Supply Chain Responsiveness and the Inventory Illusion

This paper identifies the critical dimensions of supply chain responsiveness and analyzes and evaluates the role of inventory as a driver of responsiveness. Our analysis leads us to conclude that, far from being an unequivocal driver inventory is, in many cases, a net destroyer of supply chain responsiveness. The analysis leads us to conclude that more often than not, inventory represents dead-weight that slows down a company’s response to market demand and supply conditions, and this, for a broad range of the thirteen responsiveness dimensions. Even when inventory contributes to quick market demand and supply responsiveness, it does so in the short term, and these short-term advantage become a significant disadvantage in the long term. The paper establishes a direct link between inventory and quality responsiveness. The long term responsiveness disadvantage of inventory is particularly acute in the case of a company’s quality responsiveness. Because of the critical importance of quality to competitive performance in modern markets, this paper identifies the quality impact of supply chain decisions as a promising avenue for future research.

Over the last decade or so, the design and management of the supply chain to achieve high levels of responsiveness has emerged as one of the most critical issues in the strategic positioning of a company through use of its supply chain. It is increasingly being recognized that the competitive performance of companies such as Dell, Amazon.com, Walmart and McDonald’s, to name only a few, derives substantially from the particular supply chain design that they have engineered and effectively deploy on a day-to-day basis. That responsiveness is a major driver of competitive performance is now well established. However, on the whole, inventory is accepted to be a contributor to supply chain responsiveness. In what follows, we critically evaluate the role of inventory in the achievement of responsiveness and our evaluation suggests that, contrary to accepted thinking, inventory may be a substantial net destroyer of supply chain responsiveness.

Nature of Supply Chain Responsiveness and the Impact of Inventory

Chopra and Meindl (2004) define supply chain responsiveness in the following terms; “Supply chain responsiveness includes a supply chain’s ability to do the following; Respond to wide ranges of quantities demanded; Meet short lead times; Handle a large variety of products; Build highly innovative products; Meet a very high service level; Handle supply uncertainty”. A similar framework based on lean and agile concepts avers that the appropriate design and operation of the supply chain is driven by two critical factors, the degree of product variety that the firm’s is called on to produce and the variability or uncertainty of market
Supply Chain Forum

An International Journal

Vol. 6 - N°1 - 2005

www.supplychain-forum.com

demand. According to that framework, agile supply chain designs are best adapted to less predictable environments where demand is volatile, and thus uncertain, and the requirement for variety is high. Lean supply chains are suited to high volume, low variety and predictable environments (Christopher, 1998; Christopher and Towill, 2000; Towill and Christopher 2003). Therefore, the agile supply chain is also the responsive one. These frameworks underscore the idea that responsiveness is multidimensional, which leads us to anticipate that the sources of responsiveness may be varied as well. Given this multidimensional nature of supply chain responsiveness, we conjecture that inventory may enhance some types of responsiveness while destroying others and may, on balance, create a net responsiveness disadvantage. For present purposes, responsiveness can be defined as the characteristic of the supply chain system that endows it with the capability to effectively and efficiently perform a broad range of changing requirements of the end-product market.

The nature of the ambiguity as to the role of inventory in supply chain responsiveness and the opposing conclusions that it leads to are clearly revealed from the following position taken by Chopra and Meindl (2004): “Inventory is a major source of cost in a supply chain and it has a huge impact on responsiveness. If we think of the responsiveness spectrum discussed in Chapter 2, the location and quantity of inventory can move the supply chain from one end of the spectrum to the other. For example, an apparel supply chain with high inventory levels at the retail stage has a high level of responsiveness because a consumer can walk into a store and walk out with the shirt they were looking for. In contrast, an apparel supply chain with little inventory would be very unresponsive. A customer wanting a shirt would have to order it and wait several weeks or even months for it to be manufactured”. On the other hand, they argue that; “Inventory also has a significant impact on the material flow time in a supply chain. Material flow time is the time that elapses between the time materials enter the supply chain to the point at which it is used...The logical conclusion here is that inventory and flow time are synonymous in a supply chain. Managers should use actions that lower the amount of inventory needed without increasing cost or reducing responsiveness, because reduced flow time can be a significant advantage in a supply chain.” Inventory increases flow time which is, unequivocally, a measure of decreased responsiveness. So, according to the accepted position articulated in the supply chain management literature, (Blackburn, 1991; Bozarth and Handfield, 2006; Chopra and Meindl, 2004), inventory is presented as both increasing and decreasing responsiveness. The position we take here is that supplying an order when the customer wants it is responsiveness, but having to hold high levels of inventory to do so is not and destroys the very responsiveness that inventory purportedly provides.

Inventory is a substantial net destroyer of supply chain responsiveness.

Dimensions of Supply Chain Responsiveness and the Impact of Inventory

The Overall View of Inventory and Supply Chain Responsiveness:

Table 1 outlines thirteen dimensions of supply chain responsiveness and confirms the conjecture deduced from the literature to the effect that supply chain responsiveness is a variegated concept that incorporates many dimensions (Austin et al, 1985 ; Blackburn, 1991; Bozarth and Handfield 2006; Christopher, 1998; Christopher and Towill, 2000; Fuller et al, 1993; Haas, 1994; Handfield and Pannesi 1992; Handfield and Nichols, 2002; Kivenko,1994; Monczka et al, 2005; Nickols et al, 1995; Nickols, 1996; Tryendal, et al, 1998). The framework also leads to the conclusion that supply chain responsiveness incorporates many more dimensions than what is currently recognized. What is most instructive, however, is the fact that the impact of inventory on supply chain responsiveness varies significantly depending on the dimension of responsiveness under consideration. Critical for present purposes is the notion that when it comes to supply chain responsiveness, even when it is examined at the micro-level, the role of inventory is neither neutral nor insignificant (Lee and Billington, 1992). Inventory does appear to have an impact on supply chain responsiveness but, contrary to the expectations based on current literature, the overall impact is likely to be very negative instead of positive. Moreover, the framework also instructs us that, for some variables that generate uncertainty, the impact on supply chain responsiveness for an increase in the variable may not be the same as for a decrease, and the role of inventory varies rather dramatically. For example, it is common to view demand variability, whether an increase or a decrease, as the relevant focus of analysis when it comes to supply chain responsiveness. But the framework of Table 1 tells us that these are two separate responsiveness phenomena that have radically opposite inventory impacts. Inventory has a positive impact on upward demand shift responsiveness but a largely negative impact on unpredictable downward demand shift responsiveness.

Based on an analysis of the framework of Table 1, it appears that inventory is a substantial net destroyer of supply chain responsiveness rather than being a net enhancer of it. Inventory is
### Table 1
Dimensions of Supply Chain Responsiveness and Relationship to Inventory

