Information Sharing and Supply Chain Coordination*

Fangruo Chen

Graduate School of Business
Columbia University
New York, NY 10027

October 2001
Revised November 2002
Edited January 22, 2003

*To appear in the Handbooks in Operations Research and Management Science: Supply Chain Management, edited by Ton G. de Kok and Stephen C. Graves. Part of this chapter was completed during the author's visit at the Stanford University Graduate School of Business, 2000-2001. Thanks to the OIT group of the Stanford Business School for their hospitality. The chapter has benefited from comments and suggestions made by the following colleagues: Gérard P. Cachon, Charles J. Corbett, Ton G. de Kok, Guillermo Gallego, Constantinos Maglaras, Özalp Özer, L. Beril Toktay, and Jan A. Van Mieghem. Financial support from National Science Foundation, Columbia Business School, and Stanford Global Supply Chain Management Forum is gratefully acknowledged.

1 Introduction

The performance of a supply chain depends critically on how its members coordinate their decisions. And it is hard to imagine coordination without some form of information sharing. A significant part of supply chain management research is devoted to understanding the role of information in achieving supply chain coordination. It is the purpose of this chapter to review this literature.[1]

The first part of the chapter focuses on papers that have contributed to our understanding of the value of shared information. We first consider information pertaining to the downstream part of the supply chain. The next is upstream information. Finally, we discuss papers that study the consequences of imperfect transmission of information. All the papers here adopt the perspective of a central planner whose goal is to optimize the performance of the entire supply chain.

The chapter then proceeds to discuss papers that address incentive issues in information sharing. Here it is made explicit and prominent that supply chains are composed of independent firms with private information. The goal is to understand whether or not incentives for sharing information exist, and if not, how they can be created. This section is divided into three parts. When one firm has superior information, it may hide it to gain strategic advantage or to reveal it to gain cooperation. If the former, the less-informed party may try to provide incentives for the informed to release its information. This is called screening, and it constitutes the first part of the section. If the informed tries to convey its private information, it is often the case that he has to "put his money where his mouth is" in order to be credible, i.e., signaling. This is the second part of the section. The last part of the section deals with situations where it is difficult to say if a supply chain member has more or less information: they simply have different information about something they all care about (e.g., the potential market size of a product). Here a common question is if information sharing will emerge as an equilibrium outcome in some noncooperative game.

The chapter ends with some thoughts on future research directions.

The structure of the chapter provides an implicit taxonomy for thinking about research on supply chain information sharing. Specifically, the sections and subsections provide categories so that (hopefully) every piece of relevant research finds its home. It is important to mention that the unnumbered, boldfaced headings are meant to represent examples within a category, and the examples may not be exhaustive. For example, §2.2 deals with the value of upstream information, and within this subsection are several examples (i.e., cost, leadtime, capacity information). This should not be taken to mean that these are the only types of upstream information nor that they can only come from upstream. For example, an upstream supply chain member (e.g., a manufacturer) may have some private information about demand that the (downstream) retailers don't have, so demand information is a possible type of upstream information. On the other hand, a seller may not know a buyer's cost structure, so cost information can also come from downstream. In other words, the headings without section/subsection numbers are not part of the taxonomy anymore.

A few words on how we choose the papers and what we are going to do with them. The emphasis here is modeling, not analysis. Therefore, we prefer papers with modeling novelties. And we want as much variety as possible given limited space. So if there are several papers that are close to each other in the "novelty space," we will just take one with a simple reference to the others.

[1] For a summary of recent industry initiatives to improve supply chain information flows, see Lee and Whang (2000).

Within a collection of papers, if there is no clear logical progression, we will simply review them in chronological sequence. We will often present a model without stating all the assumptions. The mentioning of results is often brief and is meant to whet your appetite so that the original paper becomes irresistable. We sometimes purposely "trivialize" a model by further specializing it (e.g., by making assumptions that a top journal referee would be hard pressed to swallow). The reader should be assured that this is done for ease of presentation and for crystallizing the main ideas without getting bogged down on details. In terms of notation, we try to be consistent with the original paper. The risk of this is that the reader may see different symbols for, say, the wholesale price. But the chapter is sufficiently modularized that we hope, the reader can easily see which symbol belongs where. Needless to say, the papers presented in this chapter reflect the author's knowledge and taste at a certain point in time, the former of which is inevitably incomplete while the latter is constantly evolving.

2 Value of Information

The perspective taken by this part of the literature is often that of a central planner, who determines decision rules to optimize the performance of the entire supply chain. The decision rules reflect the information available to the managers who implement the rules. For example, the inventory manager at a supply chain stage only has access to local inventory information, and so the decision rule (determined by the central planner) for this manager must be based on the local information. Clearly, if we increase the information available to the manager (e.g., by providing access to inventory information at other supply chain locations), the set of feasible decision rules is enlarged and the supply chain's performance may improve. The resulting improvement is then the value of the additional information. This section reviews papers that try to quantify the value of information in different supply chain settings.

2.1 Downstream Information

A significant part of the literature is interested in the value of information pertaining to the downstream part of the supply chain (i.e., the part that is closer to the end customers). We first consider papers that deal with information sharing within a supply chain. A typical setup here is one where the members of a supply chain share their information about the end customer demand, in the form of realized demand or updated demand forecasts, although other types of information are also discussed. We then review models where information sharing takes place at a supply chain's boundary, e.g., when the supply chain's customers provide advance warnings of their demands. While most researchers use the supply chain-wide costs as the performance measure, there is a stream of papers that use the "bullwhip effect" (i.e., the amplification of the order variance up the supply chain) as a surrogate performance measure. These papers are considered at the end of this subsection.

Information Sharing within a Supply Chain

Here are some papers that study the value of giving the upstream members of a supply chain access to downstream information, which can be the point-of-sale data or information about the control rule used by a downstream member. The customer demand process can be stationary or nonstationary, and the structure of the supply chain can be serial or distributional.

Chen (1998a) studies the value of demand/inventory information in a serial supply chain. The model consists of N stages. Stage 1 orders from stage 2, 2 from 3, $\cdots$, and stage N orders from an outside supplier with unlimited stock. The leadtimes from one stage to the next are constant and represent delays in production or transportation. The customer demand process is compound Poisson. When stage 1 runs out of stock, demand is backlogged. The system incurs linear holding costs at every stage, and linear backorder costs at stage 1. The objective is to minimize the long-run average total cost in the system.

