Creating the Customer-responsive Supply Chain:

A Reconciliation of Concepts

Andreas Reichhart and Matthias Holweg

Judge Business School
University of Cambridge
Cambridge CB2 1AG
United Kingdom
Tel.: +44-1223-760 583 (Matthias Holweg)
Email: m.holweg@jbs.cam.ac.uk

Forthcoming in The International Journal of Operations & Production Management
Structured Abstract

Purpose

While the concept of supply chain responsiveness (SCR) has received considerable attention in the operations management literature, mostly under the auspices of concepts such as build-to-order, mass customisation, lean and agility, so far we lack a comprehensive definition of SCR, as well as a defined relationship between ‘responsiveness’ and ‘flexibility’. Also, the frameworks at hand tend to consider only a subset of factors previously identified in the literature, and thus do not comprehensively portray the cause-and-effect relationships involved. In this paper we aim to address these gaps.

Design / methodology / approach

The paper synthesises the existing contributions to manufacturing and supply chain flexibility and responsiveness, and draws upon various related bodies of literature that affect a supply chain’s responsiveness such as the discussion of product architecture and modularisation.

Findings

We have identified four types of responsiveness: product, volume, mix and delivery, all of which can relate to different time horizons, and can be present as either potential or demonstrated responsiveness. We argue that a supply chain can feature different levels of responsiveness at different tiers, depending on the configuration of the individual nodes, as well as the integration thereof. Furthermore, we propose a holistic framework distinguishing between requiring and enabling factors for responsiveness, identifying the key relationships within and between these two categories.

Originality / value

The paper proposes a definition of four types of responsiveness which will support further empirical studies into the concept and its application. Furthermore, a holistic framework is developed that allows for cause-and-effect relationships to be investigated and dependencies to be identified.

Keywords: Supply chain management, responsiveness, flexibility, agility, build-to-order
Creating the Customer-responsive Supply Chain:
A Reconciliation of Concepts

1 Introduction

In market conditions of increasing levels of product variety and customisation, the ability to respond to customer orders in a timely fashion can provide a critical competitive advantage. Across industry sectors, such as fashion (Christopher, 2000; Storey et al., 2005), personal computers (Kapuscinski et al., 2004), consumer electronics (Catalan & Kotzab, 2003), construction (Arbulu et al., 2003), and automobiles (Holweg & Pil, 2004), companies are contemplating strategies to increase their responsiveness to customer needs by offering a high product variety with short lead-times. More recently, the discussion of mass customised products (Lampel & Mintzberg, 1996; Gilmore & Pine II, 1997), has shifted the discussion beyond the simple provision of product variety towards individually customised products. While these customer-driven or build-to-order (BTO) strategies have been implemented in the personal computer sector, with Dell being the most prominent example (see Kapuscinski et al., 2004), complex manufacturing operations, such as automotive, have been slower in adapting these strategies (Hertz et al., 2001; Holweg & Pil, 2004).

The increasing importance of BTO supply chains results from two developments: first, the number of product variants has been increasing across most industries, such as consumer electronics (Catalan & Kotzab, 2003), fashion and sportswear (Fisher et al., 1994), and automobiles (Holweg & Pil, 2004). Second, time has become a factor in competitiveness as customers are increasingly reluctant to accept long lead-times for products and services (Bower & Hout, 1988; Stalk, 1988). The former development creates severe operational problems for traditional make-to-forecast or ‘push’ strategies, as firms require large amounts of finished goods inventories to ensure customers find the specifications they are looking for. In the case of the Mercedes E-Class which is available in more than three septillion \(3 \times 10^{24}\) variations (Pil & Holweg, 2004) for example, a make-to-forecast strategy becomes virtually impossible. While BTO strategies can help overcome the first hurdle, existing auto supply chains are not sufficiently responsive to deal with the second development; impatient customers, who would like their vehicles delivered within 2-3 weeks, rather than the current six weeks (Holweg et al., 2005a).
Across industry sectors, the concept of responsiveness has been receiving increasing attention in the operations management literature, and has advanced as one of the key themes in recent supply chain research. However, we regard supply chain responsiveness not as an operations paradigm in its own right, but rather as a concept that can implicitly rest at the core of various operations strategies, such as lean thinking (Womack & Jones, 1996, as described by Hines et al. 2004), agility (Goldman & Nagel, 1993) and more recently, build-to-order supply chain management (Gunasekaran & Ngai, 2005). Common to all of these is the importance of customer-oriented pull systems, as opposed to traditional forecast-based mass production systems (Skinner, 1969; Hill & Chambers, 1991; Fisher, 1997). Responsiveness thus is a crucial aspect of build-to-order supply chains, yet not one that is confined to them, as the fashion industry with some highly responsive non build-to-order supply chains demonstrates (see for example Christopher, 2000).

We argue that the importance of responsiveness in today’s industry settings, in conjunction with a wealth of contributions that have blurred the boundaries to related concepts, such as flexibility (Slack, 1987; Upton, 1994), agility (Goldman & Nagel, 1993; van Hoek et al., 2001; Yusuf et al., 2004), and lean thinking (Womack & Jones, 1996; Hines et al., 2004), justifies a review of this body of literature, and the subsequent attempt to propose a set of clear definitions. In particular, a key shortcoming of the existing definitions of responsiveness is their unclear separation from the definitions of flexibility. Furthermore, supply chain responsiveness is a goal that can be achieved through multiple means, the appropriateness of which largely depends on various product and market related characteristics. A framework that provides a holistic approach to responsiveness, explaining the interdependencies within and between its external requirements and internal determinants, is still missing. The main objectives of this paper thus are:

(i) to propose a clear definition of supply chain responsiveness and its relationship to flexibility, and

(ii) to develop a holistic framework, capturing the individual factors that require and enable the responsiveness of a supply chain system.

The first section of the paper will define responsiveness and its relation to the concept of flexibility drawing mainly on existing definitions of flexibility and responsiveness, both in manufacturing systems and supply chains, and supported by a systems’ terminology (cf. Ackoff, 1971). The second section will review the literature on the requirements and determinants of responsiveness towards the proposition and explanation of a conceptual framework of supply
chain responsiveness. Finally, conclusions, managerial implications and areas for future research will be presented.

2 Towards a Generic Definition of Responsiveness

There is considerable ambiguity in the existing literature with regards to the differences between responsiveness and flexibility. Furthermore, both terms are often used to describe features of manufacturing systems (Slack, 1987; Gindy et al., 1999) and entire supply chains (Handfield & Bechtel, 2002; Lummus et al., 2003), and are at times used interchangeably, yet do not capture exactly the same concept. In addition, within the last decade two new terms have been introduced, ‘agility’ and ‘leagility’ (Goldman & Nagel, 1993; Naylor et al., 1999), which postulate flexibility and responsiveness in manufacturing operations, organisations and supply chains as a key tenet of a firm’s competitiveness.

Historically, the first contributions on flexibility focussed entirely on manufacturing systems, and were only later extended to entire supply chains. Responsiveness on the other hand is a term that has only more recently been introduced as a distinct, independent concept in the operations literature. Table 1 lists the key contributors to the flexibility and responsiveness streams and their respective scope (manufacturing system vs. supply chain), which will be discussed in detail below.

