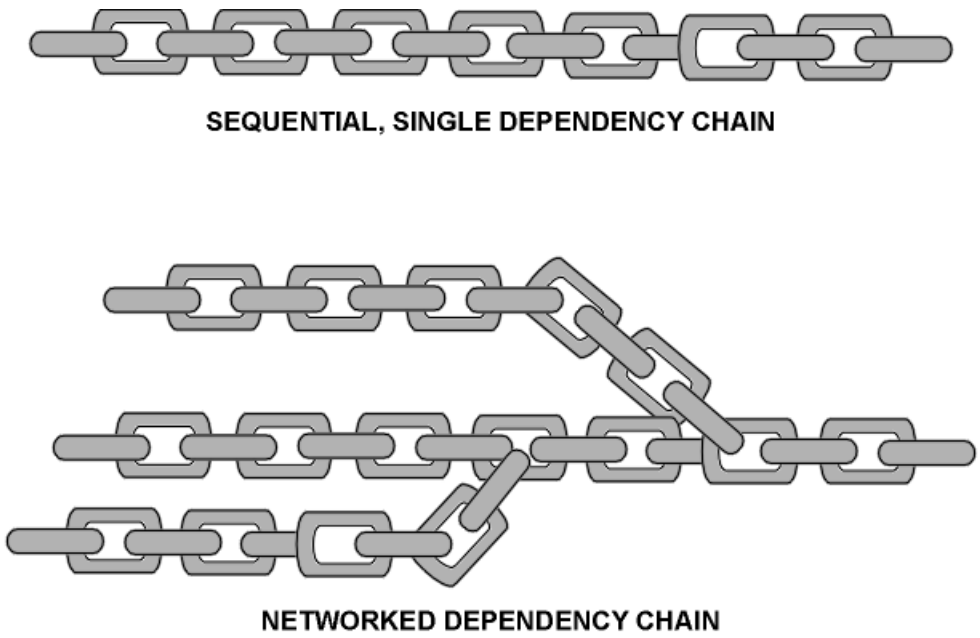



**Figure 3. Systems as Chains**

Much like a process map in lean manufacturing, the chains characterize the flow of work through the business system. Each link in the chain has a specific maximum capacity, and these capacities usually differ from one another. For example, the capacity of a punch-press and its operator to make holes in a piece of sheet metal is different from the capacity of a cutting machine to shape that sheet metal into a defined form. The cutting machine might require 30 seconds to complete one piece, whereas the punch-press might be able to process 100 pieces in the same time. Similarly, a salesperson might sell 500 units of a product in a single day, though an assembly work center might only be able to complete 200 in the same time.

Goldratt used the chain analogy to emphasize the concept of the weakest link. The strength of the entire chain is limited by the maximum strain the weakest link can stand. In the same way, the performance of a sequential flow system is limited by the least capable element of that system. The weakest link can occur anywhere in the chain of dependency. For example, limited capacity on a piece of manufacturing equipment might restrict output. Or sales might not be enough to fill up available capacity. In the first situation, the performance of the whole company (system) is limited by a physical resource. In the second, it's limited by insufficient external demand. The key word is "limited." It's the essence of the definition of *constraint:* anything that limits the performance of a system in achieving its goal.

A business system's goal, especially if it's a publicly held corporation, is usually to make money. If this isn't the company's ultimate goal, it's certainly a critical success factor. In striving to make money, the business's constraint can be a physical resource (equipment, people), market demand, material (availability, quality), a vendor or supplier, finances (cash flow), knowledge or competence, or a policy of some kind. Quality might also be considered a system constraint, though its effects usually manifest themselves in one of the other factors above.

In any chain, there are two kinds of links: the weakest one, and all the rest. Reinforcing others besides the weakest link might make us feel better, but the chain wouldn't become any stronger. If we're able to strengthen the weakest link, the whole chain would be able to accept a greater load. Similarly, in complex interdependent systems, any improvements to non-constraints don't produce any direct improvement in the continuing performance of the system. Only improvements to the system constraint will produce immediate, persistent effects.

Here's an example. Let's say that a manufacturing company has four major steps in its production process, a sales effort at the front end, and a distribution function at the back end. Let's also assume that the four manufacturing process steps are less than fully efficient. Maybe they require more time than necessary to complete work and tie up more raw material, work-in-process, and finished inventories than might really be needed. But let's also assume that this company's capacity to produce is less than fully loaded—there's slack time at every step of the process. The company's constraint is clearly insufficient external sales.

If we make the internal process steps more efficient, all we've done is create more unused capacity. If we reduce all three inventories, we haven't made any more money; all we've done is liquidate assets we've already spent money on and reduced the requirement to spend more—a one-time saving that might be substantial, but it won't count in our ledger in the next reporting period. Unless we lay off employees, we haven't saved any labor costs, because they're paid by the hour or month, whether they have anything to do or not. On the other hand, if our sales department generates more orders—enough to fill up our manufacturing capacity—we immediately make more money, whether we do anything to improve internal efficiency or not. Yes, we could certainly make more if we improve efficiency, and we'd be fools to neglect that. But if we make the system constraint our immediate priority, we have an immediate, continuing impact on the profit and cash flow of the company.

An important point to remember: constraints never completely disappear; when one constraint is broken, something else becomes the system constraint. If the sales department in the example above breaks the market demand constraint by generating a significant increase in sales, it's very likely that some aspect of the company's internal production capacity would emerge as the new system constraint. At this point, certain efforts to improve efficiency would produce immediate improvement in business performance (cash flow, profit). At some point, the internal capacity constraint might be alleviated, pushing the constraint back out to market demand again. But each time the constraint moves in response to specifically directed actions, the company's overall performance makes a measurable jump upward.

## Four Assumptions of Constraint Theory

The Theory of Constraints is based on four assumptions that apply to all systems, whether they're manufacturing or service, for-profit or not-for-profit. These assumptions lie behind all the other TOC principles and prescriptions:

1. Every system (organization) has a goal and necessary conditions that must be satisfied in order to achieve it. **(23: p.24)**
2. The system optimum is not the sum of the local optima (efficiencies). **(13: p.51)**
3. Very few variables—maybe only one—limit the performance of a system at any given time. **(23: p.26)**
4. All systems are subject to cause-and-effect.

## Five Focusing Steps

These assumptions are the logical foundation of Goldratt's prescription for system improvement, the Five Focusing Steps of TOC, which are explained in more detail in *The Haystack Syndrome* **(13: pp. 58-63)**:

1. IDENTIFY the system's constraint.
2. Decide how to EXPLOIT the system's constraint.
3. SUBORDINATE everything else to the decision in step 2.
4. ELEVATE the system's constraint.
5. Go back to step 1, but don't allow "inertia" to cause a system constraint.

To manage a system using the constraint philosophy, Goldratt created four functional tools. He conceived *drum-buffer-rope,* a finite-capacity production management methodology. **(30: pp. 103-262; 14; 23: pp. 73-135)** For project management environments, he created a scheduling tool called *critical chain.* **(11; 15; 19)**

To facilitate the measurement (metrics) and information tasks, Goldratt conceptualized a simplified financial structure **(13: pp. 14-51; 3; 20; 6; 25)** to assist in management decision-making. The structure is focused on the idea that increasing Throughput (a financial value rather than a physical output) should be the driving motivator for management action. Other writers have referred to this concept as Throughput accounting, but their definition usually falls short of considering system constraints, stopping with direct costing alone. The financial elements of this structure are used *before* a decision is made to estimate its anticipated effects, and *after* the decision is implemented to assess its success.

Each of these tools is explained in detail in the references cited. Drum-buffer-rope will be discussed in a bit more detail below, as we compare lean manufacturing with TOC.

## COMPARING LEAN AND TOC

As indicated earlier, there is a high degree of congruence between the lean production approach and the Theory of Constraints. Figure 4 shows many of their similarities.

- A whole system methodology
- Ongoing (continuous) improvement essential
- Objective: Higher profits
- Value is defined by the customer
- The value stream (supply chain) extends beyond the manufacturing plant
- Quality is essential to success
- Small production batches
- Continuous flow (rather than queue)
- Pull (make-to-order, rather than make-to-stock)
- Liberate hidden capacity
- Minimize inventory
- Work force participation is key to success

**Figure 4. Lean Manufacturing and TOC - Similarities**

### Similarities

Both lean thinking and TOC are whole-system methodologies. Lean production—the Toyota Production System, as described by Womack, Jones and Roos **(29)**—is more so than lean manufacturing alone; TOC is closer to the level of lean production than it is to lean manufacturing. Both lean thinking and TOC emphasize continuous improvement, and the objective of both is higher profits. Both methods recognize that the customer is the final arbiter of what *value* is. Lean thinking refers to the whole system of production as a *value stream.* TOC treats it as a supply chain that includes the customer. The meanings of the two are essentially the same. Quality is essential to the success of both methodologies. So is the small production batch. Both promote continuous flow, rather than large batch-and-queue. And the ideal objective of both lean and TOC is a pull system, in which the need for finished inventory is minimized because the production process is fast enough to make-to-order. Both lean and TOC liberate hidden capacity and minimize all types of inventory, especially work-in-process and finished stock. And the success of either methodology depends on the cooperation and participation of the work force, from senior management on down to the line.

## Differences

There are major and minor differences between lean production and TOC, as well. (Figure 5) The major differences are important and require some choices. The minor differences are relatively easy to reconcile. The two major differences lie in a) how each treats variability and uncertainty, and b) how each treats costs. Let's discuss the major differences first.

| LEAN | TOC |
|---|---|
| 1. Cost reduction (both fixed and variable) is the best way to profitability | 1. Costs have a point of diminishing returns; Throughput ($$$) does not |
| 2. No end to reducing effort, time, space, cost and mistakes (Perfection) | 2. Cost reduction is secondary to Throughput generation (generally, increasing sales) |
| 3. All instances of waste reduction are celebrated | 3. Only waste reduction at the constraint has an immediate impact |
| 4. Resources are typicallyorganized around specific products | 4. Resources are shared across product lines or value streams |
| 5. Does not differentiate between constraints and non-constraints;all changes (+ or -) are equally important | 5. Time lost at a constraint represents Throughput ($$) lost to the system; time saved at a non-constraint has no immediate value |
| 6. Inventory buffers are physical things | 6. Buffers are TIME, not physical things |
| 7. Emphasize single-piece flow | 7. Reduce flow quantity as much as possible without jeopardizing flow through the constraint |
| 8. No differentiation between process batch and transfer batch sizes | 8. Process batch size is different from transfer batch size |
| 9. Seeks to eliminate all variability; doesn't attempt to deal with external (market) uncertainty | 9. Accepts variation ("Murphy") and external (market) uncertainty as a way of life, and protects against both to the extent possible |
| 10. No "safety net"- everything works or nothing works | 10. Nothing ever works perfectly all the time, so plan for it |

**Figure 5. Lean Manufacturing and TOC - Differences**

## Major Difference: Cost

Lean thinking puts cost reduction—both fixed and variable— at the center of all improvement efforts.[‡] *It recognizes no end to the reduction of effort, time, space, cost and mistakes.* **(28: p. 25)** TOC suggests that there is a point of diminishing returns to cost reduction. Figure 6 illustrates this point.