The replenishment policy is of the (R,nQ) type. Each stage replenishes a stage-specific inventory position according to a stage-specific (R,nQ) policy: when the inventory position falls to or below a reorder point R, the stage orders a minimum integer multiple of Q (base quantity) from its upstream stage to increase the inventory position to above R. In case the upstream stage does not have sufficient on-hand inventory to satisfy this order, a partial shipment is sent with the remainder backlogged at the upstream stage. The base quantities are fixed and the reorder points are the only decision variables. Moreover, the base quantities, which are denoted by Q_i for stage i, $i = 1, \cdots, N$, satisfy the following integer-ratio constraint:

$$Q_{i+1} = n_i Q_i, \quad i = 1, \cdots, N-1$$

where n_i is a positive integer. This assumption is made to simplify analysis, but it also reflects some practical considerations aimed at simplifying material handling such as packaging and bulk breaking. Moreover, there is evidence that the system-wide costs are insensitive to the choice of base quantities. [2]

Two variants of the above (R,nQ) policy are considered. One is based on *echelon stock*: each stage replenishes its echelon stock with an echelon reorder point. A stage's echelon stock is the inventory position of the subsystem consisting of the stage itself as well as all the downstream stages, which includes the outstanding orders of the stage, either in transit or backlogged at the (immediate) upstream stage, plus the inventories in the subsystem, on hand or in transit, minus the customer backorders at stage 1. Let R_i be the echelon reorder point at stage i, $i = 1, \cdots, N$. Therefore, under an echelon-stock (R,nQ) policy, stage i orders a multiple of Q_i from stage $i+1$ every time its echelon stock falls to or below R_i.

Alternatively, replenishment can be based on *installation stock*: each stage controls its installation stock with an installation reorder point. A stage's installation stock refers to its local inventory position, i.e., its outstanding orders (in transit or backlogged at the upstream stage) plus its on-hand inventory minus backlogged orders from its (immediate) downstream stage. Let r_i be the installation reorder point at stage i, $i = 1, \cdots, N$. Therefore, under an installation-stock (R,nQ) policy, stage i orders a multiple of Q_i from stage $i+1$ every time its installation stock falls to or below r_i. Note that echelon-stock (R,nQ) policies require centralized demand information, while installation-stock (R,nQ) policies only require local 'demand' information, i.e., orders from the immediate downstream stage. When every customer demands exactly one unit, i.e. the demand process is simple Poisson, each order by stage i is exactly of size Q_i, $i = 1, \cdots, N$. In this case, the (R,nQ) policy reduces to the (R,Q) policy.

[2] Such insensitivity results have been established for single-location models, see Zheng (1992) and Zheng and Chen (1992). This property is likely to carry over to multi-stage models. Also, the optimality of (R,nQ) policies has been established by Chen (2000a) for systems where the base order quantities are exogenously given.

From Axsäter and Rosling (1993), we know that installation stock (R,nQ) policies are special cases of echelon stock (R,nQ) policies. The two policies coincide when

$$R_1 = r_1, \text{ and } R_i = R_{i-1} + Q_{i-1} + r_i, \quad i = 2, \cdots, N.$$

(Note that r_i is an integer multiple of Q_{i-1} for $i \geq 2$.) To see the intuition behind this result, suppose the demand process is simple Poisson so that (R,nQ) policies reduce to (R,Q) policies. Under the installation stock (R,Q) policy, orders are "nested" in the sense that every order epoch at stage i coincides with an order epoch at stages $i-1, i-2, \cdots, 1$. The installation stock at stage j after each order is $r_j + Q_j$ for all j. Consequently, just before stage i places an order, its echelon stock, which is the sum of the installation stocks at stages 1 to i, is $\sum_{j=1}^{i-1}(r_j + Q_j) + r_i$. Let this echelon stock level be R_i, $i = 1, \cdots, N$. It is easy to verify that the echelon reorder points so defined satisfy the above equalities and the resulting echelon-stock policy is identical to the installation-stock policy.

Echelon stock (R,nQ) policies have very nice properties. As a result, the optimal echelon reorder points can be determined sequentially in a bottom-up fashion starting with stage 1. Essentially, after a proper transformation, the batch-transfer model reduces to a base-stock model of the Clark and Scarf (1960) type. On the other hand, installation stock (R,nQ) policies are not as nice; a heuristic algorithm is available for determining the optimal installation-stock reorder points, based on several easy-to-compute bounds.

As mentioned earlier, echelon-stock policies require centralized demand information, while installation stock policies only require local information. The relative cost difference between the two is a measure of the value of centralized demand information. An extensive numerical study (with 1,536 examples) shows that the value of information ranges from 0% to 9% with an average of 1.75%.

Gavirneni et al. (1999) study different patterns of information flow between a retailer and a supplier. The retailer faces independent and identically distributed (i.i.d.) demands and replenishes his inventory by following an (s, S) policy. At the beginning of each period, the retailer reviews his inventory level (on-hand inventory minus customer backorders), and if it is below s, he places an order with the supplier to raise the inventory level to S. The supplier satisfies this order as much as possible. In the event the supplier does not have sufficient on-hand inventory to satisfy the retailer order, a partial shipment is made to the retailer, and the retailer obtains the unfilled part of the order from an external source. There is no delivery leadtime with both sources of supply. Customer demand arises at the retailer during the period, with complete backlogging. The focus of the analysis is the supplier, who after satisfying (partially or fully) a retailer order at the beginning of each period, decides how much to produce for the period. Production takes one period and is subject to a capacity constraint. The supplier incurs linear inventory holding costs and linear penalty costs for lost retailer orders. The objective is to determine a production strategy to minimize the supplier's costs, under various scenarios that differ in terms of the supplier's information about the downstream part of the supply chain.

The first scenario assumes that the supplier has no information about the retailer except for the orders the retailer has placed in the past. Moreover, the supplier is rather naive in assuming that the orders from the retailer are i.i.d. Under this assumption, the best the supplier can do is to follow the modified base-stock policy with the same order-up-to level in every period.[3]

[3] A modified base-stock policy with order-up-to level z works like this: if the inventory position (on-hand inventory plus work-in-process minus backorders) is less than z, produce as much as possible under the capacity constraint to increase it to z; if the inventory position is above z, produce nothing. In the context of Gavirneni et al., there are no backorders at the supplier, only lost sales, and there is no work-in-process at any review time. The optimality of such a policy has been established by Federgruen and Zipkin (1986a,b).

The second model assumes that at the beginning of each period, the supplier knows the number of periods i since the last retailer order. In addition, the supplier knows the demand distribution at the retailer, the fact that the retailer follows an (s,S) policy, and the specific policy parameters used by the retailer. Given this information, the supplier is able to determine the probability that the retailer is going to place an order in the coming period and the distribution of the order size. This influences the current production decision. It is shown that the optimal policy for the supplier in this case is again a modified base-stock policy with state-dependent order-up-to level z_i.

The third and final model assumes that the supplier has access to all the information available to her in the second model. In addition, at the beginning of each period, the supplier knows the value of j, the number of units sold by the retailer since the last retailer order. Again, a modified base-stock policy with state-dependent order-up-to level z_j is optimal. [4]

A numerical study has been conducted by Gavirneni et al. to understand the differences between the above three models. From the first model to the second, the percentage decrease in supplier costs varies from 10% to 90%; and the savings increase with capacity. From the second model to the third model, the savings range from 1% to 35 %. They also report sensitivity results on the cost savings as a function of the supplier capacity, the supplier cost parameters, the retailer demand distribution, and the retailer's policy parameters. The key observations are: 1) when the retailer demand variance is high, or the value of $S - s$ is either very high or very low, information tends to have low values, and 2) if the retailer demand variance is moderate, and the value of $S - s$ is not extreme, information can be very beneficial.