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Definition/key characteristics</th>
<th>Impact of Inventory</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Upward demand shift Unpredictable</td>
<td>Capability to satisfy high fill rates when demand unexpectedly increases.</td>
<td>Positive; Inventory helps satisfy immediate delivery (Chopra and Meindl, 2004; Lee and Billington, 1992; Monczka et al., 2005).</td>
</tr>
<tr>
<td>2. Intra-product line demand shift</td>
<td>Capability to achieve high fill rates when demand for one product in line increases at expense of another; Overall demand for line unchanged; Demand volatility exists within line.</td>
<td>Positive; Inventory increases responsiveness for product whose demand has increased Negative impact for product whose demand has decreased, because it will take longer to adjust inventory levels to the new, lower demand.</td>
</tr>
<tr>
<td>3. Unpredictable downward demand shift: Individual product</td>
<td>Capability to quickly adjust production rates and inventory levels to an unpredictably lower demand level.</td>
<td>Negative impact; High inventory means more time to bring production and inventory in line with new, lower demand.</td>
</tr>
<tr>
<td>4. Lead time shift; decrease in customer lead time</td>
<td>Capability to meet high demand/order fill rates when customers unexpectedly tighten their lead time requirements.</td>
<td>Positive impact; High levels of inventory serve as buffer. (Chopra and Meindl, 2004; Fisher, 1997; Lee, 2002; Lee and Billington, 1992). But positive inventory impact comes at expense of lower capability to adjust production/inventory levels to lower demand.</td>
</tr>
<tr>
<td>5. Market niche</td>
<td>Capability to dovetail product line to unique requirements of a large number of market niches.</td>
<td>Negative impact; broadening range of market niches derives from capability to produce, transport, stock product at point-of-sale in small lots, which drives individual item inventory down.</td>
</tr>
<tr>
<td>7. Time-to-market/Model and Design Change</td>
<td>Capability to speed up market entry of new/improved product and models with minimal impact on existing products.</td>
<td>Negative impact; Inventory increases flow time, obsolescence costs (Magretta, 1998; Magretta, 1998); High inventory slows down taking new products to market; Company is squeezed between writing off old inventory and sacrificing market entry speed.</td>
</tr>
<tr>
<td>8. Predictable demand shift (Increase)</td>
<td>Capability to increase production rates and inventory levels in anticipation of increased demand.</td>
<td>Positive impact, since less addition to inventory is required to build up to a new, higher demand level (Chopra and Meindl, 2004).</td>
</tr>
<tr>
<td>9. Predictable demand shift (Decrease)</td>
<td>Capability to decrease inventory, production levels in anticipation of decreased demand.</td>
<td>Negative impact; Inventory adds to flow time, slowing downward adjustment process.</td>
</tr>
<tr>
<td>10. Supply uncertainty</td>
<td>Capability to achieve acceptable fill rates when probability of breakdown in supply is high.</td>
<td>Positive impact, but there are usually more efficient means of protecting against supply uncertainty (Lee, 2002).</td>
</tr>
<tr>
<td>11. Quality: Feedback from market</td>
<td>Speed of quality improvement or causes of defects identified in response to customer experiences with the product.</td>
<td>Negative; Inventory lengthens quality feedback loop, which slows down speed of feedback from market and lengthens time to react to experience quality.</td>
</tr>
<tr>
<td>12. Quality: Feed-forward to market</td>
<td>Speed of signal to market, through actual use, that product has been improved.</td>
<td>Negative; Inventory increases flow time, which reduces speed with which quality improvements can be taken to market to judge experience quality.</td>
</tr>
<tr>
<td>13. Net Conversion Cycle (NCC)</td>
<td>Capability to reduce NCC and bring company to position where operations generate enough cash to sustain business without recourse to external financing.</td>
<td>Negative; Inventory adds to NCC (Stalk;1998); For most businesses, inventory is biggest component of an unfavorable NCC position. Most businesses with a favorable NCC position use pull/JIT systems which compress inventory.</td>
</tr>
</tbody>
</table>
ordinarily referred to as a driver of supply chain responsiveness, where by driver is meant that inventory has a positive impact on responsiveness. However, analysis based on the framework of Table 1 would suggest that this widely held view of the role of inventory may need to be revised, and this fundamental insight is partly supported by the responsiveness performance of JIT/Pull and lean inventory systems. Inventory has a positive impact on unpredictable upward demand shift responsiveness, predictable increase in demand responsiveness and supply uncertainty responsiveness. For two dimensions of responsiveness, intra-product-line demand shift and decrease in customer lead time responsiveness, the impact of inventory is partly positive and partly negative. For the remaining dimensions of supply chain responsiveness, the impact of inventory is substantially negative.

Inventory and Unpredictable Upward Shift in Demand Responsiveness:

There are two ways that companies create the requisite supply chain responsiveness to deal with upward, unpredictable shift in end-product-market demand. In the traditional approach, and the one has received by far and away the most attention in supply chain management domain and which is still favored by most companies, a finished goods safety stock is used to cushion the impact of unpredictable changes in demand, the level of safety stock being determined by a service level designed to optimize the tradeoff between inventory holding cost during the critical demand period, which is usually the lead time, and the stock-out or shortage cost. And since inventory in the form of safety stocks does assuage the visible, negative sales impact of upward demand shift uncertainty, this has given rise to the erroneous notion that inventory is a source of supply chain responsiveness. However, it is increasingly being recognized that dependence on inventory to deal with unpredictable, upward shift in demand is a costly and ineffective strategy. It is costly because inventory holding costs are much higher than what is recognized in traditional supply chain systems, particularly when one considers the inventory write-offs caused by technological obsolescence or design/fashion changes. Safety stock is also ineffective as a responsiveness strategy because while end-item safety stock increases immediate supply chain responsiveness in terms of high order/demand fill rates, it sacrifices other types of responsiveness, since it increases flow time.

The second approach to building a highly responsive supply chain to deal with unpredictable upward shifts in demand recognizes that safety stocks create inventory which sacrifice too much of the other dimensions of responsiveness and, therefore, that the overall impact of a strategy that uses safety stocks is negative. A company uses safety stocks when it does not possess the requisite level of responsiveness, in the first place. So, companies that are designing and deploying highly responsive supply chain strategies attempt to build in the value creating, supply chain processes themselves. Inbound logistics, Operations, Outbound Logistics, Sales and After-sales- a level of supply chain flexibility that will give them the wherewithal to respond to unpredictable, upward shifts in demand with minimal safety stocks.

As a company increases its innate level of responsiveness, safety stocks can be systematically compressed out of the system. The premise that safety stocks are deployed to mitigate the impact of an unresponsive supply chain process is clearly seen in the very model that designs a kanban/pull process, which aims to build and exploit the ultimate in terms of supply system responsiveness. The equation for the number of kanbans for coordinating supply and demand between two supply chain processes is given by \( Y \), where

\[
Y = D(T_p + T_w)(1 + \alpha) / k 
\]

\( D = \text{Demand per hour} \)

\( T_p = \text{Production time per batch (kanban container)} \)

\( T_w = \text{Wait time in hours per batch (kanban container)} \)

\( \alpha = \text{Safety factor} \)

\( k = \text{kanban container capacity} \)

The safety stock is built into the kanban process, and it is given by \( SS \), where

\[
SS = D(T_p + T_w)\alpha
\]

Given that for any state of the system \( \alpha \) is constant, equations 1 and 2 mean that the safety stock is entirely driven by \( T_p + T_w \). This is the time for the system to respond to pull signals from the downstream operation for products or parts to replenish the quantity that the downstream operation has just started to use. Since \( \alpha \) is the safety factor that assures continuity of supply in the event of supply disruption at the preceding operation, as \( T_p + T_w \) is reduced, the system becomes more responsive, and this automatically compresses the safety stock even if the safety factor remains constant. For example, when demand, \( D \), equals 100 per hour, \( \alpha \) equals 0.10 and \( T_p + T_w = 4 \), the safety stock equals 40 units. But for the same hourly demand and safety factor, the safety stock equals 10 units when \( T_p + T_w = 1 \).

All pull, kanbanized supply chain processes are structured to capitalize on that fundamental premise that safety stocks only exist because the current supply chain system is slow to respond to changes in market demand. Safety stocks minimize the disruption in supply flow, given the existing level of innate supply chain system responsiveness. This is a rather crucial idea that the people at Toyota and other advanced Japanese JIT companies discovered some time ago; Work-in-process inventory is more appropriately seen, not as a driver of responsiveness, but as a stopgap measure to assure the negative consequences of an unresponsive system. Remove the source of inflexibility and the need for safety stocks and work-in-process inventory will be greatly reduced. If inventory is a driver of
responsiveness, there is no way that a JIT/pull production system can be more responsive than its traditional counterpart.

**Predictable Increase in Demand Responsiveness**

Unpredictable increases in demand stretch a company's demand responsiveness capability to the very limit because they require both high planning and high execution competency. Predictable shifts in demand can still pose major responsiveness challenges, even if they usually require only planning system flexibility to change plans in anticipation of forecasted changes in demand. If these predictable shifts are for the long term, the major drivers relate to the factors that give a company the capability to structure new capacity and bring it on-stream quickly. Where the upward, predictable shift in demand is of an intermediate-term nature, the responsiveness challenge reduces to the usual aggregate production and inventory problem and the drivers are the hire/fire/transfer costs associated with increasing the intermediate-term production rate and level of inventory. High levels of aggregate inventory are usually viewed as a driver of intermediate-term predictable demand shift responsiveness because less addition to inventory is required to build up to the new, higher demand level.