<table>
<thead>
<tr>
<th>Flexibility – responsiveness – agility: contributors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Manufacturing System Level</strong></td>
</tr>
<tr>
<td>Zelenović, 1982</td>
</tr>
<tr>
<td>Lim, 1987</td>
</tr>
<tr>
<td>Slack, 1987</td>
</tr>
<tr>
<td>Gerwin, 1993</td>
</tr>
<tr>
<td>Gupta, 1993</td>
</tr>
<tr>
<td>Upton, 1994</td>
</tr>
<tr>
<td>Lau, 1999</td>
</tr>
<tr>
<td>Vokurka &amp; O'Leary-Kelly, 2000</td>
</tr>
<tr>
<td>Gindy et al., 1999</td>
</tr>
<tr>
<td>Matson &amp; MacFarlane, 1999</td>
</tr>
<tr>
<td>Mileham et al., 1999</td>
</tr>
<tr>
<td>Holweg, 2005a</td>
</tr>
<tr>
<td>Holweg, 2005b</td>
</tr>
<tr>
<td>Goldman &amp; Nagel, 1993</td>
</tr>
<tr>
<td>Burgess, 1994</td>
</tr>
<tr>
<td><strong>Supply Chain Level</strong></td>
</tr>
<tr>
<td>Vickery et al., 1999</td>
</tr>
<tr>
<td>Garavelli, 2003</td>
</tr>
<tr>
<td>Lummus et al., 2003</td>
</tr>
<tr>
<td>Wadhwa &amp; Rao, 2004</td>
</tr>
<tr>
<td>Lau &amp; Lee, 2000</td>
</tr>
<tr>
<td>Handfield &amp; Bechtel, 2002</td>
</tr>
<tr>
<td>Catalan &amp; Kotzab, 2003</td>
</tr>
<tr>
<td>Harrison &amp; Godsell, 2003</td>
</tr>
<tr>
<td>Randall et al., 2003</td>
</tr>
<tr>
<td>Naylor et al., 1999</td>
</tr>
<tr>
<td>Christopher, 2000</td>
</tr>
<tr>
<td>Mason-Jones et al., 2000</td>
</tr>
<tr>
<td>Herer et al., 2002</td>
</tr>
<tr>
<td>Towill &amp; Christopher, 2002</td>
</tr>
<tr>
<td>Yusuf et al., 2004</td>
</tr>
</tbody>
</table>
In this section, we aim to consolidate these concepts into a generic definition of responsiveness, and to define its relationship to flexibility. The literature on the definition of flexibility and its classification is reviewed, initially focussing on manufacturing systems and then extending it to include entire supply chains. Subsequently, the concept of responsiveness in manufacturing systems and supply chains is discussed, reviewing existing definitions. Finally, a generic definition including the relationship between flexibility and responsiveness is proposed and its implications and assumptions are discussed.

2.1 Flexibility in Manufacturing and Supply Chain Systems

The contributions on flexibility in manufacturing systems are numerous (Gerwin, 1993; Upton, 1994; Lau, 1999) and difficult to summarise (de Toni & Tonchia, 1998). In their literature review of manufacturing flexibility, de Toni & Tonchia (1998) propose a scheme for analysing this vast body of knowledge according to six aspects, only two of which are relevant for this part of the paper: (i) the definition of flexibility and (ii) the classification of flexibility. The latter is similar to what Garavelli (2003) calls the ‘object of change’, which according to him, is the most interesting aspect of the flexibility discussion from an operational perspective. It is often conceptualised based on the seminal contributions by Slack (1987) and Upton (1994), who found several types of flexibility in manufacturing systems along two and three dimensions, respectively\(^1\). Slack identifies four types of flexibility: product, mix, volume and delivery. He also identifies two dimensions for each type of flexibility: range and response. Range refers to the maximum number of different outcomes a resource with the respective flexibility type can achieve, such as the total number of different products a given machine can produce. ‘Response’ refers to the time and cost with which different values within a range can be achieved (e.g. setup time and cost for switching between two products). Slack also concludes that three types of manufacturing resources can be utilised to achieve flexibility, namely flexible technology, labour and infrastructure.

Upton (1994) is less specific with regards to the types of flexibility that exist. He identifies up to 15 different types, out of which he only exemplifies four in the case studies discussed:

\(^1\) It should be noted that Upton and Slack did not use the same terminology: Upton refers to Slack’s ‘types of flexibility’ as ‘dimensions’, whereas he refers to Slack’s ‘dimensions’ as ‘elements of flexibility’. In addition, Upton refers to Slack’s dimension of ‘response’ as ‘mobility’.
product, volume, business area and size. These 15 types are also used by Vokurka and O'Leary-Kelly (2000), while other authors report varying types and numbers (see de Toni & Tonchia, 1998 for a comprehensive review). Expanding on Slack’s (1987) framework, Upton identifies a third dimension of flexibility: uniformity, which refers to the ability of a resource to provide consistent performance throughout its entire range. In addition, Upton draws attention to a time aspect in the flexibility discussion by saying that firms can be flexible at the operational, tactical and strategic level, each one focusing on different time horizons, thereby confirming previous arguments made by Carlsson (1989) and Zelenović (1982). Upton further concludes that there are two different aspects of flexibility which he refers to as 'internal' vs. 'external' flexibility. External flexibility can be linked to achieving a competitive advantage, such as speed of delivery (‘what the customer sees’). Internal flexibility on the other hand is the internal means by which external flexibility can be achieved ('what can we do').

Drawing on Ackoff (1971) and previous definitions of flexibility (e.g. Zelenović, 1982), flexibility can thus be defined as the ability of any system to adapt to internal or external influences, thereby acting or responding to achieve a desired outcome. Following Slack (1987) and Upton (1994), a system’s flexibility is based on internal resources that can be used to achieve different types of internal flexibility, which in turn can support the system’s ability to demonstrate external flexibility to its environment. External flexibility can focus on short-term, medium-term or long-term goals and will hence require operational, tactical or strategic internal flexibility to achieve them.

The above definition based on flexibility in manufacturing systems can also be extended to entire supply chains, in the same way that traditional production and inventory control research has been extended from single to multi-tier manufacturing systems. Vickery et al. (1999) and Lummus et al. (2003) provide the two most prominent examples of this extension. Vickery et al. identify five distinct flexibilities for a supply chain which include product, volume, new product, distribution and responsiveness flexibility. As described by Lummus et al., the extension from manufacturing to supply chain flexibility follows the logical extension of a manufacturing system to a complete supply chain: ‘(..) this extension involves looking at those components that make an organisation flexible and extends them beyond the organization’s boundaries to other nodes in the supply chain’ (Lummus et al., 2003, p. 3). Hence the definition of flexibility does not change per se, only its internal types may be modified to include inter-company considerations, now that an entire supply chain is the unit of analysis, as opposed to a single node (manufacturing system) within a supply chain.
2.2 Responsiveness in Manufacturing and Supply Chain Systems

Matson and MacFarlane (1999, p. 765) define production responsiveness as ‘(...) the ability of a production system to achieve its operational goals in the presence of supplier, internal and customer disturbances’. Those disturbances clearly relate to the three types of uncertainty explained by Davis (1993), namely supply, process and demand uncertainty. McCutcheon et al. (1994) regard responsiveness as equal to the delivery lead-time for a certain product. Therefore, a manufacturing system would be regarded as more responsive, if it could deliver the same product with a shorter lead-time. Holweg (2005a, p. 6) uses a more general, yet very similar definition and argues that responsiveness is ‘(...) the ability of the manufacturing system or organisation to adapt to changes and requests in the marketplace’. A similar definition is also used by Hines (1998), who refers to the responsiveness of the supply chains studied using the supply-chain response matrix (see Hines & Rich, 1997). This approach considers the total inventory held at all stages in a supply chain in combinations with the throughput time at each stage. Hines (1998) thus suggests that those supply chains are more responsive that can switch to a new product within a shorter time period, because they can a), push new products faster through the entire supply chain and b), keep less stock of the ‘old’ product that needs to be used first. Catalan and Kotzab (2003, p. 677) define responsiveness of a supply chain in a comparable way, namely ‘(...) as the ability to respond and adapt time-effectively based on the ability to ‘read’ and understand actual market signals’.