Lee et al. (2000) study the value of sharing demand information in a supply-chain model with a nonstationary demand process. The supply chain consists of two firms, one retailer and one manufacturer. The customer demand process faced by the retailer is an AR(1) process:

$$D_t = d + \rho D_{t-1} + \epsilon_t$$

where $d > 0$, $-1 < \rho < 1$, and ϵ_t are independent random variables with a common normal distribution with mean zero and variance σ^2. Both firms know the values of the parameters of the demand process, i.e., d, ρ and σ. The retailer sees the realization of demand in each period, while the manufacturer's information depends on, well, what the retailer provides.

The retailer reviews its inventory at the end of each period. Take period t. The retailer satisfies D_t, the demand for period t, from its on-hand inventory with complete backlogging. At the end of the period, the retailer places an order for Y_t units with the manufacturer. The manufacturer satisfies this order from its own on-hand inventory, also with complete backlogging. [5] At the beginning of the next period (period $t+1$) the manufacturer places an order with an outside supplier with ample stock

[4] One can imagine that the retailer transmits his demand data to the supplier in every period via some electronic medium. The supplier can then determine the value of j and use that information in her production decisions (through the state-dependent policy). A supplier's optimal use of timely demand information from a retailer has been addressed in other papers. For example, Gallego et al. (2000) address this issue in a continuous-time model without capacity constraints. They also show that it is not always in the retailer's interests to share demand information with the supplier. Another reference is Bourland et al. (1996) who study a supply chain model with a component plant (seller) and a final assembly plant (buyer). The production cycles of the two factories do not coincide. Traditionally, information sharing occurs only when the buyer places an order. They study the impact of real-time communication of the buyer's demand data.

[5] The original assumption made in Lee et al. is that if the manufacturer's on-hand inventory is insufficient to satisfy a retailer order, the manufacturer will make up the shortfall from an external source. The analytical benefit of this assumption is that the retailer always gets its orders filled in full. But this actually complicates the demand process at the manufacturer, who is effectively operating under a lost-sales regime. It is well known that when we

to replenish its own inventory. For easy exposition, we assume that the leadtimes at both sites are zero, i.e., transportation from the outside supplier to the manufacturer or from the manufacturer to the retailer is instantaneous. (Note that if part of Y_t is backlogged at the manufacturer, that backlog will remain there until the end of period $t+1$, even though the manufacturer's replenishment leadtime is zero. This is because the manufacturer only fills the retailer's orders at the ends of periods. The case with a different sequence of events can be analyzed similarly.)

We begin with the retailer's ordering decisions. Suppose we are now at the end of period t. What is the ideal inventory level for the retailer going into period $t+1$? Since the delivery leadtime is zero, the retailer can be myopic, i.e., to minimize its expected holding and backorder costs incurred in period $t+1$. The demand in period $t+1$ is $D_{t+1} = d+\rho D_t+\epsilon_{t+1}$, which is normally distributed with mean $d+\rho D_t$ and standard deviation σ. (D_t has been realized by the end of period t.) Therefore, the ideal inventory level going into period $t+1$ is

$$S_t = d + \rho D_t + k\sigma$$

where k is a constant depending on the holding and backorder cost parameters at the retailer. (This is a well known formula for the newsvendor model with normal demand.) To derive the order quantity Y_t, suppose the retailer's inventory at the beginning of period t is S_{t-1}. Thus

$$S_{t-1} - D_t + Y_t = S_t$$

or

$$Y_t = D_t + S_t - S_{t-1}.$$

This gives us the demand process facing the manufacturer. (The value of Y_t can be negative, an unpleasant scenario, which should indicate to you the potential suboptimality of the myopic policy. But let's confine ourselves to cases where this rarely happens.)

Now consider the manufacturer's ordering decisions. Suppose we are at the beginning of period $t+1$, having just received and satisfied (completely or partially) the retailer order Y_t. What is the manufacturer's ideal inventory level for the beginning of period $t+1$? Since the manufacturer's replenishment leadtime is zero and the outside supplier has ample stock, the manufacturer can also be myopic, minimizing its expected holding and backorder costs in period $t+1$ alone. (This argument is again not water-tight because the manufacturer's myopic inventory level in one period may prevent it from reaching its myopic inventory level in the next period, i.e., there can be too much inventory. If so, the myopic policy is not optimal. Let's not worry about this here.) In period $t+1$, the retailer order is Y_{t+1}, which can be expressed as

$$Y_{t+1} = D_{t+1} + S_{t+1} - S_t = D_{t+1} + \rho(D_{t+1} - D_t).$$

Since $D_{t+1} = d + \rho D_t + \epsilon_{t+1}$, we have

$$Y_{t+1} = (1+\rho)d + \rho^2 D_t + (1+\rho)\epsilon_{t+1}. \tag{1}$$

The manufacturer's ideal inventory for period $t+1$ can be written as

$$T_t = E[Y_{t+1}] + K\, Std[Y_{t+1}]$$

combine lost sales with a positive replenishment leadtime, it is very difficult to characterize the distribution of the total demand (or satisfied retailer orders in this case) over a leadtime. This problem is not addressed in Lee et al. The same comment applies to Raghunathan (2001), to be reviewed shortly.

where K is a constant depending on the manufacturer's holding and backorder costs. Moreover, the manufacturer's minimum expected (one-period) cost is proportional to $Std[Y_{t+1}]$, the value of which depends on what the manufacturer knows about the retailer's demand process at the beginning of period $t+1$.

As mentioned earlier, the manufacturer knows the value of Y_t in any case. If there is no sharing of demand information between the retailer and the manufacturer, the latter does not see D_t. Since $Y_t = D_t + \rho(D_t - D_{t-1})$ and $D_t = d + \rho D_{t-1} + \epsilon_t$, we have

$$D_t = \frac{Y_t - d - \epsilon_t}{\rho}.$$

Plugging this into (1), we have

$$Y_{t+1} = d + \rho Y_t - \rho\epsilon_t + (1+\rho)\epsilon_{t+1}.$$

Therefore

$$Std[\, Y_{t+1} |\text{ no sharing }] = \sigma\sqrt{\rho^2 + (1+\rho)^2}. \tag{2}$$

On the other hand, if demand information is shared with the manufacturer, the latter sees the value of D_t, then we have from (1)

$$Std[\, Y_{t+1} |\text{ sharing }] = \sigma(1+\rho). \tag{3}$$

From (2) to (3), we see the reduction in the manufacturer's costs as a result of information sharing. (Recall that the manufacturer's costs are proportional to the standard deviation of its leadtime demand, i.e., Y_{t+1}.) The savings can be significant, as Lee et al. have shown by analytical and numerical results.