**Supply Uncertainty Responsiveness**

It goes without saying that supply system uncertainty is a major obstacle to the achievement of a high level of responsiveness of the overall supply chain. Moreover, because the supply system is not under the direct, administrative control of the company, supply uncertainty presents unique challenges to the management of supply chain responsiveness. The problems caused by supply uncertainty are particularly acute for JIT-In-Time/Pull/Lean supply chains, because a company cannot be highly responsive and deliver goods to market just-in-time, if the supply systems do not deliver the parts and other inputs that are required to make the product, even if not just-in-time. Just-in-time/Pull/Lean supply chains are designed to operate on the premise and operationalize the Kaizen principle that inventory is waste and should be systematically compressed from supply chain processes. These systems pursue high demand and order fill rates, not by creating high levels of inventory, but by first-of-all reducing demand variability through the Andon principles of load leveling. When reasonable load-leveling has been achieved, Just-in-Time/Pull/Lean supply chain systems seek high levels of responsiveness by building innate flexibility into supply chain planning, scheduling and execution processes. Just-in-Time/Pull/Lean supply chain systems are prototypical responsive supply chain systems, which casts doubt on the notion that inventory is a source of supply chain responsiveness. Just-in-Time systems view the so-called responsiveness given by inventory as being both illusory and costly.

Supply uncertainty comes from two sources and can affect either quantity or time. The first source derives from the fact that the customer company has changed the fundamental parameters of the supply situation by either increasing the quantities demanded from a supply system or decreasing the supply response time required of suppliers. This type of supply situation simply places greater demands, time or quantity, on suppliers. The second source of supply uncertainty derives from the fact that neither quantities nor the usual supplier lead time has changed, but there is a breakdown in the supply processes that link a company to its suppliers. In the case of supply uncertainty due to quantities, the customer increases the requirements from a base level and the supplier must adjust deliveries. In that case, a responsive supplier is one who can deliver the increased quantities within the normal lead time window. That is to say, a responsive supplier is able to deliver increased demand from its customers without sacrificing normal delivery performance. In the second type of supply uncertainty, the customer company does not change its supply quantities but requires that these be delivered according to a compressed delivery schedule, either because the customer company's own customers are demanding shorter lead times or because the customer company has recognized the strategic and marketing advantages of response time compression, and is in the process of driving supplier performance in that direction. Increasing quantities and decreasing time is a matter of the supplier having the capability to create and manage continuous improvement in supply processes through the timely and rapid addition of capacity and the capability to compress time, particularly queue time, in the manufacturing process.

There are a number of strategies for dealing with supply uncertainty and for assuring high end-product market responsiveness when there is a breakdown in supply continuity. These include safety time in time-phased requirements planning systems such as MRP, MRP-II and DRP, source redundancy or the use of multiple sources, source flexibility through the selection and certification of sources that have built innately flexible, responsive supply chain systems, the adoption of capacity strategies at the level of suppliers that load factories at less-than-full capacity, that is, the maintenance of some capacity cushion in excess of what is required to produce peak demand, the cultivation of high levels of source reliability, holding an appropriate level of safety stock of parts and subassemblies, requiring that suppliers hold appropriate levels of inventory to cope with breakdowns in their own supply processes, rerouting of supply to reallocate requirements and supply a region from warehouses or factories that are further afield even if this violates the optimal allocation of warehouses to supply chain stages.
further downstream, shipping less than full, economic size loads to speed up delivery after a breakdown and producing less than optimal lot sizes to compress manufacturing lead time and, ultimately, the total delivery time, and temporary compression of supply/transportation times in the event of breakdown in the supply system. The latter can be accomplished through the use of alternative, but usually more costly, transportation modes to accelerate delivery after a breakdown in supply, and after the supply system has recovered from that breakdown.

Maintaining supply chain responsiveness in the event of a breakdown in supply comes at a price, so all these strategies enumerated above create single-event supply chain costs that are borne within that immediate response time frame, and the cost is incurred every time action is taken to maintain responsiveness in the face of a breakdown. However, in a tightly coordinated supply chain that uses highly capable, responsive suppliers, breakdowns in supply would be the exception rather than the rule, and these suppliers would be most adept at implementing and executing the most effective strategies for maintaining high levels of supply chain responsiveness in the event of an unexpected breakdown in supply. Because supply system breakdown in that context is the exception, these short-term, emergency responsiveness strategies are less costly than those that rely on inventory. Short-term, emergency responsiveness strategies incur higher costs only during the few periods where supply disruptions occur, whereas strategies based on inventory keep materials and finished goods permanently in the system and, therefore, create permanent costs to maintain responsiveness in the face of clearly exceptional events. Although inventory has a positive impact on supply uncertainty responsiveness, other responsiveness strategies as enumerated above could provide the requisite level of responsiveness more efficiently.

**Intra-product Line Demand Shift Responsiveness: The Two-Edged Impact of Inventory**

When the overall demand for a product line remains basically constant, but demand shifts between products in the line, what we refer to as intra-product line demand volatility, this creates a double responsiveness problem for a company. To the extent that customers perceive no reduction in service when they switch between the different product options in the line, the impact of intra-product line demand volatility is assuaged significantly. However, that usually is not the case in a competitive market, because it is logical to assume that product options offered in the line are based on real and significant variations in customer requirements. Where demand of one product option is decreasing while that for another option in the same line is increasing, it invariably means that the former product is becoming less competitively attractive than the latter and that customers do not see the products in the line as being readily substitutable. In that case, the responsiveness problem for the product whose demand has increased is analogous to upward demand shift responsiveness for an individual product, as discussed below, and inventory contributes positively to responsiveness. For the product whose demand has decreased, the responsiveness problem is similar to downward demand shift responsiveness, and inventory is a destroyer of responsiveness. A company faced with intra-product line demand shift must deal with two responsiveness problems simultaneously, one requiring a ramp-up of production to meet an unpredictable upward shift in demand, as discussed above, and one requiring a deceleration of the production rate and a compression of inventory levels to adjust to a lower demand rate for the product whose demand is decreasing, as discussed below. Where the facilities that manufacture the products and move them through the supply chain are not dedicated, the responsiveness problem presents few challenges. Almost always, however, there will be some facility dedication, and such a demand shift would tax the responsiveness capability of a company. In the case of intra-product demand shift, inventory is a true double-edged sword.

**Individual Product Downward Demand Shift Responsiveness**

We define downward demand shift responsiveness as the capability of the supply chain system to adjust production levels, supply or inventory downward to bring them in line with a lower demand rate. Current supply chain management literature does not recognize a responsiveness problem for unpredictable downward shifts in demand, but the problem exists and is as prevalent, and possible as costly, as that created by unpredictable, upward shifts in demand. The reasons for this observation are not hard to uncover. In all competitive markets, systematic, prolonged increases in demand will always be followed by decreases. Moreover, the costs imposed on a firm by downward shifts in demand are potentially very high. A decreasing demand places a squeeze on profits from three different directions. First, demand contraction usually derives from increased competition which puts a squeeze on prices and profit margins. Second, decreased demand means that the company has less volume over which to dilute fixed costs, which further decreases profits. Third, in the absence of high responsiveness to drive inventory out of the supply chain faster than demand is contracting, decreased demand increases inventory levels at the very time that the product is less appealing to customers, and this exacerbates obsolescence costs, inventory financing costs, the net conversion cycle, and a wide array of other costs. In other words, downward shift in demand means that per units costs are increasing at the same time that the company is recouping less margins to absorb these costs.
Inventory has very negative consequences for the downward demand shift responsiveness of a supply chain because it significantly delays the process of adjusting supply to demand. A company that already holds a high level of inventory and comes under severe downward demand pressure is faced with the dilemma of either making dramatic cutbacks in supply/production to quickly adjust supply to demand, or implementing a more normal, slower response while holding a higher level of inventory for a longer period of time. The first response amplifies the fluctuations in the labor force and in raw materials and parts delivery from suppliers, which generate inefficiencies in labor and procurement costs.