[‡] "To achieve this purpose [profit], the primary tool of the Toyota Production System is cost reduction, or the improvement of productivity... attained through elimination of various wastes such as excessive inventory and excessive work force." (Monden, Yasuhiro. *Toyota Production System* (3rd ed), 1998)

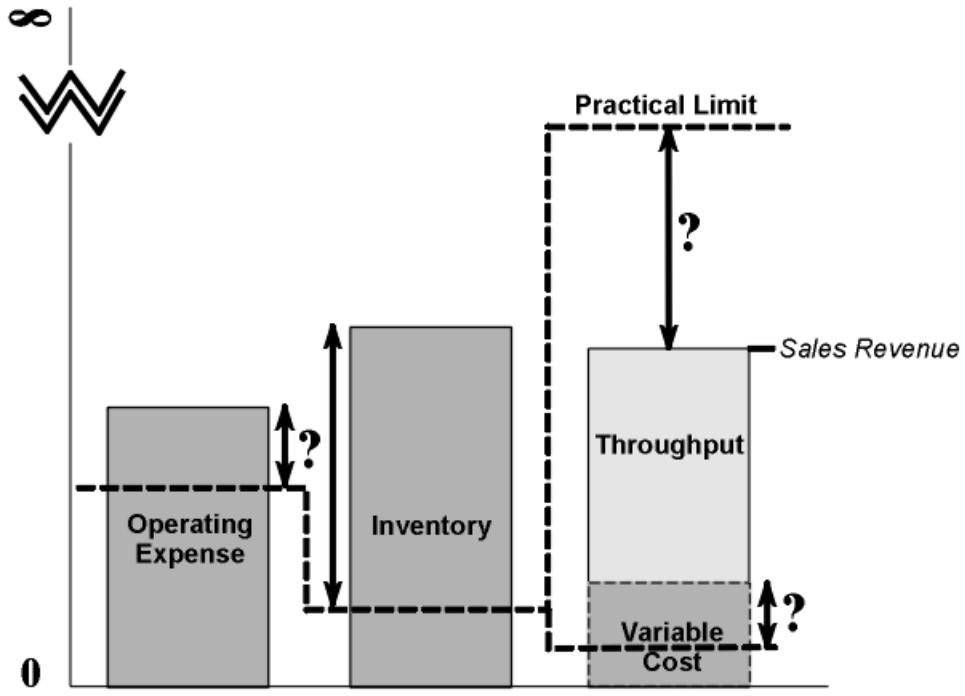



**Figure 6. Cost Reduction: Diminishing Returns**

Every commercial business incurs fixed costs to produce and sell its products or services. TOC considers these to be Operating Expenses: the sum of the costs of opening the doors for business each day, whether or not a unit of product is sold. A certain level of Operating Expense is necessary to run any business at a specific production rate. The same is true of *Inventory* (which, in the TOC way of thinking, includes capital assets such as facilities and equipment, as well). While we could theoretically reduce Operating Expense and Inventory to zero, in reality we must maintain some level of each if the business is to function at all. So there is some practical limit—well above zero—below which Operating Expense and Inventory can't be reduced without hurting the company's ability to produce value for the customer.

It's certainly possible for typical traditional mass production companies to reduce both Operating Expense and Inventory, and where such opportunities present themselves, companies should capitalize on them. But these are essentially one-time savings. They don't go on forever. Eventually, all the "fat" is gone, and only "muscle" remains. Cutting further hurts the company's ability to do its job. And at any given time, it's extremely difficult to determine what expenses are above that "practical limit" line and which ones are below it, so the knife must be wielded with care, especially as one gets closer to the practical limit.

But there's a third bar on the graph: *Throughput.* Throughput, as defined by TOC, is not the same as it is in traditional mass production or lean thinking, both of which equate it with output. In TOC, Throughput is measured in financial terms, not in units of products or materials. This distinction is necessary for effective decision-making, because not all product units are of equal value. Sometimes identical products can produce different Throughput in different circumstances (e.g., quantity discounts, different geogrpahical markets). Throughput is the

money you have left from sales revenue after you've paid all the truly variable expenses of producing products or services—for example, raw materials, commissions, variable transportation costs. While the practical limit for increasing Throughput is not infinite, it provides considerably more room for improvement than Operating Expense, Inventory, and variable cost reductions, which have discrete lower limits.

## Cost Reduction or Throughput Increase? An Example

Here's an example. Let's assume your company generates $50 million a year in sales revenues, and your capacity is about 70 percent loaded. Of that, $20 million must be paid for raw materials and other variable expenses. Let's assume that your fixed Operating Expense is another $20 million a year, and you maintain $7 million worth of inventory on hand (raw material, work–in-process, and finished). Now let's assume that you can both reduce costs and increase Throughput (the financial value). Which offers the largest potential for improving the company's financial position? Figure 7 depicts the baseline situation described above and the results of the two different options.

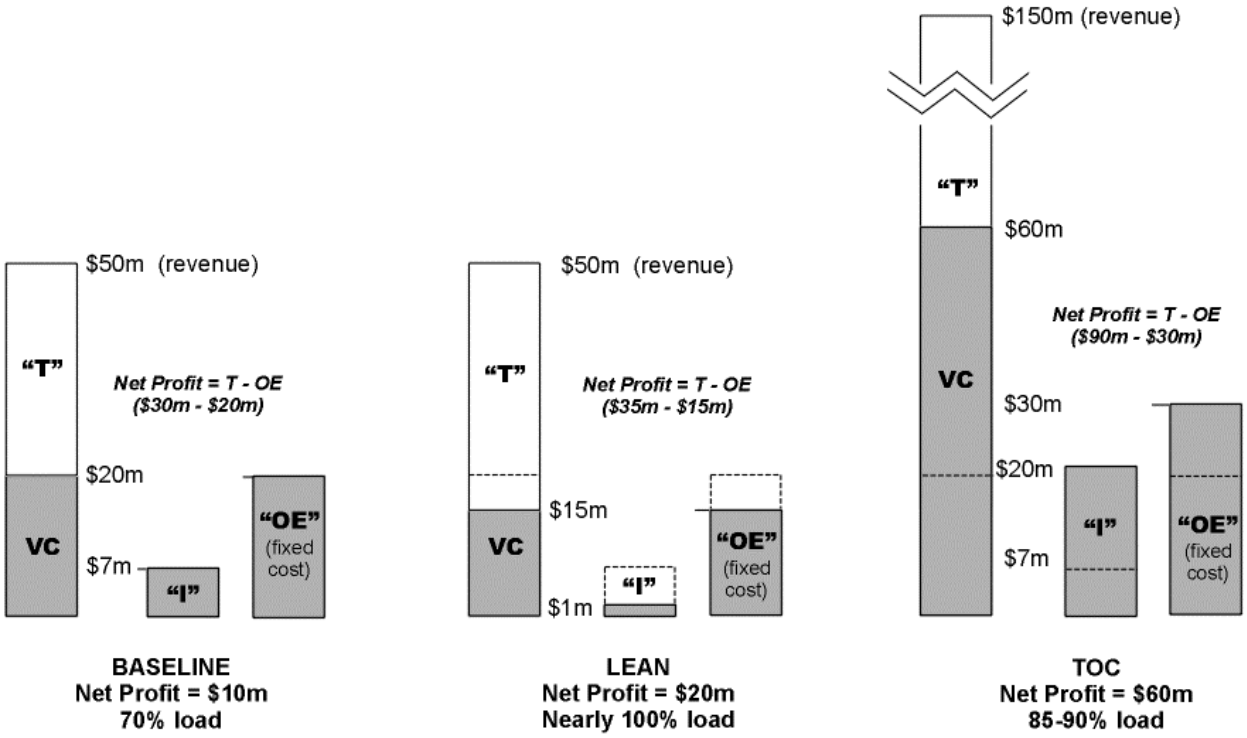



**Figure 7. Cost Reduction or Throughput Increase? An Example**

In the first year, you might be able to shave variable expense $5 million by eliminating scrap. If you apply lean thinking conscientiously, you'll more closely match your work force to the load on your resources so as to be able to achieve higher efficiency. Resource loading will approach 100 percent. So let's say you can eliminate $5 million in "waste" from your fixed expenses. Let's also assume you can run a really lean inventory without hurting yourself—you can drop that to $1 million. That's a total of $10 million in costs saved—it goes straight to the bottom line—plus another $6 million you don't have tied up in non-liquid assets.

Congratulations! You've saved money and increased your efficiency. You're now "lean."

What will you do for an encore the second year? If you did a good job "leaning up" the first year, you'll be fortunate if you can shave those numbers by another 5 percent the second year. The third year will drop even less. And you won't be sure when you go below the "fat" level and into the "muscle" of your business.

Now let's look at the potential for increasing Throughput. Let's say your $50 million in revenues each year comes from 100 customers. If you could reach out and touch them, how many potential customers like that are there in the world? 500? 1,000? 5,000? 10,000? More? Let's be conservative. We'll say that you can find 200 new customers for the same products (three times your current volume). That makes your total revenues for the year $150 million. Without doing anything "lean," your variable costs increase from $20 million to $60 million. You can assume your fixed expenses will go up some, too, because you have to hire 50 percent more employees. You don't have to buy any more equipment, because you can add shifts instead, but you might add another $10 million to your Operating Expense. And you now have to maintain $21 million in inventory, instead of $7 million.

Though the increase in total costs is $50 million, the increase in revenue is $100 million. So the potential increase in profit from emphasizing Throughput increase is $60 million, without even "leaning up." ‡ Even if these estimates are off by a factor of 30 percent, you're still $10 million ahead of the one-year saving realized by cost reductions alone.

From a practical standpoint, such a sales volume increase will probably require some aspects of quality and reliability improvement inherent in lean manufacturing, just to give you the competitive edge necessary to win your 200 new customers. But as Womack and Jones have noted, if you can deliver superior reliability, as Toyota did, you don't necessarily have to match the price of competing mass-produced products. **(29: p.64)**

The conclusion I'd like you to draw is that *it's not necessary to make a choice between lean thinking and TOC.* From the preceding example, it should be clear that the six-fold increase in profit (from $10 million to $60 million) realized by emphasizing Throughput could probably have been even higher, by perhaps $10 million, by capitalizing on opportunities to apply lean thinking as well. There are ways to productively combine the two, which we'll see later. However, the hybrid will require avoiding "extremes" in both.

## The "Red" Curve or the "Green" Curve?

The cost reduction-versus-Throughput expansion issue can be summarized in Figure 8. Some years ago, Goldratt used the terms "green curve" and "red curve" to differentiate between

---

‡ We're clearly making some assumptions here. For instance, we're assuming that you can add shifts, rather than expanding facilities. We're assuming that doing so will deliver the quantum increases in capacity you'll need to triple your output. We're assuming that your quality, price and delivery reliability are at least equivalent to your competitors in the market place.

business performance that tapered off to diminishing returns and performance that made regular upward jumps. **(24)** (Since this paper isn't published in color, the "green curve" is represented by the broken line in Figure 8 and the "red curve" by a solid line.)

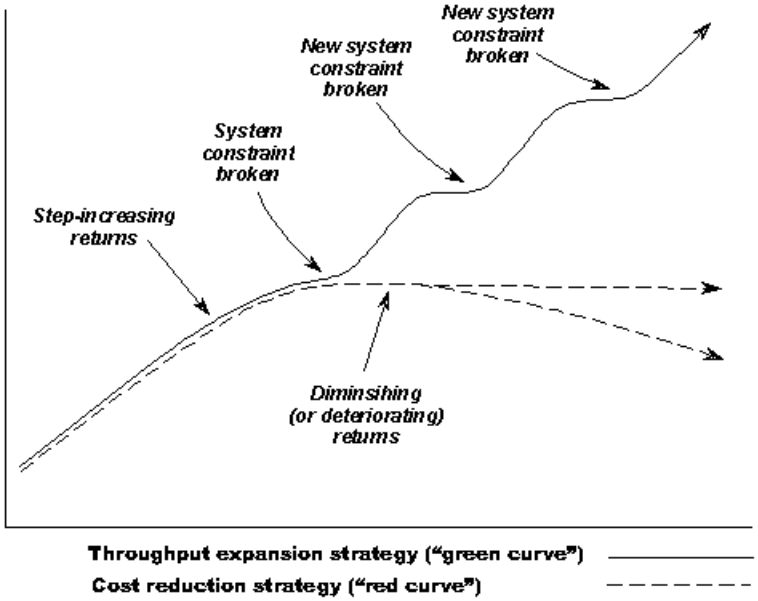



**Figure 8. Cost Reduction Versus Throughput Expansion**
**The "Red Curve" and the "Green Curve"**

Because costs can't be reduced indefinitely, without substantial attention to increasing Throughput, normally through expansion of sales, lean manufacturing's financial performance will tend to approximate the broken line in Figure 8. In other words, when all the cost savings that are practical to realize have been obtained, profitability will level off. Moreover, as environmental conditions change and solutions obsolesce, performance may actually deteriorate to some degree. However, as each successive new system constraint is broken, business performance makes a measurable upward jump, followed by the same tendency to level off into diminishing returns (albeit at a higher level). An ongoing process of finding an breaking constraints provides many more opportunities to raise financial performance over time than a single-minded devotion to reducing costs.