The above definitions, despite their differences, show considerable similarities. First, the majority of them link responsiveness to changes or outcomes required by the system’s external environment. Second, they usually include some time or effort dimension, i.e. those systems that are regarded as more responsive can adapt to these changes faster than others, a concept that matches the ‘response’ dimension of flexibility discussed above. However, Matson and MacFarlane’s (1999) definition is not fully consistent with the other definitions in as far as they link responsiveness to internal as well as external requirements (‘disturbances’). Thus, the question remains whether a system that can react quickly to internal disturbances, such as machine break downs (for manufacturing systems) or supplier problems (for supply chains), should be considered responsive or not? The majority of authors seem to link responsiveness exclusively to external events (e.g. changes in customer demand), which is supported by Ackoff (1971, p. 664), who defines the response of a system as ‘(...) a system event for which another event that occurs to the same system or to its environment is necessary but not sufficient; that
is, a system event produced by another system or environmental event (the stimulus). A system can therefore only respond to external stimuli but not to internal events alone. Responsiveness should thus be considered as a concept that is solely customer focused, and its measurability depends on where the system boundaries are drawn and thereby on the definition of the system’s customers.

As discussed above, there is no general agreement on the number of internal flexibility types in manufacturing systems, let alone in supply chains. Hence it is proposed that a system’s external flexibility, i.e. the flexibility that a customer might be ‘interested in’, consists of the four types identified by Slack (1987): product, mix, volume and delivery. Product flexibility describes the ability to introduce new products or changes to existing products. Mix flexibility is the ability to alter the product mix (within the existing product range) that the system delivers. Volume flexibility refers to the ability to change the system’s aggregated output and delivery flexibility is the ability to alter agreed delivery agreements (e.g. shortening lead-times or even changing the products’ destination). If there are in-sequence delivery arrangements, such as in the automotive component industry, delivery flexibility also includes the ability to make changes to the agreed delivery sequence.

![Figure 1: Relationship between flexibility and responsiveness](image-url)

**Figure 1: Relationship between flexibility and responsiveness**

2 The triggering of a system event without any external stimuli is referred to as an ‘act’, not a ‘response’ (Ackoff, 1971).
Figure 1 shows the relationship between flexibility and responsiveness based on the flexibility dimensions identified by Slack (1987) and extended by Upton (1994), and incorporating the four external and various internal flexibility types discussed above. Responsiveness is therefore constituted by the system’s response dimension of its external flexibility types. For the purpose of this and further studies, the following definition will be adopted:

The responsiveness of a manufacturing or supply chain system is defined by the speed with which the system can adjust its output within the available range of the four external flexibility types: product, mix, volume and delivery, in response to an external stimulus, e.g. a customer order.

The types of internal flexibility required on the other hand will be contingent upon the types of responsiveness demanded as well as upon the specific operational setting. Most importantly it needs to be acknowledged that such relationships exist, i.e. mix responsiveness might for example require flexibility in machinery. A customer would however not be interested in how his supplier is able to meet changes in product mix demands, as long as the supplier can adjust the product mix to suit the customers’ needs without negatively affecting other requirements, such as product quality. Similarly, a customer would not be interested in whether his supplier is sufficiently responsive to meet a large unforeseen order because the supplier itself is responsive (e.g. through buffer stocks of purchased components) or because the supplier’s suppliers can deliver the required components quickly. From a customer’s point of view his supplier’s manufacturing system’s responsiveness is thus the same as the supply chain’s responsiveness. Hence a shift in the focus from manufacturing systems to supply chains is driven by other factors. Traditionally, supplier flexibility was seen as an operational factor impacting upon a manufacturing system’s performance (e.g. Vokurka & O’Leary-Kelly, 2000), yet with the increasing complexity in today’s supply chains it is argued that the flexibility of manufacturing systems in a supply chain should be regarded as a factor contributing to a supply chain’s responsiveness and not vice versa, at least for firms, which want to compete on the basis of BTO supply chains.

It is also important to note that responsiveness will differ at different nodes in the system. The raw material supply node in some industries for example, is known for its unresponsiveness (in terms of large batch sizes and long lead-times), whereas tier-one suppliers are often required to be highly responsive (Holweg, 2005a). Thus a supply chain’s responsiveness can increase or decrease as one moves downstream in the supply network, similar to the way that quality levels
can increase or decrease within supply chains (see Hines, 1998). At every node in a supply chain system, the supply chain also has a potential (‘what could it do’) and a demonstrated responsiveness (‘what does it do’, see Upton, 1994). Only when these are aligned, can a responsive supply chain be created that is also cost-efficient.

Looking at the time classification of flexibility, responsiveness can be split into at least short-term and medium-term responsiveness, too. A supply chain’s short-term or operational responsiveness is its ability to adjust its output to short-term demand changes. These changes can be due to changes in the product mix (mix responsiveness), the volumes required (volume responsiveness), or the delivery sequence or timing (delivery responsiveness). New products are rarely introduced on short notice, and product responsiveness probably only exists on the medium- and long-term horizon. What exactly short-term means however depends on the industry and supply chain node under consideration; in fast clockspeed settings such as electronics (where products can have a life cycle of as low as 2 months) it will be very different from slow clockspeed settings, like automotive, where products stay in production for an average of six years (Holweg & Pil, 2004).
3 Towards a Framework for Responsiveness

3.1 Approach to Literature Review

In the following, a framework for the study of supply chain responsiveness will be developed based on a review of a broad range of relevant literature, mainly from the operations management stream. The framework serves multiple objectives: first and most importantly, the aim is to consolidate the contributions on previous frameworks (see Table 2) and individual factors into a single comprehensive framework. Second, the framework should be generic in nature in order to serve as a basis for further holistic studies for researchers from all perspectives of the epistemological spectrum. This in turn requires two-sided measurements (e.g. dependent vs. independent variables), as opposed to a mere index creation, used for example by van Hoek et al. (2001). Such a two-sided measurement is further essential for supporting the criterion-related validity (cf. Flynn et al., 1990) of any supply chain responsiveness index made up of factors assumed to increase a supply chain’s responsiveness. Last, the interdependencies within and between the external requirements and internal determinants of supply chain responsiveness need to be clearly stated, as these cause-and-effect relationships are central for building responsive supply chains.

This section will start by summarising the key contributions available on supply chain responsiveness and flexibility, categorised by their main characteristics and the respective factors considered (see Table 2). Subsequently, the individual factors related to supply chain responsiveness will be discussed. Last, these factors will be consolidated into a holistic framework of supply chain responsiveness.