The second response exposes the company to very high obsolescence costs, particularly because the products that are left in inventory would most likely be the slower moving, less technologically attractive ones that are likely to be driven to obsolescence during the adjustment period. This is exactly what happened to Lucent when demand for its product plummeted at the beginning of the last economic downturn that started around 2000. The company already had very high levels of inventory of telecommunications gear that were already technological surpassed by superior competitive products. When the downturn came, Lucent's high level of inventory drastically slowed down its adjustment process and, by the time the economy had turned around, the company's product line had become largely obsolete.

The point is that although it has not been fully recognized by supply chain management theory and practice, downward demand shift responsiveness may be as crucial to a company's competitive advantage as upward demand shift responsiveness, and inventory, far from being a contributor to downward demand shift responsiveness, destroys it.

### Lead Time Responsiveness

Supply chain lead time comes into play in the management of responsiveness at three different levels, and the impact of inventory on that dimension of performance varies from one level of the supply chain to the next. First, there is the customer or end-item demand level where lead time responsiveness refers to the speed with which products can be delivered to customers and whether the company can maintain delivery performance when customer lead time is shorter than usual or, more importantly, is being compressed due to competitive market pressures. Obviously, customer lead-time responsiveness has the most direct and significant impact on a company's market performance. Second, there is the manufacturing or in-process level where lead time refers to the speed with which products can flow through the production process. While short term, low lead time performance can be cushioned by end-item inventory that is placed between the production process and the customer delivery point, persistent, low long term manufacturing lead time performance will require increasingly higher levels of buffer inventories. This makes it a costly strategy to rely on buffer inventories to provide high levels of market end-item demand responsiveness with an unresponsive manufacturing process. Third, there is supplier-level where lead time responsiveness refers to the speed with which suppliers can deliver materials and parts or the extent to which they can maintain delivery performance in the face of a compressed lead time requirement.

#### Customer lead time responsiveness

Customers put a high value on lead time compression because a shorter lead time means both quick response to market and substantially compressed safety stocks. The latter impact is the result of at least four factors that operate in concert. Customer lead time compression reduces the forecast horizon which lowers overall forecast error, results in a shorter lead time and demand during lead time, a key factor driving safety stocks, decreases demand variability during lead time, which comes from both the reduced overall demand volatility and the shorter lead time, and a lower normal distribution z-factor. The reduced value of z comes from the fact that the supply chain can operate with a lower target service level during lead time because rapid replenishment means that the periods of shortages are less protracted which drives shortage costs down. We recall that the optimal service level as defined by the critical fractile method is given by P, where,

\[
P = \frac{G}{G + L} \quad \text{(3)}
\]

G = Shortage cost per unit
L = Inventory holding cost during lead time

Because G is usually much larger than L, sometimes by up to two orders of magnitude, which has the effect of pushing the optimal service level close to 100 percent in a competitive market, G is the principal driver of the target optimal service level. Therefore, a decrease in the shortage cost will likely drive the target service level down, leading to a decrease in z and the safety stock. A compressed lead time means that the customer can satisfy its own market demand with less inventory, and can capitalize on unpredictable increases in end-product demand through increased responsiveness, providing that the end-product production system can compress its own manufacturing lead time. Otherwise, compressed customer lead times simply mean that inventory has been backed up the supply chain, and that the customer has unloaded part of its flow time, pipeline inventory and forecast horizon onto the supplier, with no change in overall supply chain responsiveness. Consequently, true improvement in end-item lead time responsiveness depends critically on manufacturing lead time responsiveness.
Inventory has a positive impact on compressed customer lead time responsiveness. Inventory will cushion the impact of a compressed lead time, since the company can deplete its inventory on hand in order to satisfy a shorter lead time. On the other hand, inventory has a negative impact on lead time responsiveness when the requirement is for a longer lead time, because inventory slows down the process of adjusting inventory and production levels to a longer lead time. However, as we noted previously, customers place high value lead time compression so that in competitive markets the case of lead time extension is likely to be so seldom, as to make it competitively insignificant.

**Manufacturing lead time responsiveness**

For most companies, the manufacturing system is a major destroyer of overall supply chain responsiveness. Although there is much focus on the transportation system as a key constraint to overall supply chain responsiveness, the fact is that more time is consumed pushing the flow of goods through the manufacturing system than to move them from where they are manufactured to where they are consumed. It is an anomaly of modern industrial and commercial processes that many goods can be transported from locations on the planet that are farthest apart in less time than it takes to push the flow of goods from the raw materials or parts warehouse of a company through its own factories. And the biggest culprit is not the actual being transformed. It is an anomaly of modern industrial and commercial processes that many goods can be transported from locations on the planet that are farthest apart in less time than it takes to push the flow of goods from the raw materials or parts warehouse of a company through its own factories. The evidence is that queue time accounts for the lion’s share of total manufacturing lead time, accounting for as much as 90 percent, while move time and production time account for roughly 2 percent and 8 percent, respectively. That is to say, for a manufacturing lead time of 20 days, queue time accounts for 18 days, while move time and production time account for 0.4 days (3.2 hours), and 1.6 days, respectively. The irony is that, in the heart of the production system where goods are produced, products spend much more time waiting to be transformed than they spend actually being transformed.

**JIT/Pull systems** represent the pinnacle of supply chain system responsiveness, and the fundamental JIT/Pull equation underscores the role that manufacturing lead time and the production system play in driving supply chain responsiveness. According to the fundamental JIT/Pull equation, a company can implement JIT/Pull manufacturing where production actions are synchronously triggered by market demand only if the following inequality is satisfied:

\[ \text{LTCUMM} > \text{LTCUST} \]

where, \( \text{LTCUMM} = \text{lead time required by customers} \)

If this fundamental inequality is satisfied, the company can undertake JIT/Pull production and launch production in response to customer orders. For example, if \( \text{LTCUMM} = 10 \text{ days} \) while customers require 15 days, the company can accept a customer’s order, keep it as information in the order entry system for 5 days, launch production on the sixth day, move the order through the manufacturing cycle in 10 days and deliver it JIT to the customer. If \( \text{LTCUMM} = 15 \text{ days} \) and \( \text{LTCUST} = 10 \text{ days} \), the company can only satisfy demand from finished goods inventory using a push system, or by decreasing the total cumulative manufacturing lead time by keeping inventories of work-in-process from which production of customer orders will be launched. If \( \text{LTCUMM} \) is 15 days and \( \text{LTCUST} = 10 \text{ days} \), then the company must either satisfy demand from finished goods inventory or stock the equivalent of at least 5 days of manufacturing lead time in work-in-process, usually in the form of subassemblies and modules. This latter approach is exactly the one used by Dell. Where \( \text{LTCUMM} > \text{LTCUST} \), the speed with which the supply chain can respond to customer demand is slower than what customers require, making it necessary to use inventory. The greater the difference between innate supply chain responsiveness and customer-mandated responsiveness.
as expressed in LT\textsubscript{CUST}, the greater the level of inventory that must be kept by the system.

In the above example where LT\textsubscript{CUMM} is 15 days and customer lead time is 10 days, based on the experience of most companies, queue time typically would be 90 percent or 13.5 days. By achieving a forty percent compression in queue time, one could re-engineer the manufacturing system and transform it from an inventory-based push system to a zero-inventory, JIT/Pull system. Queue time compression is of paramount importance in creating a high responsiveness, JIT/Pull system, making it necessary to understand the forces that are driving queue time in a manufacturing system.