## Major Difference: Variability and Uncertainty

Another key difference between lean thinking and TOC lies in their respective treatments of variability and external uncertainty. We'll distinguish between the two this way. *Variability* is pretty much internal to the organization. It encompasses all the technical product and process factors that statistical process control is designed to address. It might also include qualitative or human factors as well. Any performance of any kind (in any functional area) that occurs within the direct authority of management—meaning, usually, internal—is subject to variability. Sometimes that variation can be estimated or predicted. Other times it can't. But it happens, and the organization has to find a way to deal with it if the system as a whole is to be controlled effectively.

*Uncertainty* is external. It includes factors outside the control of the organization, or possibly only marginally influenced. Customer behavior is one such factor. Supplier behavior might be another. Changes in market taste or demand, economic "peaks and valleys," or natural disasters might be others. Both uncertainty and variability affect the operations of a company. Generally, uncertainty causes changes in demand for the company's products or services, and variability causes changes to the company's capacity to satisfy that demand. The financial outcome for changes in each can be good, bad, or neutral. For the company to prosper, however, it's essential that it be configured to make best advantage of whatever condition (or combination of conditions) presents itself. Lean thinking and TOC diverge in their approaches to this challenge.

Let's consider variability first. Lean thinking pursues the elimination of internal variability with a vengeance. TOC does not. Rather, TOC presumes that your processes are already in statistical control and producing high quality products. In this respect there is a complementary relationship between the two methodologies.

But TOC doesn't pursue *perfection* the way that lean thinking does. The same law of diminishing returns applies to process improvement as to cost reduction. In fact, Juran emphasized this point with his *cost of quality* graph (Figure 9). If your process produces 98 percent defective products, improvement is probably both simple and profitable. But the closer a system is to producing perfectly, the lower the returns (and higher the cost) will be in improving further.

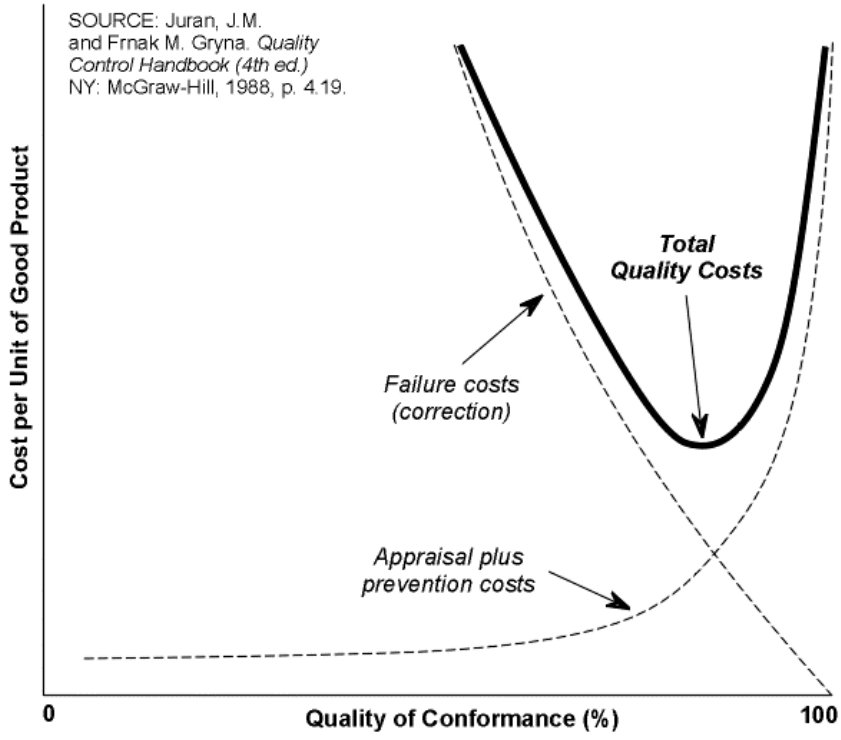



**Figure 9. Cost of Quality (from Juran)**

## "Murphy" and Buffers

Moreover, TOC recognizes the existence of "Murphy." [‡] Specifically, TOC allows for the possibility that even if all internal processes are under the strictest statistical control, unforseen events can intervene to interrupt business. TOC protects against "Murphy" with buffers. However, in contrast with traditional thinking, buffers are not *things* (i.e., inventory). In TOC, buffers are *time.*

Here's an example. Company "A," a typical manufacturing operation, would ensure that deliveries to customers were on time—even for short-notice, fast-turnaround demands—by maintaining a finished goods inventory of 1,000 units. There are no firm orders for these units—they're just "insurance" the company pays in exchange for responsiveness. When the finished inventory level dropped below this point, production might be accelerated to refill it back up to that level.

Company "B," a TOC-oriented operation, would say, "Let's schedule our firm orders to arrive at the shipping dock some *period of time* before they're due to be shipped." The length of time may vary by product or season, but it's generally just long enough to permit recovery from the worst kind of "Murphy" that could reasonably be expected.

This accounts for internal variability. What about external uncertainty? What about capricious (i.e., unscheduled, unpredictable, urgent) customer demands? For example, let's say a customer needs delivery two days after an unexpected order is received, but our company requires five days to produce the same order from scratch. TOC suggests that there are ways to accommodate such emergencies, at least in some cases. One way is to ensure that the capacity of all internal resources, even the system's least capable resource, is never fully loaded by design. This could require planning for no more than an 85 or 90 percent load at the capacity-constrained resource. In other words, leave a little protective capacity for emergencies. Another way might be for the company to maintain a small amount of physical finished inventory—but no more than necessary to cover the time difference between the company's shortest manufacturing cycle time (for an emergency or expedited order) and the customer- demanded time. In the preceding example, the company might maintain a three-day physical inventory to cover the difference between what it could actually do and what the customer expected.

The difference between the two philosophies is crucial for two reasons. First, reality is such that "working without a safety net" is never likely to be practical—and it can be fatal. And second, the level of effort (and expenditures) needed to eliminate variability and uncertainty completely rises so exponentially that begins to look like Juran's curve (Figure 9): the cost of prevention exceeds the benefits derived.

Of course, there's a middle ground here. Efforts to reduce the *muda* associated with holding inventories and improving processes can certainly contribute to a better bottom line—in TOC parlance, less Inventory and Operating Expense, and more Throughput (through lower

---

[‡] A reference to "Murphy's Law": *Whatever can go wrong will go wrong, and it will happen at the most inconvenient time.*

variable costs). And efforts to decrease internal variability undoubtedly contribute to improved Throughput (by reducing scrap and rework) and smaller buffers (less protection needed, because the variability range is narrower).

But such improvements have a point of diminishing returns. And without knowing for sure at what point the Juran curve breaks the other way, TOC maintains that it's considerably more prudent to prioritize efforts to increase Throughput first, and efforts to reduce Inventory and Operating Expense second and third (in that order). One TOC researcher, John Caspari, even suggests incentivizing Throughput protection and improvement alone, while allowing Inventory and Operating Expense to seek their own natural levels by not rewarding cost cutting in those areas. **(4)**

To summarize, lean thinking says that the need such protection should be eliminated[‡] through the perfection of processes to the point of total reliability. This is a fundamental difference from TOC, which suggests that we can't ever completely eliminate variability. And companies have even less control over uncertainty.

Some differences between lean production and TOC are more than just issues that one addresses but the other doesn't. There are some genuine philosophical differences on the usefulness of some tools or the advisability of some prescriptions. While a complete examination of all the differences is beyond the scope of this paper, two are worth mentioning. One such difference is the concept of *cycle time* and *takt* time. Another is *work balancing and one-piece flow.*

## *Takt* and Cycle Time

At the work center level, cycle time is the length of time required for a work center to complete one repetition of work on a single unit of product. For example, a fast, experienced cutting tool operator might be able to complete the cutting of a single product in one minute. *Takt* time, on the other hand, is demand-based. *Takt* comes from the German word for rhythm, or beat. **(9: pp. 69-70)** It's a ratio of the total production time available per day to the designed daily production rate. Dividing time available to produce by some kind of demand requirement input for the same period of time (firm orders or forecasts) yield the *takt* time.

For example, let's say that in one work cell an eight-hour shift, minus breaks and lunch, allows seven hours (420 minutes) of production time available. If we have all of today to produce 10 items, the *takt* time is 42 minutes per item. The scheduler uses the *takt* time to determine how many people to assign to the work cell to have the "best efficiency." To do this, the scheduler looks at the cycle time for the work cell (probably more than one manchine and operator). Let's say the cycle time is 4.2 minutes, which is one-tenth of the *takt* time. This means that the scheduler can reduce staffing in that cell today by nine-tenths, with the result that the work cell will approach 100 percent efficiency (i.e., one guy working his butt off for eight hours).

---

[‡] Womack and Jones use the term "no safety net." *The Machine That Changed the World,* p. 103.

Now, if that work cell has only three people assigned to it, the scheduler can reduce staffing in that cell to just one person for today, and deploy the other two somewhere else (where, ideally, they're qualified to do what's needed). In fact, the scheduler can even redeploy that third person for part of the day as well, after he or she finishes being "100 percent efficient" on the primary job. So *takt* time varies with the demand cycle, however long that might be, and it's not tied to actual cycle time, though the two are used jointly to calculate staffing needs. In today's tight job market, it's not likely that idle staff would be sent home without pay. If that happened, they'd look for more dependable work elsewhere. So reallocating staff that isn't needed in a primary job at the moment won't actually cause Operating Expense to go down.

In any case, a lean production process will assume that each unit of capacity (a person, or person-machine combination) will work as fast as possible, while still assuring mistake-free operation, for all of the available production time (in our example, seven hours). The end result is that a smaller total work force is needed to produce the same volume of output, meaning better efficiency.

The Theory of Constraints doesn't consider *takt* time at all, only the approximate cycle time. We say "approximate," because TOC assumes some differences in skill level among employees, leading to variability within a reasonable range around some nominal cycle time value. And TOC's focus would be on the cycle time for the entire manufacturing process, not individual cycle times for each work cell. Remember the chain analogy: only one link in the chain can be the weakest, and that link sets the pace for the entire chain: *the output of the system over time is the same as the output of the constraint.* Striving for local efficiencies at other links of the chain doesn't get the product finished faster. Neither does it save money. (The reasons for the latter statement will become clearer shortly.) TOC suggests that maximum efficiency really counts only at the capacity-constrained resource (CCR)—that step in the manufacturing process taking the most time to complete one unit of product.