In a note commenting on Lee et al., Raghunathan (2001) argues that the manufacturer can do much better in the case without information sharing. The idea is that the manufacturer can use its information about the retailer's order history to greatly sharpen its demand forecast. Let's see how this works. From $Y_t = D_t + \rho(D_t - D_{t-1})$, we have

$$D_t = \frac{1}{1+\rho}Y_t + \frac{\rho}{1+\rho}D_{t-1}.$$

Applying the above equation repeatedly, we have

$$D_t = \frac{1}{1+\rho}\sum_{i=1}^{t-1}(\frac{\rho}{1+\rho})^i Y_{t+1-i} + (\frac{\rho}{1+\rho})^t D_0$$

where it is assumed $D_0 = d + \epsilon_0$. Plugging the above into (1),

$$Y_{t+1} = (1+\rho)d + \frac{\rho^2}{1+\rho}\sum_{i=1}^{t-1}(\frac{\rho}{1+\rho})^i Y_{t+1-i} + \frac{\rho^{t+2}}{(1+\rho)^t}D_0 + (1+\rho)\epsilon_{t+1}$$

with

$$Std[\, Y_{t+1} |\text{ no sharing }] = \sigma\sqrt{\frac{\rho^{2t+4}}{(1+\rho)^{2t}} + (1+\rho)^2}.$$

It can be shown that the above expression is less than (2), suggesting that the value of information is less than what is reported in Lee et al. Moreover, as $t \to \infty$, the benefits effectively disappear.

Cachon and Fisher (2000) provide a model to quantify the value of downstream inventory information in a one-warehouse multi-retailer system. When the warehouse has access to real-time inventory status at the retailers as opposed to just retailer orders, it can make better ordering and allocation decisions. The supply chain benefits. But by how much? (We choose not to refer to the warehouse as a supplier, for the view taken here is that of a central planner.)[6]

The model consists of one warehouse and multiple, identical retailers. The periodic customer demands at the retailers are i.i.d., both across retailers and across time periods. The retailers replenish their inventories by ordering from the warehouse, who in turn orders from an external source with unlimited inventory. Complete backlogging is assumed at both the retail and the warehouse level. The replenishment leadtime at the warehouse is constant, so are the transportation leadtimes from the warehouse to the retailers. Linear inventory holding costs are incurred at the warehouse and the retailers, and linear penalty costs are incurred at the retailers for customer backorders. The objective is to minimize the long-run average system-wide holding and backorder costs (i.e., the central planner's view).

Inventory transfers from the warehouse to the retailers are restricted to be integer multiples of Q_r, an exogenously given base quantity. Similarly, orders by the warehouse (to the external source) must be integer multiples of Q_sQ_r, for some positive integer Q_s, another given parameter. The decisions for the retailers are when to place an order with the warehouse, and how many batches (each of size Q_r) to order, and the decisions for the warehouse are when to place an order with the outside source, and how many sets of batches to order (each set consists of Q_s batches, each of size Q_r).

In the scenario with traditional information sharing, the warehouse only observes the retailers' orders. Thus a replenishment policy can only be based on local information. Specifically, each retailer follows an (R_r, nQ_r) policy, i.e., whenever its inventory position (its outstanding orders, in transit or backlogged at the warehouse, plus its on-hand inventory minus its customer backorders) falls to R_r or below, order a minimum integer multiple of Q_r to increase its inventory position to above R_r. Similarly, the warehouse follows an (R_s, nQ_s) policy: whenever its inventory position (orders in transit plus on-hand inventory minus backlogged retailer orders) falls to R_sQ_r or below, order an integer multiple of Q_sQ_r units. The decision variables are the reorder points R_r and R_s.

When the warehouse is unable to satisfy every retailer's order in a period, it follows an allocation policy. It is called a batch priority policy that works as follows. Suppose a retailer orders b batches in a period. Then, the first batch in the order is assigned priority b, the second batch is assigned priority $b-1$, etc. All batches ordered in a period (by all retailers) are placed in a shipment queue,

[6] There is a large body of literature on one-warehouse multi-retailer systems. One way to categorize this literature is by looking at whether or not there are economies of scale in transferring inventory from one location to another. If the answer is no, then the focus is on the so-called one-for-one replenishment policies. The key references in this area are: for continuous-time models, Sherbrooke (1968), Simon (1971), Graves (1985), Axsäter (1990), Svoronos and Zipkin (1991), Forsberg (1995), and Graves (1996); and for discrete-time models, Eppen and Schrage (1981), Federgruen and Zipkin (1984a,b), Jackson (1988), and Diks and de Kok (1998). If there are economies of scale, then a batch-transfer policy makes more sense. The key references here are: for continuous-time models, Deuermeyer and Schwarz (1981), Moinzadeh and Lee (1986), Lee and Moinzadeh (1987a,b), Svoronos and Zipkin (1988), Axsäter (1993b, 1997, 1998, 2000), and Chen and Zheng (1997); and for discrete-time models, Aviv and Federgruen (1998), Chen and Samroengraja (1999, 2000), and Cachon (2001). For comprehensive reviews on the above literature, see Axsäter (1993a), Federgruen (1993), and Chapter 10 of this volume by Sven Axsäter. Most of the replenishment policies studied are based on local inventory information, and only a couple use centralized demand/inventory information. The objectives of these papers are typically to show how to determine the system-wide costs for a given class of policies. The desire to understand the value of demand/inventory information appeared only recently.

with the batch having the highest priority enters the queue first. (The rationale for this allocation policy is that a retailer ordering the most batches in a period is, naturally, considered to have the highest need for inventory.) When multiple batches have the same priority, they enter the queue in a random sequence. A shipment queue is maintained for each period. The retailers' orders are satisfied in the order in which the shipment queues are created and within each shipment queue, on a first-in first-out basis. Notice that the warehouse's stock allocation is based on the retailers' "needs" at the time the orders are placed.

The second scenario, called full information sharing, is where the warehouse has access to the retailers' inventory status on a real-time basis. In this case, the retailers continue to use the (R_r, nQ_r) policy described earlier. The warehouse, however, uses more sophisticated rules for ordering and allocation. The exact policy is complicated. The idea behind the new ordering policy is that the warehouse should perform some sort of cost-benefit analysis for each set of batches added to an order, with the cost being additional holding cost at the warehouse and the benefit being less delay for the retailers' orders. On the other hand, with immediate access to retailers' inventory status, the warehouse can allocate inventory (to satisfy backlogged retailer orders) based on the retailers' needs at the time of shipment.

By comparing the system-wide costs under traditional and full information sharing, one obtains a measure of the value of downstream inventory information. In a numerical study with 768 examples, it is found that information sharing reduces supply chain costs by 2.2% on average, with the maximum at 12%. [7]

Aviv and Federgruen (1998) consider a supply chain model consisting of a supplier and multiple retailers. The members of the supply chain are independent firms. In this decentralized setting, they attempt to characterize the value of sales information, which is defined to be the reduction in supply chain-wide costs if the supplier has access to real-time sales data at the retail level.[8] They then proceed to consider the impact of a vendor managed inventory (VMI) program, which comes with real-time information sharing and puts the supplier in the position of a central planner for the supply chain. Their main conclusions are based on three models: a decentralized model without information sharing, a decentralized model with information sharing, and a centralized model with information sharing.