Queue time is driven by two fundamental factors in the production stage of a supply chain. The first factor, input/output control, measures the extent to which the input loading of machine centers, transformation processes and the overall production system are matched with their output capability or capacity. This is, fundamentally, an internal manufacturing system coordination problem. Where there is not a coordinated release of work to machine centers, these will be overloaded, causing work-in-process inventory to accumulate just ahead of the overloaded processes which creates queues, the very source of queue time. Such lack of coordination that results in machine center overloading could not occur in a manufacturing system that is scheduled using JIT/Pull mechanisms. In particular, Kanban has a built-in Jidoka/Andon process that stops the process when there is a problem or that synchronizes upstream work center loading with the volume pulled by downstream work centers. Because lack of coordination is one of the distinctive features of an unresponsive supply chain, the build up of work-in-process inventory that results from it, far from being a contributor to responsiveness is, in point of fact, a symptom of an unresponsive manufacturing component of the supply chain.

The second factor driving queue time is the batch size used to launch production of parts, subassemblies and finished goods. The batch size has a direct impact on two parameters of the production system, the inventory level and the flow time, both of which have ramifications for the responsiveness of the manufacturing level component of the supply chain. The batch size directly determines the average level of cycle inventory because, according to the basic production lot sizing model, the optimal production lot or batch size is given by EPQ, where,

$$\text{EPQ} = \sqrt{\frac{2DS}{h(1-d/p)}}$$

where,

- \(D\) = annual demand
- \(S\) = setup cost
- \(d\) = demand rate
- \(p\) = production rate of item

The average inventory is given by EPQ/2. It can be shown that the EPQ has a built-in safety stock effect, which increases with the batch size (Etienne 1985 and 1987). Demand variability which determines the size of stockouts during the period prior to the order point is independent of the batch size and average inventory. With larger lot sizes there is a higher average inventory and a larger cushion for absorbing demand variability. However, the increased service level protection given by larger lot sizes is wasted or redundant and cannot be a minimum cost service level, since it is in addition to the optimal service level that the safety stock is designed to provide. So, while the additional inventory that comes with larger batch size increases responsiveness in terms of demand or order fill rates, it may do so at too high a cost, and imposes all the negative responsiveness impacts of inventory outlined previously.

The major impact of the batch size on manufacturing system responsiveness has to do with its effect on queue time (Goldratt and Cox, 1984; Goldratt and Fox, 1986). It is easy to demonstrate, as is done in Figure 1, that even where there is near perfect coordination of input and output as exists in a JIT/Pull kanbanized system, queue time will be significant if batch sizes are large. This is the crucial reason why JIT/pull systems absolutely require small lot production and theory of constraints advocates making the transfer batch significantly smaller than the production batch. One simply cannot have high responsiveness at the manufacturing system level without small lot sizes and the commensurate decrease in queue time.

Figure 1 shows demonstrates the relationship between batch size and queue time in a simple, single-facility manufacturing system in
which we assume, for simplicity, that the transfer batch is equal to the production batch and, therefore, the queue time for the whole batch is the same as that for the last unit. The complexity of the manufacturing system does not change the fundamental relationships revealed in the schematic. The batch size is tantamount to units queuing up for processing in front of the production facility, where any unit accumulates queue time as a function of the number of units ahead of it and the processing time per unit plus the setup time. This is analogous to the situation that obtains in a service facility where every unit has to wait through the service processing times of all units ahead of it. The larger the batch size, the longer units further up the queue have to wait for processing. Again assuming perfect input/output control and coordination, all units except the very first one processed accumulates queue time, and the larger the batch size, the more processing times accumulated by later units in the queue, which increases the average queue time for the batch as a whole. Figure 1 shows that when one increases the batch size from 5 to 15, average queue time increases from 1.4 to 2.4 hours.

The above analysis shows that instead of increasing manufacturing system lead time responsiveness, inventory decreases it since the level of work-in-process inventory and queue time are both driven by the same fundamental factor, the batch size. Real lead time responsiveness comes from batch size compression which drives work-in-process inventory out of the manufacturing system and dramatically reduces queue time, the largest component of manufacturing lead time. Unresponsive manufacturing systems have long setup times and schedule production of large batch sizes in a vain attempt to assuage the inefficiencies of an inflexible system. The work-in-process inventory lodged in a manufacturing system, far from improving responsiveness, destroys it. A responsive system replaces inventory with quick changeovers, timely and accurate information and strong coordination to synchronize demand requirements with manufacturing supply actions.

**Market Niche Responsiveness**

It is the capability to economically and efficiently produce, transport, buy and make the product available at the point of sale in very small lots that drives market niche responsiveness. Toyota’s mixed model processing production system is the archetypical case of a niche-responsive manufacturing company. Toyota has the capability to continuously produce on its assembly lines the daily sales mix for all of the options of a family of cars which means that the company can theoretically achieve finished goods stock turns of upwards of five hundred. Economical, small lot production and delivery of parts and subassemblies and small lot production allow the company to produce smaller batches to meet the lower demand levels of more narrowly defined, smaller volume segments of the market. The logic of the Toyota Production System says that if a company produces a product X for a broadly-defined market segment and the company, through improvements in the supply and production processes, reduces the economical lot size of the product by a factor of four, then that company can define four narrower segments based on more unique customer requirements and produce unique product options for each segment, without sacrificing cost or increasing inventory. In transportation – rapid courier delivery – UPS appears to

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Inventory Turns and the Economical Broadening of Product variety</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Pre-Inventory Compression</td>
</tr>
<tr>
<td></td>
<td>Product A</td>
</tr>
<tr>
<td>Demand</td>
<td>12000</td>
</tr>
<tr>
<td>Annual</td>
<td></td>
</tr>
<tr>
<td>Holding cost/unit</td>
<td>1</td>
</tr>
<tr>
<td>Ordering cost/order</td>
<td>160</td>
</tr>
<tr>
<td>Ordering cost, four SKUs/order</td>
<td>2.5</td>
</tr>
<tr>
<td>Required space/unit; square feet</td>
<td>1</td>
</tr>
<tr>
<td>EOQ</td>
<td>1960</td>
</tr>
<tr>
<td>Ave. Inventory</td>
<td>980</td>
</tr>
<tr>
<td>Inventory Turns/YR</td>
<td>12.24</td>
</tr>
<tr>
<td>Required space</td>
<td>1960</td>
</tr>
</tbody>
</table>
be the most niche-responsive company, while Walmart is, by far and away, the prototypical niche-responsive company in retailing. As shown in Table 2, by increasing overall inventory turns, a company can increase the range of product offerings at the point of sale, without increasing inventory, space, overhead and infrastructural costs. By quadrupling inventory turns, for example, a retail company like Walmart can increase the number of product options offered at the point of sale without increasing space costs, since the space liberated by higher stock turns can be used to offer four different options of the same product.

There are two strategies for the economical broadening of variety and increasing market niche responsiveness that results from inventory compression. The first involves using the supply chain resources liberated by inventory compression to offer products that cater to completely new market segments. Increasing product variety in this way generates the maximum level of sales because it enables the company both to enter new but related product offering segments and exploits synergy between items in the product line to the fullest. Synergy exists because of two different phenomena. The first is the substitution effect, whereby a customer who comes to a store to buy a particular product is more likely to find a good substitute if the exact product that he or she wanted to buy is momentarily out-of-stock. In that case, variety means that the company makes a sale that it would otherwise have lost. The second factor contributing to variety synergy is the complementarity effect, which captures the frequently observed phenomenon where a customer buys many more items than he/she had the intention of buying when they entered the store. The complementarity effect means that the company sells two items instead of one, and the greater the variety of products in the product offering, the more likely that is to occur. The complementarity effect means that the company can increase responsiveness by reducing inventory and simultaneously reducing cost. Wal-Mart has used that approach to systematically broaden its product offering over the years and add food retailing, pharmaceuticals, toys and automobiles to its core product offerings. We use space cost as a proxy for the overall supply chain resources and capacity liberated by inventory compression. The data presented in Table 2 show that a company that has driven stock turns up by a factor of four can use the liberated resources to increase segment scope by a factor of four, without increasing overall cost.