Table 2: Frameworks for analysing supply chain responsiveness / flexibility

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Characteristics of Framework</th>
<th>Factors Considered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kritchanchai &amp; MacCarthy, 1999</td>
<td>• Framework for comparing cross-industry responsiveness of the order fulfilment process</td>
<td>• Nature of product</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Demand</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Major impact stimuli</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Awareness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Capabilities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Goals</td>
</tr>
<tr>
<td>van Hoek et al., 2001</td>
<td>• Framework measuring a supply chain’s ‘agile capabilities’ based on five dimensions of agility</td>
<td>• Customer sensitivity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Virtual integration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Process integration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Network integration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Measurement</td>
</tr>
<tr>
<td>Catalan &amp;</td>
<td>• Responsiveness index (‘rating’)</td>
<td>• Lead time (production &amp; distribution lead-time)</td>
</tr>
</tbody>
</table>

12
Kotzab, 2003  
- based on four components grouped into two broad categories: time effective flow of goods and information & demand transparency  
- Postponement strategies  
- Bullwhip effect  
- Information exchange

Lummus et al., 2003  
- Analysis split into components and outcomes of supply chain flexibility  
- Five components and two outcomes  

<table>
<thead>
<tr>
<th>Components</th>
<th>Outcomes</th>
</tr>
</thead>
</table>
| - Operations systems  
- Logistics processes  
- Supply network  
- Organisational design  
- Information systems | - Customer satisfaction (including service and responsiveness)  
- Improved supply chain asset utilisation |

Holweg, 2005b  
- Three dimensions of responsiveness (volume, product, process)  

| Customer lead-times  
| Volume stability  
| Demand specifications (Pareto)  
| Product variety (external, internal)  
| Point of customisation  
| Product life cycle  
| Total order-to-delivery (OTD) time  
| Distribution lead-time  
| Supply chain response lead-time  
| Decoupling points |

At prima facie, there is a considerable overlap in the frameworks summarised in Table 2, yet there is neither agreement on which individual factors and concepts to include nor on how to group them. A major shortcoming of the majority of the existing frameworks is the lack of distinction between factors that require supply chains to be responsive and factors that enable them to be responsive (cf. Kritchanchai & MacCarthy, 1999). The subsequent literature review and framework development will apply this distinction, and cover ‘external requirements’, i.e. factors that require responsiveness, followed by ‘internal determinants’, i.e. factors that enable responsiveness. This grouping facilitates a structured analysis of the concept of supply chain responsiveness in at least two ways: first, it illustrates the underlying cause-and-effect relationships that exist in supply chains. For example, it is not the supply chain’s responsiveness that changes by removing external requirements, such as external demand variability, but rather the need for the supply chain to be responsive, a fact that has in the past not always been considered (see for example Harrison, 1996). Second, it can guide either customers of a supply chain or firms therein to identify ways to influence the supply chain’s suitability for its environment. This will further facilitate matching different supply chains to their environment as suggested by Fisher (1997).
The literature review will be structured according to the individual factors associated with requiring or enabling responsive supply chains. Such a structure appears appropriate, since the majority of non-conceptual contributions investigate individual factors as opposed to the holistic concept of supply chain responsiveness. The factors discussed below mainly originate from the frameworks presented in Table 2, yet the individual sections also draw extensively upon the large body of literature that focuses on those factors individually.

3.2 External Requirements

The need for supply chains to become responsive can be derived from the contributions listed in Table 2, even though they are rarely included in the frameworks. The four main areas are: i) demand uncertainty, ii) demand variability, iii) product variety, and iv) lead-time compression. Table 3 gives an overview over these factors, which also provide the structure for the following discussion of the external requirements.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Main Links to Supply Chain Responsiveness</th>
<th>Key Contributions</th>
</tr>
</thead>
</table>
| Demand uncertainty      | • Main requirement for being responsive, i.e. 100% reliable demand would considerably reduce need for responsiveness  
                          | • Important sub-category is schedule instability, particularly important for industries operating under rolling schedules | • Davis, 1993  
                          |                                 |                                                                               | • Fisher et al., 1994  
                          |                                 |                                                                               | • Harrison, 1996  
                          |                                 |                                                                               | • Griffiths & Margetts, 2000  
                          |                                 |                                                                               | • Krajewski et al., 2005  |
| Demand variability      | • Often closely linked to demand uncertainty, yet conceptually different  
                          | • Even if demand was 100% reliable, large swings (even if known) in demand could still require responsiveness | • Harrison, 1996  
                          |                                 |                                                                               | • Inman & Gonzalves, 1997  |
| Product variety         | • Product variety further increases demand uncertainty  
                          | • Product variety can directly increase the need for mix responsiveness  
                          | • High product variety increases the cost of using finished good inventories to fill orders | • McCutcheon et al., 1994  
                          |                                 |                                                                               | • MacDuffie et al., 1996  
                          |                                 |                                                                               | • Berry & Cooper, 1999  
                          |                                 |                                                                               | • Randall & Ulrich, 2001  |
| Lead-time compression   | • Directly increases need for responsiveness, as less time is available to respond to customer orders  
                          | • Indirectly increases need for responsiveness through increased demand uncertainty (changes in P:D ratio) | • Bower & Hout, 1988  
                          |                                 |                                                                               | • Mather, 1988  
                          |                                 |                                                                               | • Stalk, 1988  
                          |                                 |                                                                               | • McCutcheon et al., 1994  |

3 The contributions listed specifically link the respective factor to supply chain responsiveness.
Demand Uncertainty and Variability

A range of previous studies identify uncertainty as the main reason for being responsive (Davis, 1993; Fisher et al., 1994; Randall et al., 2003). Under conditions of reliable information about demand conditions, there would hardly be a need to be responsive. The need arises mainly from the uncertainty that stems from volume and or product mix changes in the customer demand signal. Many studies have investigated the factors that make responsiveness or flexibility a necessary feature of either individual manufacturing systems (Azzone & Masella, 1991; Matson & MacFarlane, 1999) or entire supply chains (Fisher et al., 1994; Christopher, 2000) and therein uncertainty is always mentioned as the root cause for becoming flexible or responsive, either directly or indirectly. Uncertainty itself can emanate from three different sources, namely supply uncertainty, process uncertainty and demand uncertainty (Davis, 1993). Of these, demand uncertainty is commonly regarded as the most severe type (Davis, 1993; McCutcheon et al., 1994).

An interesting sub-category of demand uncertainty is demand uncertainty under rolling-horizon planning/scheduling (Simpson, 1999; Dellaert & Jeunet, 2003; Krajewski et al., 2005), which is the prevailing method in repetitive manufacturing environments. Short-term schedule changes or even daily sequence changes for in-sequence suppliers operating under rolling-horizon schedules can occur when the updated schedule does not match the previous one. The impact of schedule instability on supply chains is very significant and it has been discussed by numerous authors (e.g. Inman & Gonzalves, 1997; Griffiths & Margetts, 2000; Krajewski et al., 2005), who consistently regard it as a main cost driver in the supply chain. Oliver et al. (1994) even conclude from their investigations into the differences between ‘world-class’ and ‘non world-class’ automotive component plants that demand stability ranked amongst the most important factors for overall plant performance.