Why doesn't maximum efficiency throughout the process matter? Notice the use of the word "maximum." There are really only two reasons why efficiency should be important in the first place: saving time and money. Saving time is only crucial if there is competition among activities for the same time. For example, if one manufacturing process, such as assembly, is used by two different product lines, and the volume of each product for which there's a firm demand comes close to consuming all of assembly's time, then every minute saved can be important. (Remember, a resource loaded to or above its capacity is a constraint.) But let's say the step after assembly is packaging, and that takes less than a fifth of the time that assembly does. And perhaps just one packer can handle the same volume as the whole assembly operation can, with time to spare. Making that packaging step more efficient (e.g., loading the packer closer to 100 percent of his or her capacity) doesn't make the whole process of manufacturing either product any faster.

What about saving money? Surely the less idle (i.e., non-value adding) time an employee has, the more money is saved. Not necessarily. Generally speaking, these days labor is not a variable cost. As long as employees aren't paid by the piece they produce, labor costs are fixed, at least over the short term. We pay permanent employees by the hour or the week. Whether our plant operates with or without a union, there is usually a minimum guaranteed work week. So, if

employees don't have any value-adding work to do for two or three hours, we still pay them for that time. We might be able to make good use of that "idle" time in other ways, such as total productive maintenance, training, or housekeeping. But unless we trim our work force to more closely match our demand—an extremely risky proposition—we won't save any real money.

## Work Balancing and One-Piece Flow

Two of the keys to success in lean manufacturing are balancing the workload in each cell (to avoid overload) and one-piece flow. The underlying assumption behind striving for a balanced workload and one-piece flow is that efficient management of all system components is required to achieve an efficient "whole system." (The sum of the local efficiencies gives the maximum system efficiency).

To balance the workload, lean-thinking managers analyze machine time, man time, and setup time in each discrete work cell. Management compares *takt* time—a reflection of demand for products—with each of these three times individually. **(9: p.71)**

If the machine cycle time (the time a machine requires to complete one piece) is greater than the *takt* time (time demanded to produce the product), available machine time must be increased by off-loading work, splitting demand, reducing the cycle time, adding equipment, or changing the processes. The latter three are longer-term efforts that may not be responsive to short-term changes in demand. If, after taking these actions, the required machine time is still greater than *takt* time, the operation will have to be balanced with in-process *kanban* inventory and/or additional shifts (although lean thinking doesn't explain where this *kanban* inventory comes from if there isn't any extra capacity to generate it).

Man time is considered separately from machine time, but the same comparison is made (man time required versus *takt* time). The assumption here is that, depending on the level and sophistication of automation, one person might be able to operate more than one machine simultaneously—a valid assumption, as long as everything works properly all the time.

Workload balancing calls for reducing, shifting, re-sequencing, combining, or eliminating individual work elements in a cell. The objective is to balance workload with *takt* time. In other words, if *takt* time represents demand on the process and the machine/man times represent capacity to fulfill that demand, then workload balancing amounts to an effort to level capacity (balance the line) at a point that matches demand. The objective of workload balancing is to facilitate one-piece flow, which will be discussed a bit more in a moment.

Lean thinking also places great emphasis on reducing setup times in every cell. Setup time reduction serves to "buy" more resource time. Spending less time setting up means that more time is available to do value-adding work (i.e., producing products that can be sold). Shorter setup times, according to lean thinking, enable the following production strategy:

> "Plan on setting up each high-volume product every day, then schedule the product mix to run accordingly. If this can't be accomplished, plan to run 2-3 days' worth at a time and hold the excess inventory until the customer or customer cell asks for it (never allow this to extend past more than a week's run). It will become very clear, very quickly, why

setup reduction is so important, when the supplier cell has to physically hold the excess inventory until the customer asks for it through a *kanban.*" **(9: p.71)**

After the machine, man and setup times are compared with the *takt* time, efforts turn to generating ideas and looking for cell designs that will make the cell more balanced compared to the *takt.* When operations are balanced to *takt* time, it's possible to realize the advantages (speed, flexibility) of a one-piece workflow, instead of running in large batch quantities. One-piece flow, as prescribed in lean manufacturing, means that products are passed one piece at a time from one operation to the next, with a first-in/first-out (FIFO) priority.

## The TOC Approach to Workflow

The overall objectives of workload balancing and one-piece flow in lean manufacturing are speed and flexibility. Lean manufacturers want orders to flow through the entire process in minimum time, and in small enough "bites" that different products can be made in the same work cells in shorter intervals of time. This is somewhat akin to a computer running different applications simultaneously through time-sharing. And the emphasis on balancing the line capacity to do so keeps the entire production process working at very high efficiencies in all work cells, which most people would consider to be a good thing.

However, workload balancing is a complex effort, and the result—if lean thinking is truly adhered to—is that each step in the production process is nearly fully loaded and operated without buffers as safety nets. A system in this state is highly vulnerable to disruption at any point. In continuous flow systems, it's all or nothing. Hence the focus on reducing variability.

> "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... By design, flow systems have an *everything-works-or-nothing works quality which must be respected and anticipated.*" **(28: p.60)** [Emphasis added]

In other words, rid the system of both common and special cause variation, and eliminate all the external uncertainty. Only in this way can flow disruptions be prevented. Accept this requirement, and expect to have to do it.

This is a noble objective, but is it realistic? Practitioners of TOC suggest that this level of perfection may never be obtainable, or only with a tremendous investment in improvement efforts. Again, the "Juran curve" (Figure 9) pertains. TOC poses (and answers) the question, "What if we could live with some reasonably achievable level of variability and uncertainty, yet still realize the objectives of speed and flexibility?"

Moreover, TOC suggests that there's an easier way to smooth the flow of work through a production system, increase the speed, attain flexibility, and in addition improve delivery reliability, all without the intensive efforts needed to balance the capacity of each cell individually. Goldratt called this method *drum-buffer-rope.* **(14: p. 96)**

## Drum-Buffer-Rope

*Drum-buffer-rope* is predicated on several basic assumptions:

1. Continuous flow at small batch sizes is better (faster, more flexible) than large batches and queues.
2. Disruptions to flow come from variability (internal) and uncertainty (external).
3. It's not practical to eliminate all disruptions to flow in a system.
4. Different process steps, work centers, or work cells inherently have different capacities (rates of flow).‡
5. Overt attempts to balance capacity bring with them a number of difficult decisions when the system is out of equilibrium with the market demand.‡‡
6. The output of the entire system can never exceed the capacity of the CCR (the weakest link in the internal chain of dependent events).
7. "Increasing efficiency" anywhere but the CCR does nothing to improve the volume or speed of the whole system's output; the same efforts at the CCR produce *immediate* benefits to the system's output.
8. The net processing time of one unit of a product is very small compared to the actual production lead time.

With these assumptions in mind, *drum-buffer-rope* postulates to the following principles:

- Balancing capacity is an exercise in futility. It's expensive and difficult, if not impossible, to adjust capacity (except within very narrow ranges) as quickly as uncertainty changes external demand on the system or "Murphy" occurs. Internal variation, especially in a balanced line loaded to or near full capacity (efficiency), drives management into a "firefighting" mode, chasing fixes to every disruption to flow, without ever knowing where the next one will occur. **(23: pp. 28-29)**

- Not all excess capacity in the system (e.g., at work resources other than the CCR) is really waste (*muda*). Some extra (protective) capacity, or time, is required everywhere—even at the CCR—to offset disruptions to flow that can't be avoided. **(7: p. 220)**

- It's neither possible nor desirable to attain high levels of local efficiency everywhere in the manufacturing chain all the time. **(23: pp. 26-31)**

So, with these principles in mind, drum-buffer-rope offers a way to realize the same objectives—speed and flexibility—as lean manufacturing without the complexity and constant capacity adjustments. Moreover, it offers an additional benefit, *reliability*, through its tolerance for and accommodation of variability and uncertainty.‡ In other words, it doesn't have to be

---

‡ Without overt attempts to balance capacity, one step/center/cell will naturally have less capacity than the rest—a capacity-constrained resource (CCR); the others will have some slack capacity relative to the CCR.

‡‡ Increasing capacity is one: to increase capacity in response to market fluctuations (and today's markets *do* fluctuate) requires improvement at every link.

‡ Mabin and Balderstone (*The World of the Theory of Constraints: A Review of the International Literature*, 2000) report average lead time reductions of 70% while due-date performances simultaneously averaged 44% improvements. Many companies achieved 100% on-time performance at the shorter lead times.

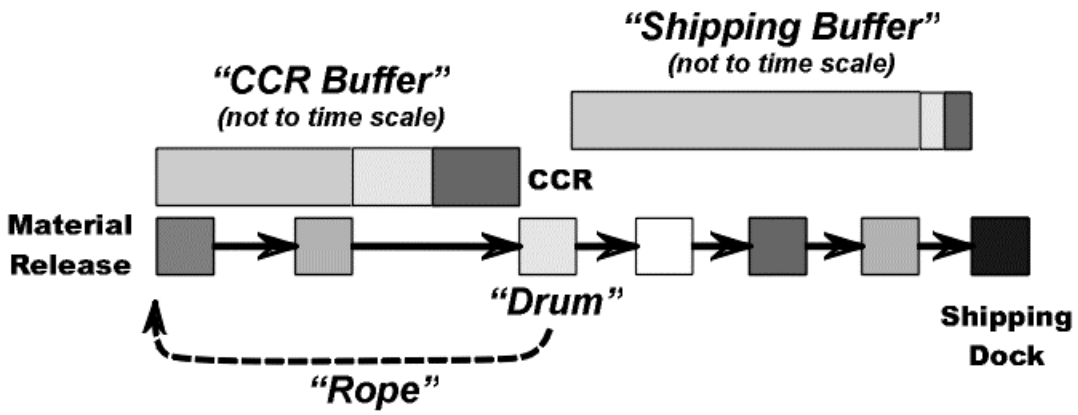
"everything-works-or-nothing-works" for speed, flexibility and reliability to be attained.

How does *drum-buffer-rope* do this? There is no shortage of books that describe how it works in detail.[‡‡] In essence, though, it works this way.

The *drum* determines the pace of the system, and, like the *takt* in lean production, it's usually driven by external demand. Sometimes, when external demand is more than the system can handle, the *drum* is the schedule for the most restricted (capacity-constrained) resource.

The *rope* is a simple communication device. It's a schedule for the release of materials into the manufacturing process. The rope's function is to prevent the production line process steps in front of the capacity-constrained resource from becoming flooded with work that the CCR can't process. While some aspects of the material release schedule—specifically, *what* is to be released—are determined by external demand, *when* they are released is controlled (and sometimes adjusted) daily by the "drumbeat," or the pace of the capacity-constrained resource. But until external demand exceeds the capability of the capacity-constrained resource, material release is immediate—the rope doesn't function as a limiter to material release.

The *buffer* is a time period protecting the two most important parts of the production system: the shipping schedule and the capacity-constrained resource. The buffer constitutes the time that a particular order should arrive at either the CCR or the shipping point *before* it's actually needed. Buffers are usually measured in hours, and they're designed to protect the manufacturing line against internal variability ("Murphy"). Figure 10 illustrates a simple *drum-buffer-rope* configuration.




Adapted from Schragenheim & Dettmer, *Manufacturing at Warp Speed*.

**Figure 10. Basic DBR Concept**

*Drum-buffer-rope* also takes a slightly different view of *one-piece flow* than does lean manufacturing. TOC/DBR differentiates between process batch size and transfer batch size.

[‡‡] Goldratt, *The Race* (1987) and The Haystack Syndrome (1990); Cox and Spencer, *The Constraint Management Handbook* (1998); Schragenheim and Dettmer, *Manufacturing at Warp Speed* (2000).