The second strategy for leveraging inventory compression to economically broaden the product offering involves targeting the same general market segment, but broadening the variety of product options that are offered. Although this strategy will have a lesser positive impact on sales growth, since it simply slices up a more broadly defined segment into a larger number of narrower ones, the company should experience some increase in sales, because the increased product variety that results has favorable internal marketing effects. Greater variety communicates an aura of broad choice and availability to the customer, and that, coupled with the positive symbiotic relationship among a large variety of different options of the same product, will likely cause overall sales for the entire family to increase, and this will reduce costs per sales dollar even further. Thus, high inventory, because it reduces the efficiency of deployment of supply chain capacity resources, decreases market-niche responsiveness and prevents a company from reaping the marketing and strategic advantages of high variety at low cost. In addition, it is evident that a company cannot achieve high levels of market niche responsiveness at low cost unless it systematically compresses inventory out of its supply chain. Inventory rather than increasing responsiveness, reduces it because of the negative impact that high levels of inventory have on market-niche responsiveness.

Technological Change Responsiveness

Accelerating technological change and the rapid obsolescence that results from it is a major concern for an increasing number of companies. Companies are faced with increasingly shorter product life cycles and the capability to respond to advances in technological is an integral part of the company’s competitive advantage. Companies either have to drive their own products to obsolescence or their competitors will do it for them. Inventory both reduces a company’s ability to respond quickly to technological change and makes it costly for it to do so. In industries that are highly fashion sensitive where competitive pressures mandate that companies change product designs frequently, high inventory levels reduce the freshness of the collection, and can be disastrous for a company’s market responsiveness. When inventory levels of items that are going out-of-fashion are high, management has the tendency of delaying somewhat the introduction of new models or designs to give the supply chain system, particularly the point-of-sale component, time to clean up the old stock. This is so because high levels of inventory that occur towards the end of the season force management to dramatically write down the old stock in an effort to entice value-conscious customers to buy what is obviously an out-of-style item. Write-downs of the order of fifty to eighty percent are not unusual in these industries. There is a natural tendency on the part of managers to delay introduction of the new, upcoming designs in an effort to mitigate the exorbitant cost of write-downs.

Time-to-Market Responsiveness

Related to technological change responsiveness is time-to-market,
or the period that elapses between the launching of a product innovation initiative in the form of a product idea, and the delivery of the first units of the product to customers. Time-to-market measures innovation speed, and the ability to improve it faster than competitors confers unsurpassed competitive advantage on an innovator. Companies that arrive to market ahead of their rivals reap first-mover advantages that include the ability to build market share, achieve economies of scale early, recoup R&D costs that can be deployed in other innovation projects in the pipeline, the ability to drive costs down through learning and experience curve effects, and the building of brand recognition and consumer familiarity with the company's version of the new product or technology ahead of competitors. Sometimes, as is the case with Microsoft's Windows and Intel's Pentium, first-mover advantages are so overpowering that later entrants into the market can take years, even decades, to assuage or reverse the competitive advantages of the innovator.

While some may argue that inventory of parts and subassemblies may help a company compress time-to-market, the reverse is actually the case. If a product is highly innovative, it most likely will make very little use of older technology, and inventory of existing parts and subassemblies will not allow a company to achieve any meaningful reduction in time-to-market. Second, the highly innovative technologies that really make the product new and give it its competitive advantage will require the procurement or development and production of new parts and subassemblies that embody the new technology. The lead time for these parts and subassemblies will determine the lead time to manufacture and deliver the new product and even though inventory of the older parts and subassemblies will decrease the lead time to acquire them and build them into the new product, these latter parts do not represent the key technological or time constraint for the new product, and matter little from the perspective of time-to-market. What is critical for time-to-market is the level of coordination of suppliers and the companies own production facilities to ensure that the new parts and subassemblies that are absolutely essential for delivering the new product to market are available on time. On the other hand, knowledge that new parts and subassemblies are available in inventory can have the unintended consequence of causing supply chain managers to pay less attention to the coordination of supply and lead to unexpected shortages.

**Quality Responsiveness**

Companies that compress time to market reap first mover advantages in terms of larger market share, higher margins, earlier and more deeply entrenched brand positions and recognition, advanced progress on the learning and experience curves, all of which drive costs down and margins up. What is most significant and often overlooked, however, are the positive and overwhelming quality and related competitive impacts associated with first-to-market and quick-time-to-market responsiveness. Because no matter what a company does and how much it is vigorous and calculating in anticipating customer requirements, a few products will not be taken to market completely fault-free. The true test of quality comes with experience that the customer/consumer gains from using the product, what is referred to as experience as opposed to search and credence quality (Darby and Karni, 1973; Heskett et al, 1990; Nelson, 1970).

The links between supply chain management and quality are many and diverse, and it is beyond the scope of this paper to elucidate all of them. Here, we focus on inventory policy and management as an enabler or hindrance to quality responsiveness. More specifically, we argue that instead of being an enabler of quality responsiveness and, by implication, overall responsiveness, inventory destroys the firm's quality responsiveness and, necessarily, overall supply chain and firm responsiveness. Supply chain quality responsiveness refers to the speed with which improved products/services and/or parts can be brought to market so that the company can begin to evaluate the extent to which the product or part is having the desired strategic impact and to measure the accuracy with which customer requirements were specified and deployed in the product. That type of responsiveness provides competitive signaling to the company or its suppliers as to whether the product is performing as expected and having the anticipated market impact, from a quality perspective, and how the product or part should be changed to meet or surpass customer requirements. Supply chain quality responsiveness is of two broad types; Quality feed-forward to market, which refers to the speed with which improved products or parts are delivered to real customers so that experience quality can start to be evaluated and quality feedback from market which measures how fast experience quality data flow back to the company or its suppliers. Additionally, to be strategically meaningful, supply chain responsiveness must include quality responsiveness as a critical dimension of supply chain performance, both at the level of supply/sourcing and at the level of market/demand. The conceptual frame of Figure 2 structures and links these quality responsiveness concepts.

**Feed-forward to Market/Source Quality Responsiveness**

Companies can employ a variety of tools, techniques and approaches to ensure that products are taken to market fast and as defect free as possible. These tools include Concurrent Engineering, Quality Function Deployment ‘QFD’, Taguchi Methods, Poka Yoke methods, Failure Mode Effect and Criticality Analysis ‘FMECA’, and competitive benchmarking.
Although these are robust tools and the overwhelming evidence is that they greatly accelerate the speed of product development, introduction, and delivery through the supply chain, the experience of even the best of companies is that few products will not arrive in the market completely error or defect free. A quality giant like Toyota which has pioneered and/or mastered nearly all the cutting-edge tools for removing defects prior to and during product manufacture, sometimes discovers defects after products hit the market and the company has had to go through the occasional recall. Experience quality is a critical dimension of the quality of most products and services, which means that it is difficult to completely anticipate and expunge all quality problems before the product is used by a broad range of customers. Feed-forward quality responsiveness has two dimensions. The first has to do with the speed with which a company can take new products or services to market, and thereby signal to customers and consumers that quality has improved. If the improvement pushes the company's quality performance beyond the strategic breakpoint (Haas, 1987), then the result will be a dramatic shift of customers towards the company's product and away from those of competitors. The second aspect of feed-forward quality responsiveness is the speed with which suppliers can signal to a company that they have improved quality, so that the company can move to integrate the improved components and parts into its own products and services and thereby capitalize on a supplier, source of competitive advantage. Whether it be customer/consumer or supplier type feed-forward quality responsiveness, is critically important to a company's ability to leverage innovation for competitive advantage.