In addition to schedule instability (or demand uncertainty in general), schedule variability (or demand variability) is mentioned in relation to responsiveness (Harrison, 1996). Here, a clarification needs to be made between two schedule attributes: schedules can be stable by not deviating from the previous schedule or forecast, and can be level (i.e. not variable), whereby the day-to-day changes are kept small within predefined boundaries (a concept also referred to as ‘heijunka’, which literally translates into ‘smooth wave’). Inman & Gonzalves (1997) and Harrison (1996) in reference to Bhattacharya et al. (1995) argue for a difference between stable schedules and level schedules with regard to their impact on supply chains, yet the existing literature on schedule variations (both instability and variability) has often neglected to clearly differentiate between these two concepts (see for example Liker & Wu, 2000).
Both demand uncertainty and variability are inherent to most operations and can require different types of responsiveness depending on their nature (i.e. product mix vs. volume vs. delivery changes). Thus, depending on the type of demand uncertainty or variability, different internal capabilities might be required, at least on a manufacturing system level (Suarez et al., 1995). The impact of uncertainty can be observed in many supply chains. Hewlett-Packard for example, relied on high inventory levels to buffer against demand uncertainty in their supply chain until they realised that some product configuration decisions could be postponed by using a more modular product architecture (Davis, 1993; Feitzinger & Lee, 1997). While they could not reduce the demand uncertainty originating from end customers, they moved to a less costly strategy to deal with the problem. In a true build-to-order supply chain however, such as in the case of the Volvo Car Group (Hertz et al., 2001), investments in flexible manufacturing (e.g. through mutable support structures) are generally more suited than late configuration or postponement (Pil & Holweg, 2004).

Product Variety

One of the prevailing streams in the literature on the management of product variety is on its implications on the firm’s wider performance (Lancaster, 1990; da Silveira, 1998; Ramdas, 2003; Pil & Holweg, 2004). The need for managing product variety results both from the competitive importance of product variety in today’s markets (McCutcheon et al., 1994; Lampel & Mintzberg, 1996; Gilmore & Pine II, 1997), and its potential financial impact for product development activities and manufacturing operations. As Fisher (1994) and Randall & Ulrich (2001) explain, demand uncertainty is amplified by product variety, as the same aggregated demand is split over more SKUs (stock keeping units), leading to an increase in the aggregated errors associated with each forecast. In addition, product variety increases the need for mix responsiveness as the range of the external mix flexibility increases (Berry & Cooper, 1999), and customers are not willing to accept longer lead-times. Those problems have also led firms to rethink the level of product variety that is really demanded by their customers (Fisher et al., 1994; MacDuffie et al., 1996).

Holweg & Pil (2004) differentiate between three dimensions of product variety. First, external variety (also referred to as product proliferation by some authors: see Gupta & Srinivasan, 1998) refers to the number of SKUs (including their variations) available to a firm’s customers at any point in time. Second, internal variety refers to the complexity within a firm’s manufacturing processes and can be approximated by the number and variety of components
required for manufacturing a given product. Clearly, a high level of modularisation (Starr, 1965) helps limit the internal variety under a given external variety. Internal variety or rather product architecture should however be seen as an internal determinant of responsiveness and will be discussed later. Last, *dynamic variety* mainly refers to shortened product life cycles, i.e. it refers to the speed with which consumers will be given access to new products. Several authors (Davis, 1993; Fisher *et al.*, 1994) have argued that the demand for a product will always be harder to predict at the beginning of its life cycle, as no past demand pattern for this product is available, and because customers may react unexpectedly to the new product. Higher dynamic variety thus further increases demand uncertainty. Despite being primarily an external requirement, product variety can also directly inhibit supply chains from being responsive since it makes the use of finished goods buffer stocks more costly (Holweg & Pil, 2004).

In practice, the product variety that a company in a given industry wants to offer to its customers should determine the supply chain strategy (Fisher, 1997), but it is at the same time also restricted by its supply chain capabilities. Dell for example, a company known for its responsive assemble-to-order supply chain, is able to offer virtually any possible variation of a PC or laptop computer to its customers, because the computer is only assembled once the exact specifications are known (Holweg & Pil, 2004). Most of Dell’s competitors however, who sell made-to-forecast computers through retail outlets, are forced to limit product variety to a few variations. This means that a customer who would like to have a faster CPU might also have to take (and ultimately pay for) a larger hard drive, even though it might not be required

*Lead-Time Compression*

Time based competition (Bower & Hout, 1988; Stalk, 1988; McCutcheon *et al.*, 1994) by definition increase the need to be responsive, because the firm or supply chain is given less time to respond to new orders or changes in existing ones. Mather (1988) provides another explanation for why lead-time compression requires additional responsiveness. Using the P:D ratio, a concept dating back to work by Shingo (cf. Shingo, 1989), he explains how the forecasting horizon becomes longer if the customer lead-time ‘D’ decreases in relation to the production lead-time ‘P’. The longer the time horizon that needs to be forecasted however, the less reliable the forecast becomes (Mather, 1988; Randall & Ulrich, 2001), which in turn increases demand uncertainty.
3.3 Internal Determinants

The internal determinants identified during the literature search can be grouped into two broad categories. The first are factors that focus mainly on individual nodes within the supply chain (i.e. manufacturing systems). Secondly are factors that deal with the integration of supply chain partners. The former will be referred to as ‘operational factors’, whereas the latter are commonly called the dimensions or factors of ‘supply chain integration’ (Morash & Clinton, 1998; Lee, 2000). The existing literature on supply chain responsiveness (see Table 2) tends to focus on supply chain integration, even though it is also acknowledged that single node factors, such as manufacturing responsiveness, will contribute to the responsiveness of the overall supply chain (see for example Lummus et al., 2003). The identification and discussion of the operational factors thus mainly relies on a broad review of the operations management literature. While the discussion of supply chain integration also draws upon such a broad review, the concept itself has been discussed by various authors already (Morash & Clinton, 1998; Lee, 2000; Bask & Juga, 2001; van Hoek et al., 2001), yet rarely are the same terminology or grouping of sub-factors used. Van Hoeck et al. (2001) for example, use the terms ‘virtual integration’, to describe information sharing and the integration of information systems, ‘process integration’, to describe the alignment of processes to manage change and overcome internal uncertainties, and ‘network integration’ to describe the focus on common goals as opposed to single firm competition. Lummus et al. (2003) on the other hand use the categories ‘information systems’, ‘logistics processes’ and ‘supply networks’ to cover broadly the same content. For the following discussion, the grouping will be based on Lee (2000) and Bagchi & Skjoett-Larsen (2002). Thus supply chain integration will be split into ‘information integration’, ‘coordination and resource sharing’, ‘organisational integration’ and a fourth integration factor that emerged during the literature review: ‘spatial integration and logistics’. All factors identified are summarised in Table 4.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Main Links to Supply Chain Responsiveness</th>
<th>Key Contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demand anticipation</td>
<td>The accurate anticipation of demand can help supply chains to respond to customer requirements faster</td>
<td>Fisher et al., 1994</td>
</tr>
<tr>
<td></td>
<td></td>
<td>de Treville et al., 2004</td>
</tr>
</tbody>
</table>

The contributions listed specifically link the respective factor to supply chain responsiveness.
<table>
<thead>
<tr>
<th>Supply chain integration</th>
<th>Manufacturing flexibility</th>
<th>Inventory</th>
<th>Product architecture / postponement</th>
<th>Information integration</th>
<th>Coordination and resource sharing</th>
<th>Organisational integration</th>
<th>Spatial integration and logistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Traditional focus of the flexibility and responsiveness discussion • Can directly reduce the production lead-times and change-over times for products in supply chains</td>
<td>• Can both increase and decrease the responsiveness of supply chains • Often used as buffer against uncertainty • Closely linked to the decoupling point, which is a common criterion for classifying supply chain strategies, such as build-to-order supply chains</td>
<td>• Determines to a large extent where the decoupling point can be placed and thus how responsiveness can be achieved • Determines manufacturability and internal product variety / complexity</td>
<td>• Can help reduce internal demand amplification and eliminate delays due to slow information flows • Eliminating unnecessary demand uncertainty and variability facilitates a better focus on end customer demand</td>
<td>• Removes delays and unnecessary activities in supply chains and leverages synergies • Reduces demand variability and uncertainty</td>
<td>• Can increase the responsiveness and general performance of supply chains in various ways • Particularly important impact on trust, which is required for a variety of interactions between supply chain members</td>
<td>• A reduction in transport lead-times directly increases the responsiveness of supply chains • Can further strengthen process coordination and organisational integration</td>
</tr>
</tbody>
</table>