Like lean manufacturing, DBR recommends transferring units of product from one process step to another in very small quantities (a single piece, if possible, but very small numbers if that's not really practical). But DBR recognizes that working on larger quantities of the same kind of part, without changing setup, is usually critical only at the capacity-constrained resource. Since other resources naturally have more capacity, they can tolerate more setup changes—obviously, not an infinite number, but certainly more than the CCR can.

Figure 11 shows the load profile of two typical resources (it doesn't matter whether you call them work centers or cells). One is a CCR, the other is not. Notice that the non-CCR has much more "non-productive" time available that can be devoted to doing more setups, periodic maintenance, training, or whatever management deems appropriate. There is also protective capacity available at these resources to make up for lost time or delays, either at this resource or elsewhere in the system. Now notice the difference in the load profile on the CCR. There is almost no unused capacity at all. Nearly all its time is taken up doing value-adding work. If external demand surges, even such things as periodic maintenance and training would be deferred in the short term to allow more "up" time. And the number of setup changes would be held to the absolute minimum needed to meet schedules. To do otherwise would be to compromise the Throughput (financial value) generated by the whole system. However, if external demand isn't fully loading the CCR (the weakest link in the internal chain of events), there is time available to do more "discretionary" setups even at the CCR—in other words, smaller process batches. But once the CCR—or any other resource—completes work on a single unit of product, that unit is eligible to be moved on to the next workstation, just as it is in lean manufacturing under a *kanban* system.

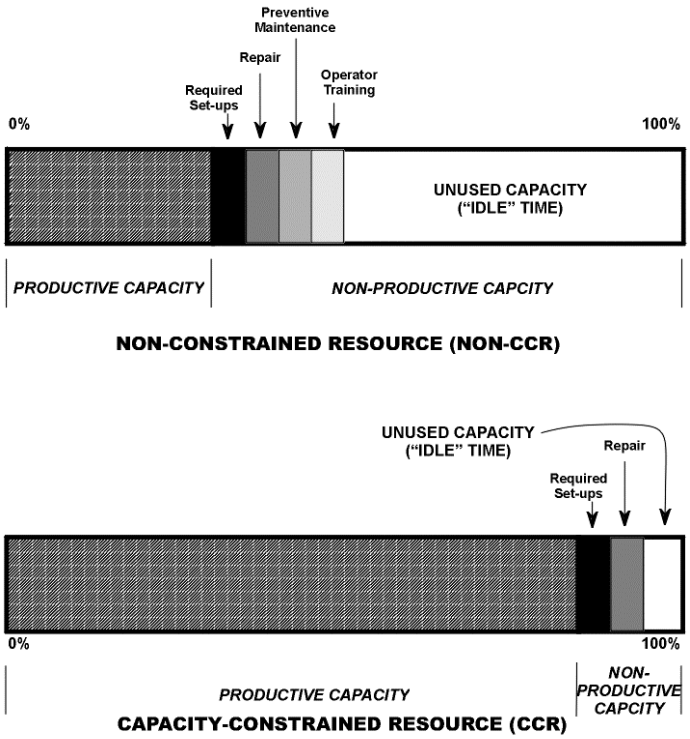



**Figure 11. Resource Load Profile**

To summarize the differences between TOC and lean thinking where work balancing and one-piece flow are concerned, it's safe to say that both aspire to the same objectives: speed and flexibility. Lean requires every link the in chain to run at the same pace, and it adjusts staffing of the links frequently to keep individual efficiencies as high as possible. In other words, it tries to keep the line as close to perfectly balanced as possible. Any variability at all will bring this system to its knees.

TOC, on the other hand, encourages each link to run at a pace independent of the others, using what's come to be known as the "road runner" approach: full speed or full stop. The net effect of this is that "idle" time is easier to aggregate for other purposes (periodic maintenance, housekeeping, training, etc.) As might be expected, TOC doesn't put a high premium on maximum efficiency at most links in the chain—only at the capacity-constrained resource. The reason that TOC tolerates local "inefficiencies" (the resultant "slack" at non-constraints) is that it considers them to be inconsequential to the overall financial performance of the system. The protective capacity this slack represents provides a way to accommodate variability for less-than-perfect environments.

TOC also accounts for the disruptions that are beyond management control, either those resulting from internal variation or external uncertainty. Lean manufacturing is likely to be more difficult to configure and manage. Its lack of "safety nets" make it intolerant of disruptions ("everything-works-or-nothing-works"), and its obsession with high efficiency everywhere craves continual adjustments everywhere as well. Consequently, for lean to succeed, you have to watch everything every day. For TOC to succeed, only a few key points in the system (the CCR and the buffers) are important to watch continually. Everything else can be checked occasionally.

## Measures of Success in Lean and TOC Environments

As one might expect, when two management philosophies emphasize different objectives—cost reduction and high efficiency in lean manufacturing, increasing Throughput in TOC—each one embraces different types of "success metrics." While it won't be possible within the scope of this paper to examine the measures of both lean manufacturing and TOC in detail, some of each are worth mentioning. The references cited provide more detailed explanation of each measure of success.

Indicators specific to lean manufacturing indicators include: **(9: 15)**

| | | |
|---|---|---|
| Lead time | Productivity | Process DPPM |
| Work-in-process | Delivery | Linearity |
| Travel | Set-up time | Increase workload |
| Density | Process yield | Increase work volume |
| Down time | Up time | Pilot new project |

Some of these are system-level measures. Other are process measures. Still others are work cell (or work center) measures. Virtually all of them encourage cost-reduction efforts.

Measurements unique to the Theory of Constraints include: **(13, 23, 25)**

Throughput ($$$)
Inventory ($$$)
Operating Expense ($$$)
Throughput/Unit of the Constraint's Time (T/CU)
ΔT−ΔOE ($$$)
Buffer level (time)
Manufacturing cycle time
Quoted lead time
Throughput Dollar Days
Inventory Dollar Days

All of these are system-level metrics. Other typical measurements such as set-up time and machine cycle time are used to calculate production capacity for planning purposes, just as they are in traditional mass production and lean production systems. However, such measurements are not routinely used to gauge system performance. TOC operates on a "simpler is better" assumption whenever possible, so the fewer indicators management must watch to ensure system success, the better.

It should be noted that any measurement in any management philosophy can be misapplied or abused, to the detriment of the overall system. An obsession with localized measurements runs the risk of "winning the battle but losing the war." For this reason, TOC prescribes keeping management's attention on the system-level measures listed above, rather than on measures of local efficiency.

## The Minor Differences

Lean production includes the major elements of what has come to be known as "total quality management" in North America. The emphasis on "perfection" described earlier is particularly prominent in product and process quality. The Theory of Constraints doesn't address quality, per se. It assumes either a) quality is not the overriding constraint of your system at the moment; or b) quality will be identified as the overriding constraint in the first of the Five Focusing Steps. If your operations don't enjoy a high level of quality, TOC will suggest that you employ the appropriate quality tools to remedy that situation as part of Focusing Steps 2-4. However, TOC doesn't encompass quality tools as part of its basic philosophy the way lean production does. In this respect, the quality improvement tools of lean production fit very nicely with TOC. Some of these tools include: **(9: p.5)**

- Poka-yoke (mistake-proofing operations)
- Statistical quality control (SPC)
- Process capability
- Continuous improvement
- Failure Modes and Effects Analysis (FMEA, both product and process)
- Line stop

Beyond quality alone, lean production provides some discrete logistic management tools that TOC doesn't, but which aren't particularly at odds with the prescriptions of TOC. Some of these tools include: **(9: p.5)**

- Cell design (meaning, in this case, establishing work centers around natural work groups)
- Team roles/responsibilities/rules
- Graphic work instructions
- Visual controls
- Single-minute exchange of die (SMED)
- Five "S" ‡

## How Difficult is Lean to Apply?

It took Toyoda and Ohno more than 20 years to "ramp up" the Toyota Production System in Japan. **(30: p.62)** Even allowing for the fact that they were inventing much of it as they went along, this is still a very long time by the standards of most western companies. And Toyoda and Ohno were working in a somewhat benign cultural environment, where the power of unions to neutralize what management wanted to do was pretty much kept in check. Moreover, the success of lean production, or any other philosophy that depends on employees' willing—even eager—cooperation, contribution, and sometime even subordination of personal desires, is especially challenging to bring about.

Employees respond only when there exists some sense of reciprocal obligation, a sense that management actually values skilled workers, and will make sacrifices to retain them. **(30: p.99)** Toyoda and Ohno were able to achieve this. But doing so exacted a price that not many western companies are willing to pay: guarantees of lifetime employment; a willingness not to lay people off, even in economic downturns; company-provided housing and schools, and other benefits that engendered both loyalty to the company and the camaraderie needed to make team work succeed.

Moreover, Japanese companies benefit from a sympathetic business environment not found in the west. One characteristic of this environment is weak labor unions. Another is a general acceptance by stockholders that they stay in their place, which means: "Be quiet, just keep collecting dividends, be happy with the returns we give you, and don't involve yourselves in what we're doing." And a third is the rise of the *keiretsu,* the interlocking system of ownership among the large industrial companies and their suppliers.‡ The existence of the *keiretsu* smooths the horizontal integration required by lean production in ways that are hard to measure, but which remove many obstacles to effective supplier relations. Womack and Jones cite the example of Nippodenso, at $7 billion the largest supplier in the world for electrical and

---

‡ The five "S's" are *seiri, seiton, seiso, seiketsu,* and *shitsuke,* which roughly translate into English as sifting, sorting, sweeping, standardize, and sustain. The first three terms refer to general housekeeping in the work cell. The last two terms refer to the self-discipline of workers to make the first three happen, and the responsibility of management to see that they do. From Feld, W.M: *Lean Manufacturing Tools, Techniques, and How To Use Them.***(9)**

‡ For more information on *keiretsu,* see Karel Van Wolferen's *The Enigma of Japanese Power* **(27)**.

electronic systems and engine computers. **(30: p.61)** Over 60 percent of Nippodenso's business is supplying Toyota. Toyota holds 22 percent of the equity in Nippodenso. Other Toyota suppliers hold an additional 30 percent, and Robert Bosch (Germany) holds 6 percent. Virtually all of Toyota's suppliers are embedded in this kind of interlocking ownership. In the U.S. in particular, such relationships might draw the immediate attention of the Department of Justice's Anti-Trust Division.

Does this mean that lean production is too difficult to attempt in the west? If "lean production" means a "cookie-cutter" imprint of the Toyota Production System, probably so. However, aspects of what Toyoda and Ohno did are certainly transferable to companies outside Japan. And, as with other management philosophies, this has been done with varying degrees of success. Major Japanese companies (Honda, Toyota, Nissan, NEC, Sony, etc.) have subsidiary operations ("transplants") in the U.S., where the work force doesn't have the job security prevalent in Japan, and where the benefits of *keiretsu* are limited to those parts coming from suppliers in Japan.

Other companies (not Japanese-owned) have chosen instead to selectively apply aspects of the Toyota Production System. The external supply chains in most North American companies aren't as "tightly wrapped" as those in Japan, sometimes complicating the communication, coordination, and performance between parties. Some companies choose only to emphasize the lean manufacturing aspect of lean production, devoting less attention to lean design and lean supply chains. By overlaying these partial efforts on the cultural and business environment in North America, it's no wonder that fully embracing lean thinking is so problematic. With this in mind, let's look at the challenges faced by one North American company.