Inventory destroys both types of feed-forward quality responsiveness. Inventory adds to flow time and slows down the process of taking new products or improved supplier components to market and reduces the degree of competitive advantage that can be derived from innovation. Inventory decreases product freshness, a crucial dimension of quality for perishable products and those with limited shelf lives, and causes high levels of product obsolescence and the related costs, particularly where products are fashion-driven, as is the case for a companies like Doré-Doré (Wong and Hammond, 1991), Benetton (Signorelli and eskett, 1984) and Nike (Austin et al, 1985), or where product life cycles are short, as in the case of chip manufacturing and high technology consumer electronics. In the case of Doré-Doré (Audris, 1991), Benetton (Signorelli, 1984) and Nike (Austin, 1985), markdowns of 25-50 percent are not unusual when inventory remains at the point of sale after the close of a particular season. Considering that for Benetton, like most companies, the net profit margin is about 4 percent of sales while the gross profit margin is 40 percent of sales, the whole net profit of the company comes from the last 10 percent of sales, the 10 percent beyond the breakeven point, and it only takes the average markdown of 37.5 percent on a mere 10.6 percent of sales to wipe out the entire net profit of the company.

So, a company can spend enormous resources and effort to push sales beyond the breakeven point only to see the benefits destroyed by the high markdowns associated with high inventory levels and low inventory turns. Benetton, like many other companies, simply cannot afford to have low inventory turns and the corresponding high levels of markdowns. For pharmaceutical compounds, storage time decreases product potency and efficacy, and regulatory authorities mandate that expiration dates be clearly indicated on these products and that they be destroyed upon reaching these expiration dates. Moreover, high levels of inventory and the associated low inventory turns expose a product to defects introduced by the warehousing and inventorying processes themselves, because these processes are not quality-neutral, contrary to what many companies believe. Not only do products deteriorate in storage and the higher the level of inventory, the greater the storage time and associated deterioration, but supply chain managers often overlook the fact that storage is actually a quality impacting process whose operations cause defects. The greater the storage time, the more the supply chain system exposes inventory to defects that are caused by the storage process itself.

A company that is turning its own finished goods inventory fifty-two times a year can take improved products to market in one week, while a company that is turning inventory four times a year must wait three months to draw down existing inventory of the product before it can take quality improvements to market. Ten weeks can be like an eternity in the world of innovation driven, global

<table>
<thead>
<tr>
<th>Quality Responsiveness Target/Source</th>
<th>Source/supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feed-forward</td>
<td></td>
</tr>
<tr>
<td>Responsiveness Type</td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td></td>
</tr>
<tr>
<td>Speed of taking improved quality to market, thus signaling to customers that quality has changed and increasing their expectations</td>
<td>Speed of which suppliers bring new and improved components/parts to market so customers 'manufacturers' can deploy these to improve their products</td>
</tr>
<tr>
<td>Speed of which company receives feedback from customers as to true quality performance of product based on experience of use</td>
<td>Speed of signaling to suppliers true quality of components/parts based on experience in production, after-sales support of product, as basis for improving them</td>
</tr>
</tbody>
</table>
competition. Similarly, inventory destroys the ability of a company to take new supplier quality to market, because the company must draw down the supply of supplier parts inventory before it can replace the old parts with the new, quality improved ones. The farther downstream the company is from the supplier that made the innovation, the more inventory slows down feed-forward supplier quality responsiveness, and the more such inventory is damaging to the company's ability to leverage the competitive advantage from supplier innovation. Analogously, the farther upstream the company is from the final customer/consumer, the greater is the damage inflicted by inventory on a company's customer feedforward responsiveness. Figure 3 provides some computations that support these fundamental insights. As a base case scenario, we assume inventory turns of twelve and three-hundred-and-sixty-five per year for work-in-process inventory and for the product at the end-consumer/customer level, respectively, and of six per year for all other levels of the supply chain. The corresponding flow times due to inventory are shown in the figure. For a second case scenario, we double the inventory turns at all levels except at the customer level. The calculations show that for the inventory turns of the base case scenario, which are well within currently achieved levels for most industrial and commercial supply chain operations, inventory slows down supplier quality feedforward responsiveness by thirteen months and feedforward customer/consumer quality responsiveness by nine months. In either case, inventory slows down feedforward quality responsiveness enough to destroy a company's innovative efforts and is sufficient to completely eliminate a company's first mover advantages in most modern-day, competitive markets. A delay of between nine and thirteen months to take innovations to market gives competitors enough time to develop and implement competitive responses, except where the innovator has substantial innovation advantage that is protected in a solid patent that cannot be circumvented in the short to intermediate run. Similarly, a doubling of inventory turns from the base scenario predictable compresses quality feedforward/feedback time by half.

Feedback from Market Responsiveness

For the vast majority of products, experience quality is the key to building and exploiting first-mover advantages from innovation. When a company takes new products to market, it needs to quickly demonstrate that the new product is, in fact, an improvement and worthy of the user's switching efforts and costs. No matter what a company promises prior to and during new product launching and ramp-up to market demand, there is no innovation until the customer/consumer experiences the quality and that information diffuses rapidly through the market. In some cases, and even where the company has a vigorous QFD program, the new product will have bugs that were not identified during product testing and these defects are discovered when customers start to use the product (HBS, 1989). In the final analysis, the only true, final and conclusive product test occurs when real customers use the product in real consumption situations. Consequently, rapid quality feedback from the market is the key to isolating design defects and flaws, new process weaknesses and production quality control system inadequacies that cause defective new products to be taken to market. Anything that slows down quality feedback from the market also destroys a company's quality responsiveness advantage and seriously blunts its innovative efforts.

The arguments that support the case that inventory destroys quality feedback responsiveness are similar to those that make the case for the rather dramatic negative impact that inventory has on feedforward quality responsiveness. The calculations of Figure 3 make the case. In fact, the inventory induced delay of eleven months in the case of supplier feedback quality responsiveness is the same as the inventory induced delay.
delay for the supplier feedforward quality responsiveness; and the inventory-induced delay of seven months in the case of the manufacturer’s feedback quality responsiveness is the same as the inventory induced delay of seven months for the manufacturer’s feedforward quality responsiveness. This is so because neither the supplier nor the end-product company can get feedback from the market until the product actually reaches the end customer/consumer. Before such feedback can take place, the supply chain system must deplete the inventory of the old product and, according to the calculations illustrated in Figure 3, that depletion process will take eleven months in the case of supplier-produced innovation and nine months in the case of the manufacturing company innovation. Therefore, inventory destroys a supplier’s and manufacturer’s ability to take products to market fast and thereby to feed forward evidence of product quality improvements to market. Similarly, inventory destroys a company’s ability to receive rapid, evidence-based quality feedback from the market. Inventory may create the impression that the company, by having the product readily available, is more responsive to the market. However, all inventory does is give the company the ability to make an immediate sale for an old product. It does so at the greater competitive cost of slow quality responsiveness for new, improved products, which is far more destructive of a company’s long term competitive position than the fact of losing an immediate sale on an old product.

Model/Design Change Responsiveness

In an increasingly large number of industries, model change is the norm. Television sets, household appliances, automobiles, motorcycles, computers, home furnishings, fashionable and trendy clothing, sports equipment and travel accessories are examples of industries where sales are driven by the rapid introduction of new models. Contrary to widely accepted notions, inventory offers no help to a company competing on model/design change responsiveness and is a destroyer of a company’s capability to deliver new designs to market fast and efficiently. The processes that first spot the emergence of new model/design requirements and the market trends that give rise to them are driven by the company’s market intelligence and capacity to stay close to and listen to customers, and have nothing to do with the existence or non-existence of inventory. Once the need for new models/designs has been established, the processes that produce design/model concepts quickly and manufacture working models and test their feasibility within a compressed time frame have to do with the collective competence lodged in the company’s design team, the small scale manufacturing capability to produce working models that can be refined by subsequent design efforts. These competencies have nothing to do with inventory. Once a design/model has been finalized and accepted for taking to market, how fast it actually reaches the market has very little to do with the level of inventory held by a company. Instead, the key capabilities for doing so include the existence of reliable, responsive suppliers that are committed to supporting the company’s innovation efforts, the rapid establishment of Bills of Material and routings for building the product, vigorous planning of the tooling/retooling effort, a state-of-the-art production and materials planning system that accurately estimates material requirements, a procurement system that assures tight coordination of the supplier network to assure timely delivery of requisite materials, a robust production planning and scheduling system that puts the resources in place to execute the sales plan, a well-honed execution system that executes plans and schedules reliably and efficiently, and a free flow of information between marketing, procurement, production and logistics that assures seamless coordination of the supply chain processes that deliver the product to market. These are not inventory issues, but planning and coordinating issues. Taking new models/designs to market quickly and efficiently requires tight coupling of the supply chain processes. The level of inventory in a supply chain seeks to decouple its successive parts so that each operation can be performed independent of the others. Because inventory consumes both time and money, a decoupled supply chain is also a slow, unresponsive and high cost one. A quick response supply chain is designed and operated to use vigorous planning and tight coordination to deliver customer requirements in timely fashion at minimum cost. Managing the supply and demand activities of a company as a supply chain instead of a set supply/inventory bin means undertaking supply actions in timely fashion so that the delivery of supply operates in near-perfect synchronicity with the realization of demand.