**Demand Anticipation**

One of the most obvious enablers for a supply chain to be responsive is for its members to be able to anticipate their actual demanded output. Fisher et al. (1994) suggest ‘accurate response’ systems to counter demand uncertainty by differentiating which of the products offered can be forecasted more accurately than others. For easy-to-forecast products, less demanding supply chain strategies can be used, saving flexible resources for hard-to-forecast products (Fisher, 1997). De Treville et al. (2004) investigate different demand anticipation and information exchange strategies and develop what they refer to as the three levels of relative supply lead-time (RSLT), a concept based on Mather’s P:D ratio (Mather, 1988). They conclude that improving demand anticipation is only part of the story and that the overall P:D ratio must be improved to become more responsive. This they link to required improvements in combined production lead-times.
Manufacturing Flexibility

Supply chain responsiveness by definition requires the individual manufacturing systems to be responsive (i.e. flexible on their external response dimension). In our literature review however, we did not find any previous study which differentiated the factors of manufacturing flexibility by their impact on any specific dimension. Nevertheless, a reduction in the system’s throughput time (the ‘P’ in Mather’s P:D ratio) is certainly a determinant of responsiveness as it allows for products to be delivered to customers without the need for market anticipation or costly inventories of finished goods (see also Hines, 1998). A decrease in the throughput time of the overall system can be achieved by various measures, such as a reduction in logistics lead-times (Davis, 1993), faster information processing (de Trelle et al., 2004) and factors directly linked to manufacturing flexibility like shortened machine changeover times (Shingo, 1989; Mileham et al., 1999). A large body of knowledge on manufacturing flexibility and its positive effects on supply chains and lead-time reductions is available (Lim, 1987; Slack, 1987; Gerwin, 1993; Upton, 1994; Blumenfeld et al., 1999; Gindy et al., 1999), and was reviewed inter alia, by de Toni & Tonchia (1998) and Beach et al. (2000). Manufacturing flexibility should however not be seen as the only solution for achieving system flexibility as pointed out by Lau (1999) who further explains that managers should also include other functions in their company as well as their suppliers, an argument that has formed the basis for supply chain management since its inception in the 1980’s (e.g. Houlihan, 1985).

Inventory

Despite a tendency across industries to become ‘lean’ and operate with less stock (Womack et al., 1990; Womack & Jones, 1996), inventory buffers still exist in many places in supply chains (Davis, 1993; Hines, 1998; Holweg & Pil, 2004). Turnbull et al. (1993) reported that when required to deliver just in time to their customers, smaller suppliers were still relying on inventories instead of synchronised production, an approach that Womack & Jones (1996) were also observing: ‘Most of the applications of JIT, even in Japan, have involved Just-in-Time supply, not Just-in-Time production’ (Womack & Jones, 1996, p. 88). As Davis (1993) further explains, more than half of the total stock in Hewlett Packard’s supply chain existed as a buffer against demand uncertainty. Buffer stocks can clearly increase a supply chain’s volume, mix or delivery responsiveness, even if they might reduce a supply chain’s product responsiveness (Hines, 1998), yet Griffith & Margetts (2000) conclude that such buffer stocks can also easily
offer a false sense of security, while hiding problems in the underlying processes – a message strongly supported by related research from the system dynamics field.

Shingo (1989) further differentiates between two types of stock: ‘naturally’ occurring and ‘necessary’ stock. The latter becomes necessary due to process inefficiencies, such as stock produced to offset weakness in the production process in terms of quality, or stock produced due to a P:D ratio of greater than one. However, many recent contributions argue that in the case that the P:D ratio cannot be reduced to below one, decoupling points (Hoekstra & Romme, 1992; Olhager, 2003) and thus inventory are required in supply chains (Naylor et al., 1999; Christopher, 2000). The decoupling point in a supply chain is the inventory point at which end customer demand meets forecast driven production. As explained by Olhager (2003), market-related, production-related and product-related factors have to be considered in the positioning of the decoupling point, while Bozarth & Chapman (1996) primarily argue for a contingency on product architecture. On the one hand, a decoupling point that is close to the end-customer will lead to shortened lead-times and thus partially increased responsiveness. On the other hand, increased inventory holding costs will be incurred due to an increase in value added together with higher product variety in later stages of the supply chain (Holweg & Pil, 2004).

Product Architecture

Product architecture and the related concepts of modularisation (Starr, 1965; Ulrich, 1995; Baldwin & Clark, 1997), postponement and late configuration (Feitzinger & Lee, 1997; Lee & Tang, 1997; Pagh & Cooper, 1998; Pil & Holweg, 2004), are often linked to the mass customisation debate (Lampel & Mintzberg, 1996; Gilmore & Pine II, 1997). Adjusting the product architecture is seen as a way to employ decoupling points to offer a wide variety of products to end customers while reducing inventory holding costs for products with a P:D ratio of greater than one. While the personal computer industry has implemented BTO strategies through an assemble-to-order approach, this is not feasible in sectors that lack standardised component interfaces, such as in automotive (Holweg & Pil, 2004). For the latter, either a purchase and build-to-order strategy (Hoekstra & Romme, 1992) or late configuration / postponement is used. Postponement refers to the postponement of specification decisions in the production process to reduce initial product variety and for the main variant explosion to occur once the demand for the exact specifications is known, e.g. through a customer order (Pagh & Cooper, 1998), which can in some cases happen after the final assembly and is then usually referred to as late configuration (Holweg & Pil, 2004). Daugherty & Pittman (1995) for
example, report the increasing use of late configuration in distribution centres which they thus refer to as ‘accessorization centers’.

While late configuration can increase a supply chain’s responsiveness, decision makers need to be aware that a scenario in which the decoupling point is positioned after the final assembly operation, is no longer considered a BTO supply chain and hence does not yield the associated benefits. Furthermore, the product architecture is linked to both manufacturability and internal product variety, which can inhibit the responsiveness of manufacturing operations and therefore the responsiveness of the supply chains they are part of (Fisher et al., 1994; Christopher, 2000; Holweg & Pil, 2004). As a result, supply chain partners increasingly make product development decisions jointly (Petersen et al., 2005) and ideally align them with the respective operations strategy and supply chain structure (Fixson, 2005).

**Information Integration**

The detrimental effects of a lack of demand visibility on the performance of the supply chain has been pointed out by Forrester (1958) almost half a century ago. Building upon Forrester’s bullwhip effect, Towill (1997) explains how demand uncertainty (especially with regards to aggregated volumes per period) further upstream in supply chains is created to a great extend by downstream supply chain members, referring to Burbidge’s work on the effects of re-order levels on supply chains (Burbidge, 1961). Subsequent studies were conducted by Sterman (1989) and Lee et al. (1997) confirming these findings. Inman & Gonzalves (1997) however report that schedule fluctuations can also be reduced by individual supply chain partners, similar to a phenomenon that Hines (1998) refers to as ‘quality filters’ with regards to quality improvements in supply chains.