## Lean Manufacturing: The CAMI Story

In August 1986, General Motors (Canada) and Suzuki Motor Company (Japan) announced a joint venture called CAMI Automotive. CAMI opened its doors for business in Ingersoll, Ontario, in 1989 as a model of lean production. It was intended to be an ideal marriage. Suzuki would obtain needed capital, access to GM's large network of North American dealers, and some shelter from the unfavorable appreciation of the Japanese yen and fears of North American protectionism. GM would have the benefit of the proven track record of the largest producer of small cars in Japan and adherence to U.S.-mandated Corporate Average Fuel Economy (CAFE) standards until its own fuel-efficient cars could be introduced. GM also wanted to use CAMI as a showcase plant to demonstrate to its employees what it called "synchronous manufacturing"—the nuts and bolts of lean production. **(22: p.12)**

But a funny thing happened on the road to CAMI's future. Within three years the wheels seemed to be coming off. In September 1992, after months of intensive negotiations, the Canadian Auto Workers (CAW) approved a strike against CAMI with a 98.9 percent vote, the first time a North American Japanese "transplant" had experienced a strike. At the root of the strike were the working conditions at the CAMI plant. **(22: p.3)**

There were multiple grievances behind the strike, but among the most important were the lean procedures, aspects of management by no means unique to the CAMI plant alone. A major

longitudinal study of CAMI in 1990-91 revealed the reasons why the problems with lean procedures arose.‡

During this period, turnover, absenteeism, and repetitive stress injuries proliferated, in spite of the new, more autonomous and satisfying work place that lean production was supposed to engender. One member of a six-person stamping team on a 10-hour shift at CAMI observed, "When you have downtime, you're supposed to be '5S-ing', but nobody does that anymore." **(22: p.47)** But why not? Are the CAMI workers just lazy slackers? Apparently not.

The absence of "safety nets" and the idea that "everything works or nothing works" is fundamental to lean production. And, as Womack and Jones have observed, "it is the dynamic work team that emerges as the heart of the lean factory." **(30: p.99)** In other words, no-safety-net, everything-works lean manufacturing can't happen without conscientious, motivated work teams. Lean requires proactive cooperation and willing acceptance of additional responsibility to make *kaizen,* 5S, total productive maintenance, and other lean initiatives work. Teams were also intended to provide another pole in the lean tent. Given the problems created by lean staffing policies (compounded by injuries and absenteeism), teams were expected to provide "lateral controls," in the form of peer pressure, to operate as a kind of fail-safe mechanism to promote behaviors consistent with company goals *in the absence of worker commitment to the company and its values.* **(22: p.102)** [emphasis added]

But what do team members see as the outcome of lean manufacturing methods, as executed by teams? One worker at the New United Motors Manufacturing Incorporated (NUMMI) plant (a joint venture of GM and Toyota) in Hayward, California, observed: "We're supposed to go to management and tell them when we have extra seconds to spare. Why would I do that when all that will happen is that you'll take my spare seconds away and work me even harder than before." **(1: p.100)** It should be noted that NUMMI's rigidly standardized jobs have an average cycle time of only 60 seconds. **(22: p.127)** Notwithstanding the job rotation and team work inherent in lean work cells, some researchers have used the term "management by stress" to characterize a system in which employees are subjected to relentless pressure from the pace of work, the absence of buffers and relief workers, managers, and their own team members. **(22: p.9)**

Even the much-heralded *kaizen* is more of a boon to management than it is for the job autonomy or enrichment of the team worker. Mazda workers at its Flat Rock, Michigan, plant initially endorsed *kaizen,* based on promises of job enrichment, but "when it became obvious...that their suggestions were only increasing their own workloads, enthusiasm for *kaizen* all but disappeared." **(10: p.161)** In fact, based on observations and interviews, Parker and Slaughter conclude that *kaizen* translates into "super-Taylorism." They found that almost invariably management dictates the process, the basic production layout, and the techniques to be

‡ Rinehart, Huxley and Robertson conducted comprehensive surveys of 100+ employees and team leaders from throughout the CAMI plant, as well as interviews with managers and union representatives. These surveys/ interviews were conducted four times at six-month intervals over an 18-month period beginning in March 1990 (before the strike). The same questions were asked each time, and the differences in responses noted and tabulated. Results of the study were published in *Just Another Car Factory? Lean Production and Its Discontents.* **(22)**

used. These in turn largely determine job requirements and design, leaving little latitude for team-initiated modification. **(21: p.19)**

A survey of Mazda workers revealed that suggestions from the manufacturing floor, which were carefully screened by managers and engineers, were approved only when they met productivity and cost-down objectives. **(2: pp.1-37)** Such observations aren't confined to North American Japanese "transplants." As far back as 1979, in a study of Toyota Auto Body, Cole described managers' "firm control of the innovative process." Management initiated the problem-solving agenda pursued by workers and rigorously controlled and guided their "activities into channels which flowed toward the achievement of basic management goals." ‡ **(5: p.217)** Guess what those goals might have been. (If you said "cost reduction and increased productivity," you get a gold star!)

What can we conclude from all this? First, the lean manufacturing concept is heavily dependent on the willing, proactive contributions of its component teams. Second, the contributions of these teams are almost exclusively focused (by management emphasis) on changes that both "lean up" the operation (increase productivity, decrease cost) and increase the intensity of work and stress level imposed on the work force. And third, over an extended period, this "continuous improvement" syndrome eventually extinguishes individual motivation to work toward the company's goals. Employees' attention in that direction is sustained only by the informal lateral control inherent in the work team—and, in the case of the Japanese workers, the strong dependency on the company created by job security, socio-economic, and other cultural factors. In western companies (if not in Japanese), the end result of this continual can be worse relations with unions and strikes.

Figure 12 (a, b, c and d) depicts a hierarchy of lean manufacturing intermediate objectives. Each vertical layer in the hierarchy depends on the achievement of objectives in the layer beneath it. As you can see, the diagram is split into four connected parts to simplify viewing. Figure 12a might be called the *human factors* page. Figures 12b and 12c are the *logistics* and *quality* pages, respectively. And Figure 12d shows the overall *business objectives*. Notice the sequence of connections. The human factors page interlocks with both the logistic and quality pages. The logistics and quality pages describe the lean production policies and tools, but the humans in the system must execute those policies and tools if the business objectives are to be achieved. You can establish all the policies and procedures you want, but if the people involved don't care about following them—or worse, if they resist doing so, either passively or actively—the tools and policies will be no more than hollow shells. Moreover, there might be other serious obstacles to achieving some of these intermediate objectives. For example, even if it's possible to eliminate redundant job classifications (block 113), what do you do with the people who performed those jobs? Lay them off to reduce costs or assign them to other work (no cost reduction benefit in this)? Does this become an unpalatable choice, i.e., sacrifice employees or forego cost savings?

---

‡ Even Japanese researchers reached the same conclusion. M. Nemoto, in *Shinshakaihatsu no saizensen* (Tokyo: Nikkagiren, 1992), observed that 80-90 percent of all *kaizen* activity is undertaken by team leaders and foremen.