The largest part of the inventory held by companies derives from the failure to manage responsiveness in the supply chain so as to achieve that high level of supply and demand seamless synchronization that is demanded by today’s customers and competitive conditions. That synchronization invariably means driving inventory out of the supply chain processes and replacing it with information. It is much more costly to hold and manage inventory than it is to generate and manage information, so reliance on inventory as opposed to information-driven coordination and coupling to achieve high levels of product availability is a very costly strategy that destroys much of the responsiveness that it purportedly creates. In any area of supply chain performance, quick response means coordination and coupling. Inventory is decoupling, failure to coordinate slow, costly response. It is not a surprise that companies that are on the verge of bankruptcy are always awash in inventory. The reason is simple enough; the demand for their products plummeted and they failed through coordinated supply chain actions
to drive production and/or buying down to stay abreast of their lower demand levels. Lack of responsiveness always shows up as excess inventory.

Net Conversion Cycle (NCC) Responsiveness

In a real sense, the overall responsiveness of a company measures how fast it can convert market opportunities into profits and cash that can be partly returned to shareholders by way of a dividend payout and partly redeployed in the business in pursuit of more growth opportunities. Once a product/service has been successfully taken to market, a large part of that overall responsiveness is the speed with which the business can generate cash from normal, ongoing operations, what is referred to as the Net Conversion Cycle. Mathematically, the net conversion cycle in given by NCC, where,

\[ \text{NCC} = DI + DR - DP \]  \hspace{1cm} (7)

and,

\[ \begin{align*} 
DI &= \text{Days inventory} \\
DR &= \text{Days receivables} \\
DP &= \text{Days payables} 
\end{align*} \]

There is good evidence that the NCC is a major barometer of the overall market responsiveness of a company and its ability to generate cash to sustain its strategic development. Examples of benchmark supply chain management companies that also have very favorable NCCs include Amazon.com, Walmart and Dell. The evidence is also very convincing that inventory is the single largest contributor to poor NCC responsiveness. IBM's NCC is far inferior to Dell's, and the inventory turns of the former is also much lower than that of the latter. Walmart's NCC is far superior to Kmart's and the latter's inventory turns are also much lower.

Figure 4 further reinforces the idea that inventory turns are the key factor that drives the difference in the relative NCC responsiveness of companies. Days payables are usually driven by competitive forces in the supplier market, in particular, the bargaining power of suppliers relative to their customers. Thus Walmart, because of its huge size, has a very favorable bargaining power position and is quite successful at leveraging it to obtain very favorable financing terms from its suppliers. The receivables turnover is usually driven by the nature of the business and the competitive forces in the company's market. In most retailing, business is conducted on the basis of cash or credit card financing, which means that days-receivables are financed for at most one day. Inventory turnover, on the other hand, is the only one of the NCC drivers that depends almost entirely on management actions and is directly under the control of management of the supply chain. While a company usually has limited influence on the level of days payables and receivables and is severely constrained by what the competitive forces dictate, inventory turnover and the commensurate days inventory constitute the one driver that can be used to mitigate the unfavorable effects of supplier and end market forces on days payables and days receivables. A company that wants to create a favorable NCC responsiveness position relative to its competitors is pretty much constrained to do so by building a superior inventory turns advantage, until ideally, the NCC becomes negative. In consequence, companies that do a superior job of designing a supply chain process and planning, coordinating, executing and controlling supply chain activities in such a way as to achieve inventory turns that are superior to those of their competitors will almost always achieve NCC responsiveness superiority. Clearly then, far from having a positive impact on net conversion cycle responsiveness, inventory is its Achilles' heel, the one factor that, if not mitigated, will destroy a company's NCC responsiveness, a key barometer of competitive and financial superiority.

Conclusion and Implications for Supply Chain Management Research

The fact that the impact of inventory on supply chain responsiveness is largely negative makes the view widely held in theory and practice, the analysis undertaken here reveals that inventory is an important net destroyer of supply chain responsiveness, negatively affecting all of the dimensions of supply chain responsiveness except predictable upward shift in demand, unpredictable upward shift in demand and customer lead time compression. However, using inventory to provide high responsiveness when customers compress their lead time may not be the best strategy since the largest part of lead time comes from queue time that comes from large lot production. In that case, inventory masks the true unresponsiveness of an inefficient manufacturing planning, control and execution system. In most cases, instead of increasing speed of supply chain response, inventory represents dead weight that slows down the response process. In consequence, a firm
can only validly rely on inventory as a driver of supply chain responsiveness if that firm competes in market segments that are experiencing large predictable and unpredictable upward shifts in demand and customer lead time compression that is temporary. These upward demand shifts can only exist on a sustained basis at the beginning of the growth cycle for unusually cyclical industries, or for industries that are starting the growth phase of the product life cycle. In both these cases, the reliance on inventory as a driver of supply chain responsiveness is temporary and will, in fact, expose a company to severe cost and obsolescence penalties once the factors that are powering high increases in demand will have dissipated. A company that is using inventory as a means of increasing supply chain responsiveness must be very careful to monitor the pattern of demand growth to spot when the trend is changing and thereby deploy massive effort to reduce inventories quickly. Otherwise, inventory becomes a true double-edged sword that destroys responsiveness and reduces rather than enhances competitive advantage.

The issue of quality feedback and feedforward responsiveness has received very little attention in the literature, but it may prove to be critical in the future as companies are increasingly required to compete on quality. The insights developed here may open up a new avenue of research in supply chain management relative to the quality responsiveness impact of traditional supply chain decisions. Our results support the notion that supply chain decisions have real, direct and lasting quality impact and must be factored into a company's overall quality strategy and competitive strategy, particularly the differentiation subset of generic strategy and time-based strategies (Porter, 1980; Porter, 1985; Porter, 1986; Schmenner, 1988; Merrills, 1989; 34, 1990; Austin et al, 1991; Magretta, 1998). Our analysis leads to the anticipation that the effectiveness of differentiation strategies that are anchored on quality, variety, compressed time-to-market, quick model changeovers and rapid delivery of economical, small lots are compromised or at least significantly blunted by high levels of inventory. Although some of the conclusions that we have drawn on the impact of inventory on supply chain responsiveness challenge many of the notions and relationships currently accepted in the supply chain management literature, the extensive anecdotal and analytical evidence we have presented, buttressed by some insight gleaned from existing frameworks, lead us to conclude that there is much need for further research on the subject. We admit, however, that the results presented here are preliminary in nature and more definitive and stronger conclusions will depend on future empirical research. Future research may very well conclude that as competition intensifies and companies recognize the enormous negative impact of inventory on quality, response time and obsolescence cost, management is pushed to develop innovative ways to structure, configure and more tightly coordinate their supply chain systems while simultaneously expanding and deepening the deployment of existing, well-proven approaches like JIT/Pull systems, Channel Coordination, Vendor-Managed Inventory and Postponement Strategies. We conjecture that these innovations are changing the nature and fundamental understanding of the role of inventory in supply chain responsiveness.

References