We begin with the base model. There are J retailers. Customer demands are stochastic and occur at the retail sites only. The retailers monitor their inventories periodically. The demand processes at the retailers are independent. Demands in different periods at the same retailer are i.i.d. according to a retailer-specific distribution. When demand exceeds on-hand inventory at a retailer, the excess demand is backlogged. The retailers replenish their inventories from the supplier, who in turn replenishes its own inventory through production. The transportation leadtimes from the supplier to the retailers are constant but may be retailer-specific. The production leadtime at the supplier is also constant. The production quantity that the supplier can initiate in a period is subject to a constant capacity constraint. Each retailer incurs a fixed cost for each order it places with the

[7]Cachon and Fisher also provide a lower bound on the system-wide costs under full information, and compare that with the costs under traditional information sharing. This does not change the picture on the value of information in any significant way, meaning the proposed full-information policy is near-optimal. Moreover, they show, again via numerical examples, that significant savings can be had if the leadtimes or batch sizes (due to fixed ordering costs) are reduced, which may be expected from better information-linkup. Similar findings have been reported in Chen (1998b).

[8]This paper therefore deviates from the mainstream approach of quantifying the value of information in centralized models.

supplier, linear inventory holding costs, and linear penalty costs for customer backorders. These cost parameters are stationary over time, but they can be retailer-specific. The supplier incurs linear holding costs for its on-hand inventory and linear penalty costs for backorders of retailer orders. This latter cost component is a contract parameter, which is given exogenously and represents a revenue for the retailers. There are no fixed costs for initiating a production run at the supplier.

The replenishment policies at the retailers are of the (m, β) type, whereby the retailer reviews its inventory position every m periods and places an order to increase it to β. The values of the policy parameters can be retailer-specific, with (m_j, β_j) for retailer j, $j = 1, \cdots, J$. Let M be the least common multiple of $m_1, \cdots, m_J$. Therefore, the joint order process of the retailers regenerates after a grand replenishment cycle of M periods. In general, the replenishment cycles of the retailers are not coordinated. Aviv and Federgruen consider two extreme arrangements in this regard. One is called "peaked," in which all retailers order at the beginning of a grand replenishment cycle, and the other "staggered," in which the retailer cycles are spread out to achieve a smooth order process for the supplier. (The staggered pattern is clearly defined if the retailers are identical. Otherwise, one needs to spell out what is meant by "smooth." Staggered policies have been proposed and studied by Chen and Samroengraja (2000) in one-warehouse multi-retailer systems.)

Given that the demand process at the supplier is cyclical, it is reasonable to expect that the supplier's production policy is also cyclical. Aviv and Federgruen assume that the supplier follows a modified base-stock policy with cyclical order-up-to levels, whereby the supplier initiates a production run, subject to the capacity constraint, in period t to increase its inventory position to β^m, if period t is the mth period in the grand cycle, $m = 0, 1, \cdots, M - 1$. [9] In the event that the supplier cannot satisfy all retailers' orders in a period, an allocation mechanism is given that is based on some measure of expected needs of the retailers. A shipment can be sent to a retailer even though it is not the retailer's ordering period. (Recall that the retailers order intermittently.) Of course, this occurs only when the supplier backlogs an order (or part of it) from the retailer and fills it in a subsequent period.

The firms minimize their own long-run average costs in a noncooperative fashion. Ideally, the solution to this noncooperative game can be obtained by using some established equilibrium concept. Since this is intractable, Aviv and Federgruen take a two-step approach: first, the retailers optimize assuming the supplier has ample stock, and then, given the retailers' decisions, the supplier optimizes. This completes our description of the base model.

The second model retains the above decentralized structure but assumes that the supplier observes the realized demands at the retail sites immediately. This information allows the supplier to better anticipate the orders that the retailers are going to place in future periods. As a result, the supplier can use a state-dependent, modified base-stock policy, where the state now includes not only where in a grand cycle the current period is but also a summary of relevant sales information from the retail sites.

Finally, the third model assumes that a VMI program is in place, which provides the supplier with immediate access to the sales data at the retail sites and gives the supplier the rights to decide when and how much to ship to each retailer. It is assumed that the VMI contract is such that it is in the supplier's interests to minimize the total costs in the supply chain. Aviv and Federgruen propose a heuristic method to solve this centralized planning problem.

[9] They also consider a policy with a constant order-up-to level. For proofs of the optimality of the cyclical base-stock policies in single-location settings, see Aviv and Federgruen (1997) and Kapuscinski and Tayur (1998).

A numerical study shows that the average improvement in supply chain costs from the first to the second model is around 2%, with a range from 0% to 5%. Most of these savings accrue to the supplier. From the second model to the third, the average improvement is 4.7%, with a range from 0.4% to 9.5%. It is also found that the value of information sharing and VMI increases, as the degree of heterogeneity among the retailers increases, the leadtimes become longer, or the capacity becomes tighter. Finally, the system tends to perform better with staggered retailer cycles rather than the peaked pattern.

One of the key drivers for production-inventory planning decisions is demand forecast. In any given period, the firm determines a set of predictions for the demands in future periods based on its information about the operating environment and planned activities. As time progresses and new information becomes available, the firm revises its demand forecast. From the standpoint of production-inventory planning, an important question is how to integrate the evolving demand forecast into the planning decisions. Below, we summarize several papers that address this question.

Gullu (1997) studies a two-echelon supply chain consisting of a central depot and N retailers. The depot serves as a transshipment center where an order arriving at the depot from an outside supplier is immediately allocated among the retailers. (Thus the depot does not hold inventory.) Customer demand arises only at the retailers, with unsatisfied demand fully backlogged. The objective is to determine a depot replenishment/allocation policy that minimizes the system-wide costs. A unique feature of the model is that each retailer maintains a vector of demand forecasts for a number of future periods, and this vector is updated from one period to the next. Gullu considers two models, one that utilizes the demand forecasts in the replenishment/allocation decision and the other that ignores the forecasts. By comparing the two models, one sees the value of demand information (contained in the forecasts).