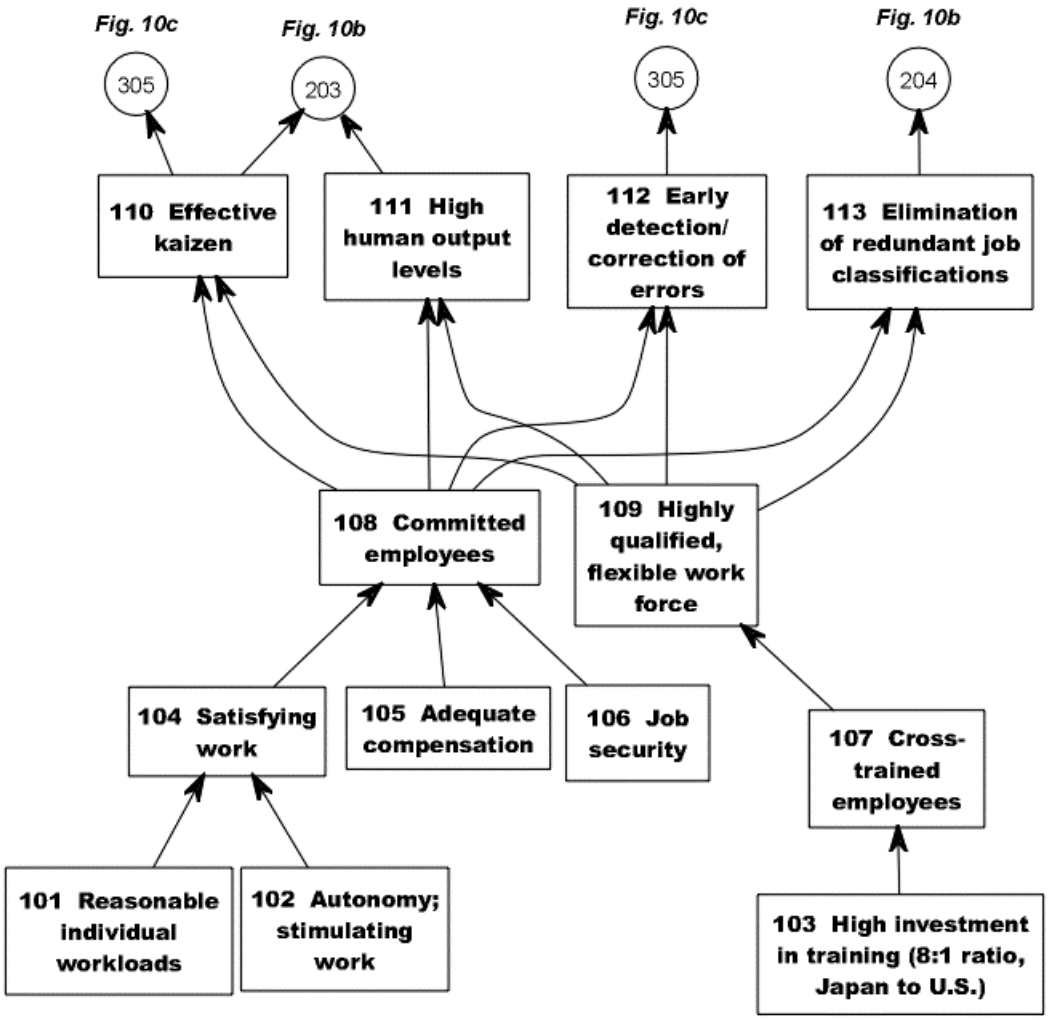



**Figure 12a. Lean Production Intermediate Objectives (1) Human Factors**

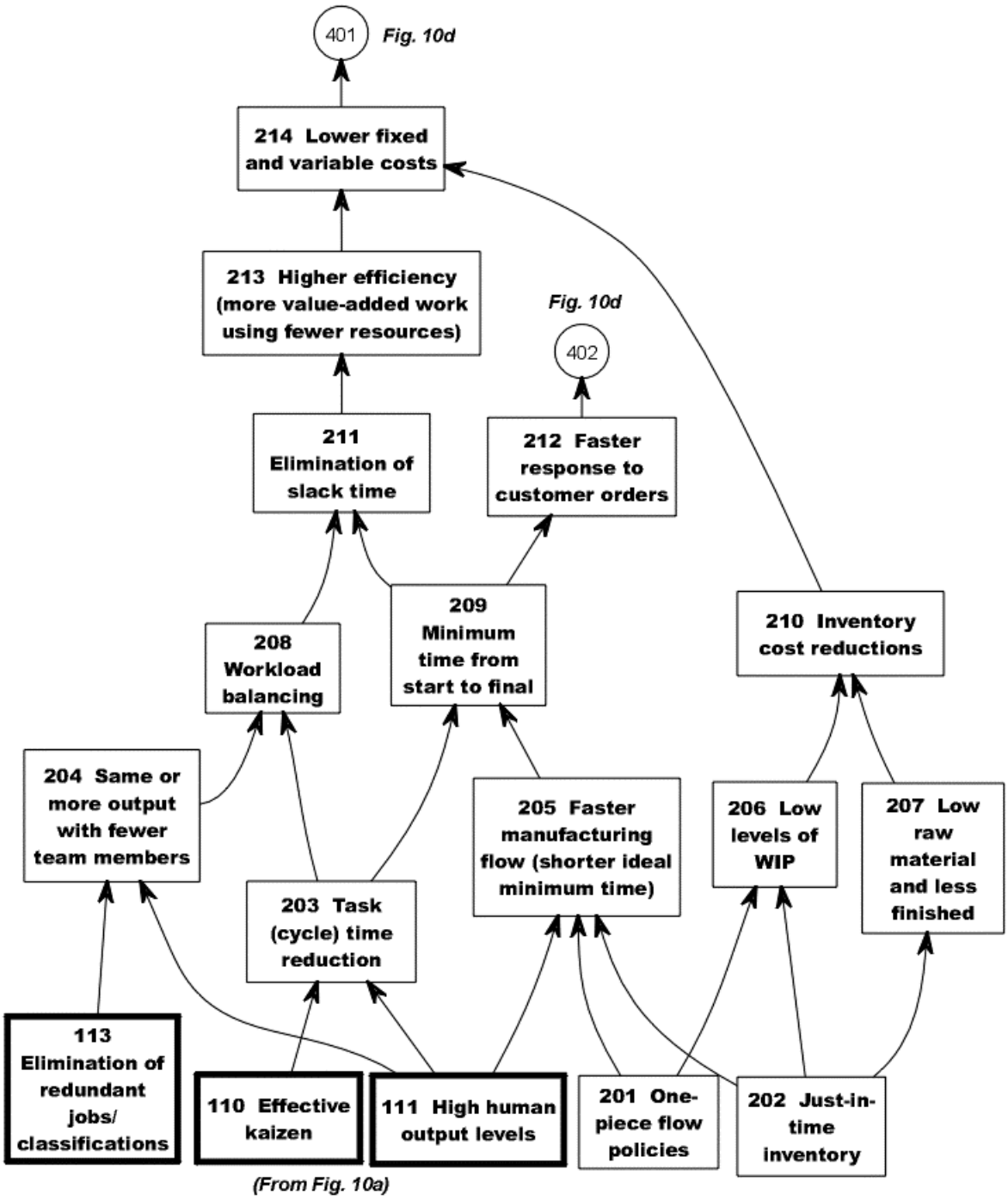



**Figure 12b. Lean Production Intermediate Objectives (2) Logistics**

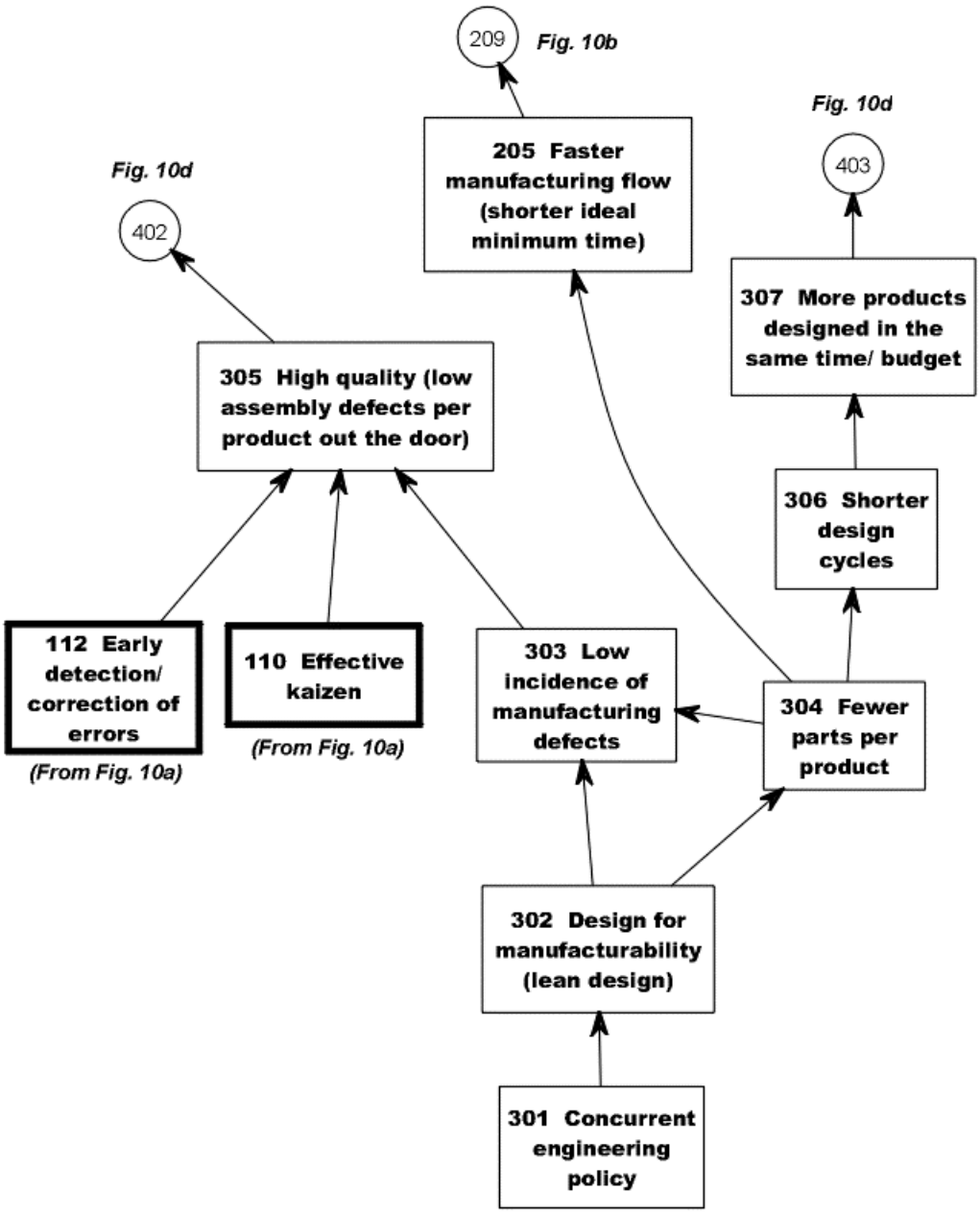



**Figure 12c. Lean Production Intermediate Objectives (3) Quality**

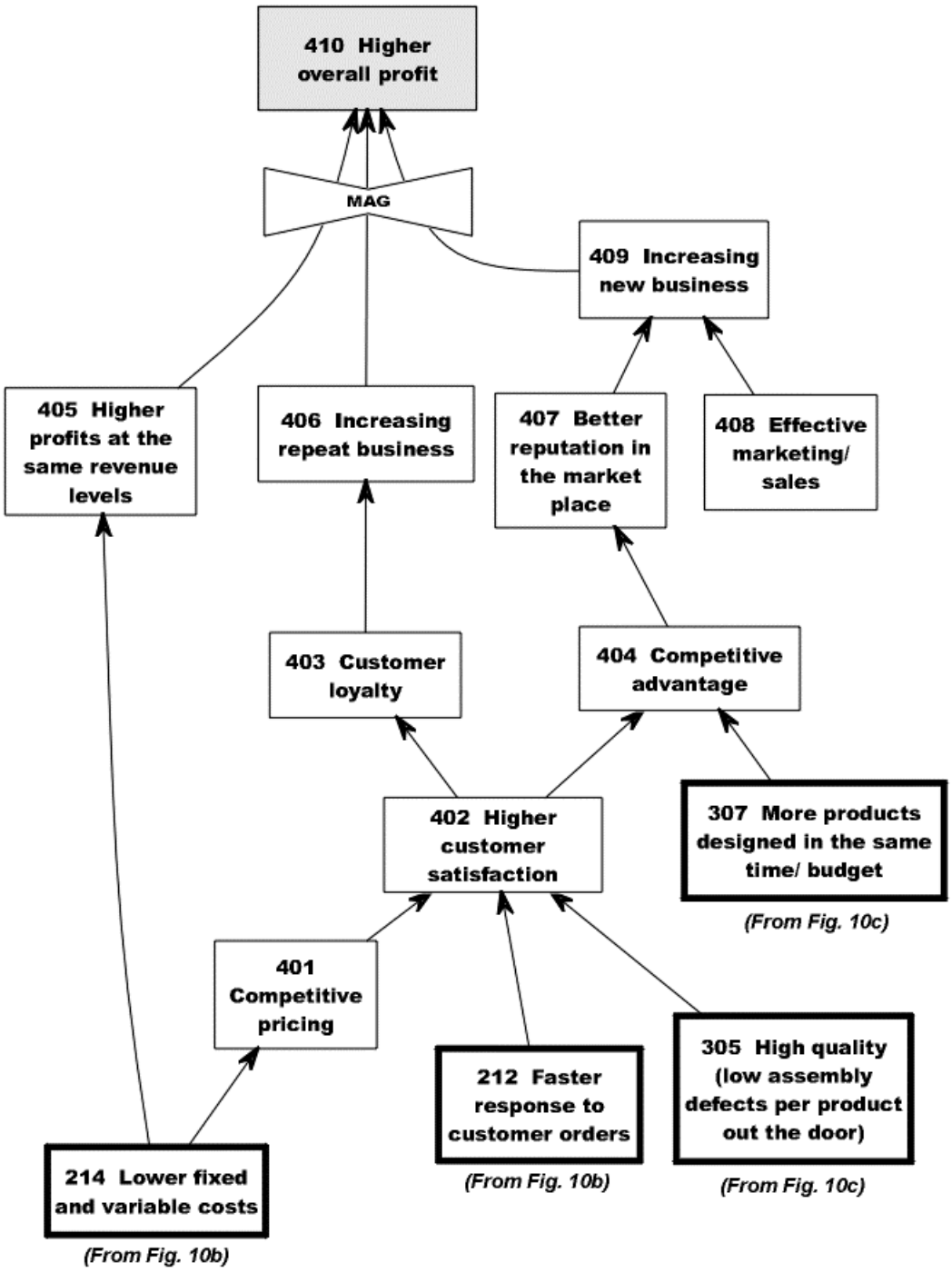



**Figure 12d. Lean Production Intermediate Objectives (4) Business Objectives**

Earlier we saw that in the case of some of the Japanese "transplants," the lean manufacturing paradigm had a rocky introduction. The Canadian Auto Workers union even struck CAMI over these policies and their effects. Though the strike has long since been put to bed, the "honeymoon" between management and labor is probably over. It would be interesting

to see the results of another survey conducted now, such as Rinehart, Huxley and Robertson concluded in 1992, to determine whether (and to what degree) the work force is still pursuing continual quality improvement and elimination of *muda.*

It's likely that these plants and other places where lean manufacturing has been operating for several years have "leaned up" about as far as they can go, and the point of diminishing returns has been reached. Naturally, this raises the question: *Where to go from here?*

## Beyond Lean Manufacturing: The Next Step

As is clearly indicated by Figure 6 *(Cost Reduction Diminishing Returns)*, the well eventually runs dry in reduction of Operating Expense and Inventory. Another way of saying this is that *not many companies have saved their way to prosperity.* **(18: p. 25)** So how does a company keep the "continuing" characteristic in continuous improvement?

One way to do it is merge aspects of lean thinking with the Theory of Constraints. As Figure 6 also indicates, the potential to expand Throughput is considerably greater than the potential to be realized from cost and inventory reductions. TOC is designed with Throughput expansion in mind. However, merging the two philosophies requires a conscious decision to re-prioritize both objectives and efforts to some degree, and to accept the idea that some aspects of lean thinking might create as many (or more) problems than they solve. Specifically, the overarching emphasis on cost reduction and maximizing local efficiency everywhere in the system needs to be rethought. Proper application of constraint thinking and TOC principles can help a company avoid the human factor "downside" experienced by CAMI and other comparable organizations.