Suggestions for improvements have been made by a variety of authors (Bagchi & Skjoett-Larsen, 2002; Gunasekaran & Ngai, 2004), which are commonly referred to as ‘supply chain collaboration’ (Holweg et al., 2005b) or ‘virtual supply chain’ (Christopher, 2000), and are targeted at creating transparency or visibility of both demand and capacity information in the supply chain without time delays. Information sharing is usually achieved through the increased use of information technology or a closer integration between supply chain partners (Bagchi & Skjoett-Larsen, 2002) in order to facilitate two-way communication. Various authors also take a more critical view on the extent to which information systems can solve supply chain problems and increase their responsiveness, pointing out that other inter-organisational aspects, such as trust (Mclvor et al., 2003; Akkermans et al., 2004) and further process coordination and
organisational integration (Burbidge, 1961; Lee, 2000) should accompany a mere information exchange.

**Coordination and Resource Sharing**

Coordination and resource sharing refers to how processes, value-adding steps and related decisions are coordinated and potentially rearranged across firm boundaries and how internal or external resources are shared to add value to products at interfaces in supply chains. Few studies have looked at the extent to which supply chain partners coordinate their processes, yet some examples can be found in related publications. Burbidge (1961) for example describes how misaligned re-order levels can create demand variability and uncertainty in supply chains. Li & Liu (2006) further demonstrate that supply chain partners can benefit from a co-ordination of quantity discount policies. One additional recent development is vendor managed inventory (VMI), an arrangement under which suppliers take responsibility for maintaining stock levels at their customers’ sites, thereby relieving their customers of re-ordering decisions (Holweg et al., 2005b).

The realignment of value-adding tasks in some supply chains is described by Lee (2000) using the computer industry as an example. Collins et al. (1997) describe how similar shifts in the automotive value chain can contribute to increased supply chain responsiveness and general performance by leveraging core competencies and realigning complexity. The coordination of processes and achievement of synergies can be further supported by outsourcing processes at the interfaces to third parties, such as logistics service providers (Spencer et al., 1994).

**Organisational Integration**

The supply chain literature features many contributions on the effects of organisational integration on supply chain performance and its contributing factors (Suarez et al., 1995; Lamming, 2000; Droge et al., 2004), and the ways in which organisational integration can be achieved (Rich & Hines, 1997; Bagchi & Skjoett-Larsen, 2002; Perona & Saccani, 2004). Hill & Chambers (1991), Suarez et al. (1995), Liker & Wu (2000) and Droge et al. (2004) all agree that close organisational integration positively influences manufacturing and supply chain performance and responsiveness. The activities that constitute organisational integration can range from extensive communication to close cross-supplier integration in the case of Japanese
supplier associations. This is seen as one of the competitive advantages of Japanese vehicle manufacturers (Rich & Hines, 1997). Suppliers can also employ engineers at customer sites to facilitate better communication and faster problem solving (Ikeda, 2000). Bagchi & Skjoett-Larsen (2002) provide a concise overview of the most common characteristics of organisational integration which include joint design teams, process and quality teams, joint performance measurement and problem solving, amongst others. As Wagner (2003) points out though, one needs to differentiate between supplier integration during product development and operational execution. Sako & Helper (1998) further conclude that some forms of integration can facilitate inter-firm trust, a concept further discussed below.

Spatial Integration and Logistics

Spatial integration can take many forms, from loose co-location to formalised supplier parks (Larsson, 2002). Many factors have been discussed in relation to spatial integration, such as trust and commitment (see for example Collins et al., 1997), problem solving capabilities (see for example Salerno & Dias, 2002), labour relations (see for example Sako, 2003), and synergies (see for example Larsson, 2002). The most benefit however seems to come from logistical proximity which reduces transportation lead-times and costs to a minimum, often supported by dedicated infrastructure, such as conveyor belts. The latter are of increasing importance due to just-in-time or even in-sequence deliveries (Frigant & Lung, 2002; Larsson, 2002). In addition, spatial proximity can facilitate cross-firm cost sharing of logistics costs, for example through 'milk-run' deliveries (Holweg & Pil, 2004), without negatively affecting the supply chain’s responsiveness. Rather, a reduction in transport times is likely to occur which will increase responsiveness, yet such an effect is contingent upon the form and functionality of such spatial configurations within the wider supply chain, also related to the positioning of the decoupling point, especially in supply chains that source components from overseas (Reichhart & Holweg, 2005). Spatial integration in the form of local logistics centres or formalised suppliers parks are particularly common at decoupling points in build-to-order or assemble-to-order supply chains, like those at Dell (Kapuscinski et al., 2004) and Volvo Cars (Larsson, 2002).

3.4 Framework Consolidation

Based on the discussion of the external requirements and internal determinants of supply chain responsiveness, the framework in Figure 2 is suggested as a conceptualisation of supply
chain responsiveness. In addition to the factors discussed above, it also features some ‘relational factors’. Relational factors govern or simply influence the relationship between the supply chain and its customers. Depending on the particular supply chain and the characteristics of the transactions carried out between the partners, those factors will be more or less formalised (see for example Williamson, 1979). Relational factors are neither external requirements nor pure internal determinants. They can be influenced by both parties and impact upon the need, as well as the ability, to be responsive. Relational factors often mitigate certain uncertainties inherent in markets and trading relationships in order to benefit from associated savings.

Under rolling schedules for example, there are two important tools for limiting demand uncertainty contractually. First, so-called ‘frozen horizons’ (usually 4-6 days for automotive manufacturers), in which the production schedule does not change, can substantially reduce the need for very short-term responsiveness (Karlsson & Norr, 1994; Harrison, 1996; Inman & Gonzalves, 1997; Holweg & Pil, 2004). Furthermore, in order to control the period prior to the frozen horizon, a so-called ‘quantity-flexibility’ (QF) contract can be used to set upper limits for possible changes, thereby allowing the supply chain to focus its resources efficiently on those changes that are allowed (Tsay & Lovejoy, 1999; Sethi et al., 2004; Krajewski et al., 2005).

For other scenarios time-based pricing has been suggested as a tool for reducing demand uncertainty and improving capacity utilisation, thereby allowing a supply chain to be more responsive to last minute urgent orders and to recover the costs thereof (Holweg & Pil, 2004; Jiang & Geunes, forthcoming). Time-based pricing however is a tool that has remained largely unused in most traditional manufacturing industries despite its advantages for BTO supply chains and its success in many service industries, such as air travel (Holweg & Pil, 2004).

Finally, trust and mutual commitment, though having been partially addressed already, constitute a relational factor in their own right. Handfield & Bechtel (2002) provide the most direct study on the relationship between trust and responsiveness and show a positive correlation between the two. They also confirm previous arguments by Dyer (1994) of links between trust and dedicated assets, which can further contribute to a supply chain’s responsiveness. Sako (1998) has shown that inter-firm trust has a positive relationship on the business performance of automotive suppliers, and in particular on responsive just-in-time delivery arrangements.

Finally, a control factor is of particular importance for quantitative studies, as two different supply chains with the same external requirements don’t demonstrate the same level of responsiveness if one of the supply chains can deliver products more reliably and with better
quality. Therefore, delivery reliability (commonly measured as percent of on-time and in-full shipments) and product quality (commonly measured as defects per million units) were added to the framework. Depending on the particular industry setting, these specific measures can be replaced as long as some control factor measures the supply chain’s ability to meet the requirements placed upon it.