The evolution of demand forecasts is described by the martingale model of forecast evolution (MMFE). [10] Let D_t^j be retailer j's demand forecasts for periods $t, t+1, \cdots$ at the end of period t, $j = 1, \cdots, N$. That is,

$$D_t^j = (d_{t,t}^j, d_{t,t+1}^j, \cdots)$$

where $d_{t,t}^j$ is the realized demand for period t (thus not really a forecast) and $d_{t,t+k}^j$, $k \geq 1$, is the retailer's forecast, made at the end of period t, for the demand in period $t+k$. In the additive model, D_t^j is obtained by adding an error term (or adjustment) to each relevant component of D_{t-1}^j, i.e.,

$$\begin{aligned} d_{t,t}^j &= d_{t-1,t}^j + \epsilon_{t,1}^j \\ d_{t,t+1}^j &= d_{t-1,t+1}^j + \epsilon_{t,2}^j \\ &\vdots \end{aligned}$$

Let $\bar{\epsilon}_t^j = (\epsilon_{t,1}^j, \epsilon_{t,2}^j, \cdots)$ and $\bar{\epsilon}_t = (\bar{\epsilon}_t^1, \cdots, \bar{\epsilon}_t^N)$. Gullu assumes that $\epsilon_{t,k}^j = 0$ for all $k > M$ for some positive integer M, for all t and all j. In other words, the new information collected during period t only affects the demand forecasts in M periods (i.e., the current period and the next $M-1$ periods). Moreover, it is assumed that $\bar{\epsilon}_t$, for all t, are independent and have the same multi-variate normal distribution with zero mean. It is, however, possible that for a given t, the different components of $\bar{\epsilon}_t$ are correlated. This allows for modeling of demand correlation over time and across retailers.

[10]For the development of the MMFE model, see Hausman (1969), Graves et al. (1986, 1998), and Heath and Jackson (1994). Hausman (1969) and Heath and Jackson (1994) also consider a multiplicative model.

Finally, the initial forecast for the demand in a period that is more than M periods away is a constant, i.e.

$$d^j_{t,t+k} = \mu^j, \ \ \forall t \text{ and } \forall k \geq M.$$

(Thus μ^j is the mean demand per period at retailer j.) Consequently,

$$d^j_{t,t} = d^j_{t-1,t} + \epsilon^j_{t,1} = \cdots = \mu^j + \sum_{i=1}^{M} \epsilon^j_{t-i+1,i}$$

Letting $\sigma^2_{j,i}$ be the variance of $\epsilon^j_{t,i}$, we have

$$Var[d^j_{t,t}] = \sum_{i=1}^{M} \sigma^2_{j,i}$$

Note that at the end of period $t-1$, the conditional variance of $d^j_{t,t}$ given $d^j_{t-1,t}$ is only $\sigma^2_{j,1}$. In other words, as the demand forecast for a fixed period is successively updated, the variance for the demand in that period is successively reduced. This reduction in demand uncertainty in turn leads to improvement in supply chain performance.[11]

Under the above demand model, Gullu considers two scenarios depending on whether or not the demand forecasts are used in the depot's allocation decision. The depot's replenishment policy is the order-up-to S policy, i.e., in each period, the depot places an order with the outside supplier to increase the system-wide inventory position to the constant level S.[12] The key analytical results are that the use of demand forecasts leads to lower system-wide costs and if and only if the backorder penalty cost rate is higher than the holding cost rate (identical cost rates are assumed across retailers), a lower system-wide inventory position. (This latter result is well known for the newsvendor model with normal demand.) There are also various asymptotic results for some special cases, which we omit.

Toktay and Wein (2001) consider a single-item, single-stage, capacitated production system with an MMFE demand process. Random demand for the item arises in each period. Demand is satisfied from the finished-goods inventory, which is replenished by a production system with capacity C_t in period t. That is, the system can produce up to C_t units of the product in period t, with C_t in different periods being i.i.d. normal random variables with mean μ and variance σ^2_C. If demand exceeds the finished-goods inventory, the excess demand is fully backlogged. Let I_t be the (finished goods) inventory level at the end of period t. The system incurs holding and backorder costs equal to $(hI^+_t + bI^-_t)$ in period t, where h and b are the holding and backorder cost rates. Let P_t be the production quantity in period t. Thus $P_t = \min\{Q_{t-1}, C_t\}$, where Q_{t-1} is the number of production orders waiting to be processed at the end of period $t-1$. At the end of each period t, R_t new orders are released to the production system. Thus $Q_t = Q_{t-1} - P_t + R_t$. The objective is to find a release policy $\{R_t\}$ to minimize the expected steady-state holding and backorder costs.[13]

Let D_t be the demand in period t. The demand process is stationary with $E[D_t] = \lambda$. (Thus we need $\mu > \lambda$ for stability.) Let $D_{t,t+i}$ be the forecast for D_{t+i} determined at the end of period t, $i \geq 0$.

[11] Updates of demand forecasts do not always make them more accurate, see Cattani and Hausman (2000) for both empirical and theoretical evidence.

[12] It is possible that the supply chain's performance can be improved if the depot's replenishment decision takes into account the demand forecasts at the retail level. This should be investigated. If you are familiar with Eppen and Schrage (1981), then you can see that the Gullu model is basically the Eppen-Schrage model with forecast evolution. The analysis is also similar to Eppen-Schrage's.

[13] What queueing folks call a release policy is called a replenishment policy by inventory folks.

Thus $D_{t,t}$ is the realized demand in period t. It is assumed that nontrivial forecasts are available only for the next H periods, i.e., $D_{t,t+i} = \lambda$ for all $i > H$. Define $\epsilon_{t,t+i} = D_{t,t+i} - D_{t-1,t+i}$, $i \geq 0$. It is then clear that $\epsilon_{t,t+i} = 0$ for al $i > H$. Thus $\epsilon_t = (\epsilon_{t,t}, \epsilon_{t,t+1}, \cdots, \epsilon_{t,t+H})$ is the forecast update vector whose value is observed at the end of period t. The forecast update vectors (for different periods) are assumed to be i.i.d. normal random variables with zero mean.

Toktay and Wein consider two classes of release policies: one ignores the forecast information, and the other utilizes it. In the former case, $R_t = D_t$. Under the initial condition that $Q_0 = 0$ and $I_0 = s_m$, the proposed release policy leads to $Q_t + I_t = s_m$ for all t.[14] To integrate the demand forecasts into the release policy, consider

$$R_t = \sum_{i=0}^{H} D_{t,t+i} - \sum_{i=0}^{H-1} D_{t-1,t+i} = \sum_{i=0}^{H-1} \epsilon_{t,t+i} + D_{t,t+H} = \sum_{i=0}^{H} \epsilon_{t,t+i} + \lambda.$$

Under this release policy and with proper initial conditions, one can show that for all t

$$Q_t + I_t - \sum_{i=1}^{H} D_{t,t+i} = s_H$$

for some constant s_H which can be controlled by setting the initial inventory level.[15] A key finding of the paper is the observation that excess production capacity, as measured by $\mu - \lambda$, and demand information, as contained in the demand forecasts, are substitutes. (Other studies on capacitated problems with forecast evolution include Gullu (1996) and Gallego and Toktay (1999).)

Aviv (2001) considers a supply chain model with one retailer and one supplier. Customer demand arises at the retailer, who replenishes its inventory from the supplier, who in turn orders from an outside source with ample stock. The two members of the supply chain independently forecast the customer demands in future periods and periodically adjust their forecasts as more information becomes available. The retailer and the supplier are modeled as a team in the sense that they share a common objective to minimize the system-wide costs, but they do not necessarily share their demand forecasts. Aviv studies the following three scenarios. In scenario one, the two members neither share their demand forecasts nor use their own demand forecasts in making replenishment decisions. In scenario two, they still do not share their demand forecasts, but now they each integrate their own forecasts in their replenishment decisions. In scenario three, they share their demand forecasts and use the shared information in their replenishment decisions.