## INTEGRATING LEAN MANUFACTURING AND THEORY OF CONSTRAINTS: A PRESCRIPTION

What follows is a general prescription for combining lean thinking and the Theory of Constraints. It's not a detailed "cookbook;" rather, it's more of a conceptual framework. The details will undoubtedly require customizing in individual organizations. The important thing is to remain faithful to the overall concepts.

Moore and Scheinkopf suggest that the first two steps should be to adopt a Throughput perspective and to define the boundaries of the system to be improved, determine its purpose, and analyze how that purpose is measured. **(18: p.32)** In other words, they suggest foregoing intensive cost and inventory reduction, instead working to gain internal control of the system before attempting any major "forward-looking" initiatives. In many companies, gaining control is an improvement effort in and of itself. And it is one that is uniquely suited to the quality tools and methods associated with lean thinking.

Once the system is in reasonably stable control, tools of lean such as the *kaikaku* process can be used to chart the value stream and dispense with steps that are obviously and totally unnecessary, and easy to eliminate. The reason for doing so is not to reduce costs, but to quickly reduce the number of dependencies in the company's chain. Fewer dependencies mean less opportunity for "Murphy" to do his dirty work. Another key action at this point is to provide for some protective capacity. **(18: p.32)** Accepting the idea that protective capacity is a good thing

supports the Throughput concept, allows for shorter buffer times, and enables less disruption of workflow through the system.

Once a Throughput perspective has been established and the system defined, the Five Focusing Steps of constraint theory can be applied. Applying the first step (Identify the system's constraint) presupposes acceptance of the "chain of dependent links" concept of systems, but the process of determining the value stream should help establish the system's weakest link.

Exploiting that weakest link, the second of the five steps, is a natural point at which to apply lean techniques. Reducing setup times *at the constraint,* perhaps using Single Minute Exchange of Dies (SMED) or other techniques, contributes immediately to liberating hidden or trapped capacity to increase Throughput. At non-constraint, setup reduction might be less important, except where it consumes so much time that it threatens to turn a non-constraint into a constraint. That, too, would be a candidate for setup time reduction. Other kinds of *muda* might include use of skilled labor at the constrained resource to do unskilled or menial tasks, while those same skills might be in short supply. You should be seeing a pattern here: The best application of lean techniques, including *kaizen,* is at the capacity-constraining resource. Anywhere else, it not only ineffective, it might actually be harmful (e.g., turning a non-constraint into a system constraint).

Moore and Scheinkopf cite the importance of really understanding the Throughput (financial) value of contemplated or rejected actions. **(18: p.33)** Let's say the system's hourly financial output, governed by the optimum pace at which the constraint operates, is $1,300 per hour. If the constraint operator, following perfectly valid lean thinking, spends an hour of his or her time performing "5-S" functions, that's an opportunity loss of $1,300. On the other hand, hiring a non-skilled worker at $10.50 per hour to do the constraint operator's "5-S" functions for that hour alone frees the constraint operator to generate an additional $1,300 in Throughput for the company. For the other seven hours of the non-skilled worker's time doing the same functions at non-constraints, he or she could be relieving the non-constraint operator to perform Total Productive Maintenance, attend training, or train replacement workers—all functions that could immediately contribute to increased Throughput. However, keep in mind that unswerving commitment to truly lean thinking would probably allow the sacrifice of that $1,300 in Throughput in the interest of saving the non-skilled worker's $10.50 per hour.

The third step in TOC's Five Focusing Steps is to *subordinate* everything else to the system's constraint. This means maintaining protective capacity at non-constraints, so that they can always catch up from a disruption in time to preclude "starving" the system constraint. That means they won't be totally efficient. But since increasing efficiencies everywhere doesn't actually change the amount of money spent (unless protective capacity work force is trimmed), the cost savings are illusory.

Moreover, maximizing efficiency isn't practical unless capacity is fairly well balanced throughout the system. And balancing capacity with workload may drive efficiency up, but it can create chaos when "Murphy" strikes, especially if protective capacity has been eliminated. (And if it hasn't no cost saving accrues.) Moreover, as the CAMI case illustrates, driving the work force to extremely high efficiencies everywhere through high intensity workloads can have

devastating effects on the human element of the business equation so crucial to success.

However, the *kanban* concept of lean manufacturing is comparable to TOC's subordination idea, which serves to limit the introduction of material into the manufacturing process to a rate that the capacity-constrained resource can handle. The chief difference is that lean seeks to balance the line, where TOC suggests not attempting to do so (and to deliberately introduce a constraint if the line is already balanced—it makes managing flow so much easier!)

The fourth of TOC's Five Focusing Steps, *elevating* the constraint, which invariably means spending more money (additional capital equipment, hiring more employees) in order to make even more money to more than offset the added expenditure. This is, quite simply, an expansion of capacity, and both lean thinking and TOC would defer this decision until it was definitely determined that the existing system was as efficient as it could be, while still guaranteeing the protective capacity needed to respond to both internal variation and external uncertainty.

## CONCLUSION

Figure 13 groups the various lean production tools and methods with the Five Focusing Steps to which they apply. It should be clear that there is substantial overlap between the lean thinking paradigm and the Theory of Constraints paradigm. In summary, TOC provides a useful system-level framework for directing lean thinking efforts where they will do the most good (the system constraint) and avoiding the pitfalls of applying them where they will do harm.

**1. IDENTIFY the system's constraint (TOC)**
- Identify the value stream (Lean)
- Product/quantity assessment (Lean)
- Process mapping (Lean)
- Routing analysis (Lean)
- Capacity determination (TOC)
- Cell layout/design (Lean)
- Standard work (Lean)
- Roles and Responsibilities (Lean)

**2. Decide how to EXPLOIT the system's constraint (TOC)**
- Kanban sizing (Lean)
  - Transfer batch sizing (TOC)
- One-Piece flow (Lean)
  - Process batch sizing (TOC)
- Backward plan (TOC)
- "Drum" (TOC)
- *SMED -CCR only (Lean)
- *Poka-Yoke -CCR only (Lean)
- *Kaizen -CCR only (Lean)
- Graphical Work instructions (Lean)

**3. SUBORDINATE everything else to the decision in Step 2 (TOC)**
- Kanban pull signal (Lean)
  - "Rope" (TOC)
- "Buffer" (TOC)
- **5S housekeeping - Non-CCR (Lean)
- **SMED - Non-CCR (Lean)
- **Total Productive Maintenance - Non-CCR (Lean)
- **Kaizen - Non-CCR (Lean)
- **Training (Lean, TOC)

**4. ELEVATE the system's constraint (TOC)**

**5. GO BACK to Step 1, but beware of inertia (TOC)**

***NOTE 1:*** *Activities indicated by a single star (*) are a combined engineering and system operator effort at the CCR. These activities are performed outside of normal production schedules (off-shifts, weekends, if possible) to minimize CCR downtime. These are HIGH PRIORITY efforts required to maximize available CCR capacity.*

***NOTE 2:*** *Activities indicated by a double star (**) are largely system operator initiated and managed efforts at all non-CCRs (i.e., everywhere else except the CCR). These activities are performed during normal shifts, if possible, during idle time between production jobs. These are LOWER priority efforts, unless a non-CCR is in danger of becoming a CCR.*

**Figure 13. Integrating Lean Thinking and Theory of Constraints**

## REFERENCES


1. Adler, Paul S. "Time and Motion Regained." *Harvard Business Review,* Jan-Feb 1993.

2. Babson, Steve (ed.). *Lean Work: Empowerment and Exploitation in the Global Auto Industry.* Detroit: Wayne State University Press, 1993.

3. Caspari, John. "Ch. 8A - Theory of Constraints," in Keller-Bulloch-Shultis, *Management Accountant's Handbook (4th ed.) 1993 Supplement.* NY: John Wiley & Sons, 1994.

4. Caspari, John."TOC Based Bonus System--a Constraints Accounting Approach." http://members.home.net/casparija/aweb/bonus.htm; http://members.home.net/casparija/aweb/bonus.htm#t3.

5. Cole, Robert E. *Work, Mobility and Participation: A Comparative Study of American and Japanese Industry.* Berkeley: University of California Press, 1979.

6. Corbett, Thomas. *Throughput Accounting.* MA: The North River Press, 1998.

7. Cox, James F., III, and Michael S. Spencer. *The Constraints Management Handbook.* FL: St. Lucie Press, 1998.

8. Dettmer, H. William. *Breaking the Constraints to World-Class Performance.* Milwaukee, WI: ASQ Quality Press, 1998.

9. Feld, William M. *Lean Manufacturing: Tools, Techniques, and How to Use Them.* Boca Raton, FL: St. Lucie Press, 2001.

10. Fucini, Joseph, and Suzy Fucini. *Working for the Japanese: Inside Mazda's Auto Plant.* NY: The Free Press, 1990.

11. Goldratt, Eliyahu M. *Critical Chain.* MA: The North River Press, 1997.

12. Goldratt, Eliyahu M. *The Goal* (2nd ed.). NY: The North River Press, 1992.

13. Goldratt, Eliyahu M. *The Haystack Syndrome: Sifting Information Out of The Data Ocean*. NY: The North River Press, 1990.

14. Goldratt, Eliyahu M., and Robert E. Fox. *The Race.* NY: The North River Press, 1987.

15. Leach, Lawrence P. *Critical Chain Project Management.* Boston: Artech House, 2000.

16. Mabin, Victoria J., and Steven J. Balderstone. *The World of the Theory of Constraints: A Review of the International Literature.* FL: St. Lucie Press, 2000.

17. Monden, Yasuhiro. *Toyota Production System: An Integrated Approach to Just-in-Time* (3rd ed.). Norcorss, GA: Engineering & Management Press, 1998.

18. Moore, Richard, and Lisa Scheinkopf. "Theory of Constraints and Lean Manufacturing: Friends or Foes?" (Unpublished paper) Chesapeake Consulting, Inc., 1998.

19. Newbold, Robert C. *Project Management in the Fast Lane: Applying the Theory of Constraints.* FL: St. Lucie Press, 1998.

20. Noreen, Eric, Debra Smith, and James T. Mackey. *The Theory of Constraints and Its Implications for Management Accounting.* MA: The North River Press, 1995.

21. Parker, Mike, and Jane Slaughter. *Choosing Sides: Unions and the Team Concept.* Boston: South End Press, 1988.

22. Rinehart, James, Christopher Huxley, and David Robertson. *Just Another Car Factory?* Ithaca, NY: Cornell University Press, 1997.

23. Schragenheim, Eli, and H. William Dettmer. *Manufacturing at Warp Speed: Optimizing Supply Chain Financial Performance*. Boca Raton, FL: St. Lucie Press, 2000.

24. Sheridan, John. "Throughput With a Capital T." *Industry Week.* March 1991

25. Smith, Debra. *The Measurement Nightmare: How the Theory of Constraints Can Resolve Conflicting Strategies, Policies and Measures.* FL: St. Lucie Press, 2000.

26. Spear, Steven, and H. Kent Bowen. "Decoding the DNA of the Toyota Production System." *The Harvard Business Review,* September-October 1999, pp. 97-106.

27. Van Wolferen, Karel. *The Enigma of Japanese Power.* NY: Vintage Books (Random House), 1990.

28. Womack, James P., and Daniel T. Jones. "From Lean Production to the Lean Enterprise." *The Harvard Business Review,* March-April 1994, pp. 93-103.

29. Womack, James P., and Daniel T. Jones. *Lean Thinking: Banish Waste and Create Wealth in Your Corporation.* NY: Simon and Schuster, 1996.

30. Womack, James P., Daniel T. Jones, and Daniel Roos. *The Machine That Changed the World*. NY: Harper Perennial, 1991.