Given the assumption that responsive supply chains incur higher costs than less responsive supply chains (Fisher, 1997), the two-sided approach presented in the above framework can facilitate the design of more cost-efficient supply chains. It is important to understand the framework as a tool designed to be applied to every important interface in a supply chain, starting with the end customer and going upstream. At every interface the affected supply chain partners, usually under the lead of the OEM, can evaluate how much responsiveness is really required and adjust the external requirements and thereby the supply chain’s demonstrated responsiveness accordingly. As a second step, they have to determine how this responsiveness can be best achieved through an appropriate combination of the internal determinants available. The latter translates into a supply chain’s potential responsiveness.
responsiveness. The first step largely reduces costs in the supply chain by reducing the need to be responsive in the first place, while the second step adjusts and reconfigures the supply chain to deliver the required level of responsiveness cost-efficiently. The 'relational factors' can support these efforts either way.

3.5 Framework Application and Recommendations for Future Research

The primary purpose of the above supply chain responsiveness framework is to support future academic studies, but it can equally be applied to managerial problems. In our view key questions for future research include the study of responsiveness in different industry settings, and in particular responsiveness profiles of whole supply chains (i.e. the responsiveness levels at different points in a given supply chain). The framework provides a clear set of definitions and the foundation for such research without limiting its application to a particular stance within the epistemological spectrum. By differentiating between potential and demonstrated responsiveness and the respective grouping of the identified factors into internal determinants and external requirements, we have provided the basis for a two-sided measurement of this complex concept. The framework is not limited to quantitative approaches, as it does not prescribe any quantitative measure to be used for the individual factors. As such, future researchers are free to either use the framework for qualitative investigations, or if desired, refer to the quantitative studies referenced for existing measures.

A second main benefit is that the framework is specific enough to focus the researcher's attention on the key areas that enable and require responsiveness while not limiting investigations to individual nodes in the supply chain. Due to this it can be used to measure responsiveness at single nodes as well as to create responsiveness profiles of entire supply chains. The main areas that this framework can assist future research with are:

i) The trade-offs between responsiveness and costs: The framework allows for a comparison of internal determinants (and associated costs) against actually demonstrated responsiveness levels. A quantitative study capable of confirming the impact of individual factors on a supply chain's responsiveness is still amiss, and this kind of study would extend the research into trade-offs in operations management (cf. Skinner, 1969; Boyer & Lewis, 2002) from factory to supply chain level. While there seems to be wide-spread agreement that there are trade-offs between responsiveness (or more generally, flexibility) and cost-efficiency in supply chains (e.g. Fisher, 1997; Randall et al., 2003), we are not aware of any empirical study that
examines to what extent entire supply chains can push their performance frontiers (cf. Schmenner & Swink, 1998) to become more competitive.

ii) Improve **understanding of inventory** in the supply chain: A well-defined responsiveness concept offers new possibilities for modelling supply chains. Instead of modelling production processes within individual echelons in detail, one could conceptualise each node in terms of its responsiveness. This can facilitate more quantitative research into explaining the operational need for inventory through the responsiveness at individual inventory locations in supply chains.

iii) **Process coordination**: Another area for future investigations is how decisions and process changes at individual echelons impact on responsiveness-related costs in the upstream supply chain, an area where the debate has so far remained rather conceptual, anecdotal or narrative in nature. While process coordination amongst supply chain members has been identified as an important aspect of supply chain integration (see section on ‘external requirements’), little empirical evidence exists that illustrates and more specifically quantifies the savings that can be achieved through process coordination. Unless means are developed for this purpose, individual firms will remain reluctant to bear the initial costs of process changes in the light of unknown cost savings in other supply chain partners’ operations.

For practitioners, the framework can serve two functions. First, it facilitates an understanding of how individual nodes in supply chains interact in determining the overall system’s ability to respond to customer demand, showing which factors, and thus decisions, can increase or decrease this responsiveness of the overall supply chain. This includes an understanding of which actions can be taken to reduce the need for responsiveness at every node (for example through the use of ‘frozen horizons’ or level scheduling). Second, it can serve to benchmark the responsiveness of supply chains within an industry, bearing in mind that cross-industry benchmarks would have to further consider differences in end product architecture and complexity (Kritchanchai & MacCarthy, 1999). Practitioners can also use the framework to evaluate how they can eliminate unnecessary requirements from their suppliers by removing deficiencies from their own processes and by making more favourable product architecture decisions for the entire supply chain.
4 Conclusions

In this paper, the literature on operational responsiveness and flexibility was reviewed and consolidated, taking a particular focus on concepts that relate to creating a customer-responsive supply chain system. While seminal contributions have been made (e.g. Slack, 1987; Upton, 1994) which have developed our understanding of the underlying theoretical concepts, the definitions used are at times contradictory and generally do not provide the crucial distinction between the ‘responsiveness’ and ‘flexibility’ concepts. Furthermore, while the individual cause-and-effect relationships provide some insights into selected linkages between the key variables, they do not provide the holistic view required to describe this complex concept. Based on our review, we hence propose that:

i) There are different types of responsiveness, both in terms of the \textit{unit of change} (product, volume, mix and delivery responsiveness) and in terms of the \textit{time horizon} affected (short, medium or even long-term responsiveness).

ii) A supply chain can (and is even likely to) exhibit different levels of responsiveness, depending on where in the supply chain its responsiveness is measured.

iii) One needs to distinguish between factors that \textit{require} a supply chain to be responsive and those that \textit{enable} it to be responsive. This leads to the conclusion that at any point in a supply chain, the supply chain will have a \textit{potential} and \textit{demonstrated} responsiveness.

Moreover we argue that, based on the framework developed, individual supply chain partners and their integration with each other can either increase or decrease an overall supply chain’s responsiveness. This proposition leads to the conclusion that there are different ways to increase the responsiveness of supply chains, some of which will be more costly than others, yet the real costs and benefits of increasing a supply chain’s responsiveness cannot be assessed by considering individual echelons only. As a consequence, a supply chain’s responsiveness can only be understood by investigating the responsiveness of its individual members and their interactions, yet such investigations should be specific about the type of responsiveness studied. In particular, we would like to argue for a two-part approach to building responsive and cost-efficient supply chains, a combination crucial to the expansion of BTO supply chains: first, a reduction of requirements at each echelon, and second, an adjustment of the internal determinants of responsiveness in the upstream supply chains. Such an approach will also
ensure that – at every echelon – the supply chain’s potential and demonstrated responsiveness are aligned, avoiding unnecessary costs.

We hope that the definition and framework will inform the further study of supply chain responsiveness by providing both a clarification of the concepts involved, and a guideline to conduct more quantitative, yet holistic investigations in this area.

References


Burbidge, J.L., 1961 "The new approach to production" Production Engineer, Vol 40 No 12, pp. 769-784


Pagh, J.D., Cooper, M.C., 1998 "Supply Chain Postponement and Speculation Strategies: How to Choose the Right Strategy" *Interfaces*, Vol 34 No 5, pp. 394-403


Sethi, S.P., Yan, H., Zhang, H., 2004 "Quantity Flexibility Contracts: Optimal Decisions with Information Updates" Decision Sciences, Vol 35 No 4, pp. 691-712


Towill, D.R., 1997 "FORRIDGE - Principles of good practice in material flow" Production Planning & Control, Vol 8 No 7, pp. 622-632

Towill, D.R., Christopher, M., 2002 "The Supply Chain Strategy Conundrum: To be Lean Or Agile or To be Lean And Agile?" International Journal of Logistics: Research and Applications, Vol 5 No 3, pp. 299-309

Tsay, A.A., Lovejoy, W.S., 1999 "Quantity Flexibility Contracts and Supply Chain Performance" Manufacturing & Service Operations Management, Vol 1 No 2, pp. 89-111