Aviv uses an MMFE demand model. The demand in period t, d_t, is the sum of a constant and a stream of random variables representing adjustments to the forecast of d_t made at different times leading up to period t. Specifically,

$$d_t = \mu + \epsilon_t + \sum_{i=0}^{\infty} (\epsilon^r_{t,i} + \epsilon^s_{t,i})$$

[14] In traditional inventory lingo, Q_t is the outstanding orders, while I_t is the inventory level. The sum of the two is the inventory position. The proposed policy is thus a base-stock policy whereby the inventory position is maintained at a constant level.

[15] Again one can draw some connection to inventory theory here. As noted above $Q_t + I_t$ is the inventory position at the end of period t. So the second release policy corresponds to a modified base-stock policy that is based on an "adjusted inventory position." The adjustment is the total forecasted demand in the next H periods. From inventory theory, the "optimal" adjustment should be based on the demand during the replenishment "leadtime." The problem is that there is no leadtime here, only capacity. The proposed release policy seems to draw an equivalence between "leadtime" and the forecast horizon in a capacitated production system. If capacity is tight, then the "leadtime" should be long, and the opposite holds if capacity is ample. But that has little to do with the forecast horizon.

where μ is a constant, $\{\epsilon_t\}_{t\geq 1}$ are i.i.d. normal random variables, the components of the vector $\{(\epsilon^r_{t,i}, \epsilon^s_{t,i})\}_{i=0}^{\infty}$ are independent and each a bi-variate normal and the vectors (for different t) are i.i.d., and $\{\epsilon_t\}_{t\geq 1}$ are independent of $\{(\epsilon^r_{t,i}, \epsilon^s_{t,i})\}_{t\geq 1, i\geq 0}$. All the random variables have zero mean. (Different notation is used in Aviv.) As a result, $\{d_t\}$ is a sequence of i.i.d. normal random variables with mean μ. At the beginning of period τ, for any τ, the retailer privately observes the vector $\{\epsilon^r_{t',t'-\tau}\}_{t'\geq\tau}$, and the supplier privately observes the vector $\{\epsilon^s_{t',t'-\tau}\}_{t'\geq\tau}$. Therefore, by the beginning of period $t-k$, $k \geq 0$, the retailer has observed the value of $\sum_{i=k}^{\infty} \epsilon^r_{t,i}$ whereas the supplier has observed the value of $\sum_{i=k}^{\infty} \epsilon^s_{t,i}$. It is easy to see that as they get closer and closer to period t, i.e., as k decreases, the supply chain members have less and less uncertainty about d_t (or better and better forecast for d_t). In scenario three, the supply chain members share their private information. This enables them to improve (by unifying) their demand forecasts.

In a numerical study, Aviv found that integrating the forecast updates in the replenishment decisions reduces, on average, the supply chain costs by 11%, and information sharing between the retailer and the supplier brings in an additional reduction of 10%.

It is worthwhile to note that demand forecasts can take other forms with different patterns of evolution. For example, a parameter of the demand distribution may be unknown. Beginning with a prior distribution for the unknown parameter, one can sharpen the estimate of the parameter after each observation of demand. The production/inventory decisions can be made to dynamically reflect the new information that becomes available as time progresses. Alternatively, demands in different periods may be correlated and the data on early sales can be used to update the forecasts for the later sales. The following are additional papers that incorporate adjustments of demand forecasts: Scarf (1959, 1960), Iglehart (1964), Murray and Silver (1966), Hausman and Peterson (1972), Johnson and Thompson (1975), Azoury and Miller (1984), Azoury (1985), Bitran et al. (1986), Miller (1986), Bradford and Sugrue (1990), Lovejoy (1990, 1992), Matsuo (1990), Fisher and Raman (1996), Eppen and Iyer (1997a,b), Sobel (1997), Barnes-Schuster et al. (1998), Brown and Lee (1998), Lariviere and Porteus (1999), Dong and Lee (2000), Donohue (2000), Milner and Kouvelis (2001), and Ding et al. (2002). On the other hand, the demand process can be modulated by an exogenous Markov chain; the state of the exogenous Markov chain determines the current period's demand distribution. For inventory models with Markov-modulated demands, see Song and Zipkin (1992, 1993, 1996a), Sethi and Cheng (1997), Chen and Song (2001), and Muharremoglu and Tsitsiklis (2001).

Advance Warnings of Customer Demands

When members of a supply chain share information, no new information is created; only existing information moves from one place to another. In some situations, however, customers can, and willing to, provide advance warnings of their demands. These warnings represent new information for the supply chain. And the question is how to exploit such information.

Hariharan and Zipkin (1995) study inventory models where customers provide advance warnings of their demands. Customer orders arise randomly. Each order comes with a due date, a future time when the customer wishes to receive the goods ordered. They call the time from a customer's order to its due date the *demand leadtime*. The customer does not want to receive delivery before the due date. Deliveries after due dates are possible, but undesirable. They show that demand leadtimes are the opposite of supply leadtimes in terms of their impact on the system performance.

A simple model illustrates the basic idea. Suppose customer orders arrive according to a simple Poisson process. Each customer orders a single unit. The demand leadtime is a constant l, i.e., an order at time t calls for a "demand" at time $t+l$. Demand is satisfied from on-hand inventory, with complete backlogging. Inventory is replenished from an outside source with ample stock, and the supply leadtime is a constant L. There are no economies of scale in ordering.

If $l \geq L$, then one can satisfy all customer demands perfectly without holding any inventory. Here is how to achieve that. Whenever a customer order arrives, wait $(l-L)$ units of time before placing an order for one unit with the outside supplier. This replenishment unit will arrive just in time to satisfy the customer's demand (at time $t+l$).

Now suppose $l < L$. From basic inventory theory, we know that the inventory position at time t should "cover" the total demand during the supply leadtime (i.e., the leadtime demand). Note that at time t, the total demand in the interval $(t, t+l)$ is already known due to advance ordering, whereas the demand in $(t+l, t+L)$ remains unknown (it is a Poisson random variable with mean $\lambda(L-l)$ where λ is the arrival rate of customer orders). Let d_t be the known part, and D_t the unknown portion. The leadtime demand is $D_t^L = d_t + D_t$. Therefore, the inventory position at time t should consist of d_t and a buffer inventory S for protection against the uncertain part of the leadtime demand D_t. The inventory level at time $t+L$ is $(d_t + S) - D_t^L = S - D_t$. Therefore, the expected holding and backorder cost rate at time $t+L$ can be written as

$$E[h(S-D_t)^+ + b(S-D_t)^-]$$

where h and b are the holding and backorder cost rates. Let S^* be the S value that minimizes this cost expression. If we set the inventory position at time t to $d_t + S^*$, then we know that the expected holding and backorder costs one supply leadtime later are minimized. If this inventory position can be achieved for all t, then the system's long-run average costs are minimized and we have an optimal policy. Here is a proof. Assume at $t = 0$, the inventory position is $d_0 + S^*$. (If it is lower than this target level, order enough to make up the shortfall; otherwise, just wait until the inventory position at some time τ equals $d_\tau + S^*$.) Then, whenever a customer order arrives, order one unit from the outside source. This is just like the one-for-one replenishment policy used in the conventional system without advance ordering, with a caveat that replenishment orders are based on customer orders not customer demands. It is a simple matter to see that the inventory position at any time coincides with the ideal target. The key point of the above exercise is that a system with demand leadtime l and supply leadtime L is essentially the same as the conventional system with supply leadtime $(L-l)$. Thus, demand leadtime is the opposite of supply leadtime, an elegant characterization of the value of (one type of) demand information.

Hariharan and Zipkin also study advance ordering in other models, where the supply leadtime is stochastic or where the replenishment process consists of multiple stages. We omit the details.

One limitation of the Hariharan-Zipkin construct is that all customers come with the same demand leadtime. This assumption is relaxed in Chen (2001a), the customer population is divided into M segments. The customers from segment m are homogeneous and provide a common demand leadtime l_m, $m = 1, \cdots, M$. In this multi-segment case, it is still rather straightforward to characterize the value of advance ordering, even in a multi-stage serial inventory system. But the main objective of Chen (2001a) is to study the incentives required by the customers in order for them to willingly offer advance warnings of their demands and how these incentives can be traded off against the benefits of demand information embodied in the advance orders. We will review Chen (2001a) in greater detail in §3.1.

Gallego and Özer (2001) provide a discrete-time version of the above multi-segment model of advance ordering. Time is divided into periods. In each period t, a demand vector is observed:

$\vec{D}_t = (D_{t,t}, \cdots, D_{t,t+N})$, where $D_{t,s}$ is orders placed by customers in period t for deliveries in period s and N is a constant (positive integer) and is referred to as the information horizon.[16] (Therefore, the customer population effectively consists of $N+1$ segments.) For this demand process, Gallego and Özer prove optimal policies in a single-location model with or without fixed ordering costs. They consider multiple scenarios where the planning horizon can be finite or infinite and the cost parameters can be nonstationary. The main result is that if there are fixed order costs, the optimal policy is a state-dependent (s,S) policy; otherwise, the optimal policy is a state-dependent base-stock policy. But what is the state? Define for any $s \geq t$

$$O_{t,s} = \sum_{\tau=s-N}^{t-1} D_{\tau,s}$$

which represents what we know at the beginning of period t about the demand in period s. As in any inventory model, we care about the total demand during the supply leadtime, which is assumed to be a constant L. Define

$$O_t^L = \sum_{s=t}^{t+L} O_{t,s}$$

which is what we, standing at the beginning of period t, know about the future demands in periods $t, t+1, \cdots, t+L$. (Therefore, the leadtime demand is total demand over $L+1$ periods; the extra period is simply due to the convention that orders are placed at the beginning of a period and costs are assessed at the end of a period.) The modified inventory position at the beginning of period t is simply the inventory level (on-hand minus backorders) plus outstanding orders minus O_t^L. (Therefore, the known part of the leadtime demand has been taken out of the inventory position. This is just for control purposes, of course, as we will see.) The state of the inventory system consists of the above modified inventory position plus what we know about the demands beyond the supply leadtime, i.e.

$$\vec{O}_t = (O_{t,t+L+1}, \cdots, O_{t,t+N-1}).$$

The optimal (s,S) policy has control parameters that are dependent on $\vec{O}_t$ and operates based on the modified inventory position. That is, at the beginning of period t, if the modified inventory position is at or below $s(\vec{O}_t)$, order to increase it to $S(\vec{O}_t)$; otherwise, do nothing. When there are no fixed order costs, the action is simply ordering to increase the modified inventory position up to $S(\vec{O}_t)$ every period. For this latter case, and when the problem is stationary, the base-stock level no longer depends on $\vec{O}_t$. This makes intuitive sense.

Other related studies include Gallego and Özer (2000), Özer (2000), Karaesmen et al. (2001), and Özer and Wei (2001). These papers show how advance demand information can be used to improve performance in various production/distribution systems with or without capacity constraints.

A mirror image of customers providing advance demand information is the decision maker postponing a decision until after customers have placed their orders. This is, e.g., the case when a firm switches from a make-to-stock regime to a make-to-order regime. The postponement reduces

[16] The reader may notice that this demand model, where customers place orders in advance of their requirements, resembles the MMFE model considered earlier. In fact, strictly speaking, the demand model with advance orders can be considered a special case of the forecast evolution model. The only, perhaps superficial, distinction is that the updates in the advance-orders model represent actual customer orders, whereas the updates in the MMFE don't have to be. Moreover, the advance-orders model assumes no order cancellation (i.e., the updates are always nonnegative). No such assumption has been detected under the MMFE framework.

the uncertainty confronting the decision maker, improving the quality of the decision and thus performance. For more on the impact of the postponement of operations decisions, see the cases of Benetton (by Signorelli and Heskett, 1984) and Hewlett-Packard (by Kopczak and Lee, 1994) and the papers by Lee and Tang (1998) and Van Mieghem and Dada (1999) and the references therein. Refer to Chapter 5 of this volume by Hau L. Lee and Jayashankar M. Swaminathan for extensive discussions on postponement strategies.

The Bullwhip Phenomenon

The bullwhip effect refers to a phenomenon where the replenishment orders generated by a stage in a supply chain exhibit more volatility than the demand the stage faces. Recently there has been a flurry of activities on the bullwhip effect. We review this part of the literature here mainly because information sharing (e.g., sharing of customer demand information) is often suggested to combat the undesirable effect.

Many economists have studied the bullwhip phenomenon; they are interested in it because empirical observations refute a conventional wisdom that inventory smoothes production. A firm carries inventory, the conventional wisdom goes, which serves as a buffer to smooth out the peaks and valleys of demand. This in turn generates a relatively stable environment for production. So production should be smoother than demand. Unfortunately, industry data point to the other way. Why? Possible explanations include: the use of (s,S) type of replenishment policies, the presence of positive serial correlation in demand, etc. See Blinder (1982, 1986), Blanchard (1983), Caplin (1985), and Kahn (1987). Other explanations call for industrial dynamics and organizational behavior (Forrester 1961) and irrational behavior on the part of decision makers (Sterman 1989).

We focus on the operations management literature, which has provided some new insights into the bullwhip phenomenon. The general approach in this literature is to first specify the environment (e.g., revenue/cost structure, characteristics of the demand process, etc.) in which a supply chain member operates, and then show that when the supply chain member optimizes its own performance, it generates orders that are more volatile than the demand process it faces. The implicit message here is that the supply chain member should not be blamed for the bullwhip effect; it is the environment that has created the observed behavior. Sometimes, one can change the environment, with potential benefits for some members of the supply chain or the supply chain as a whole.

Lee et al. (1997a,b) exemplify the above approach. They have identified four causes for the bullwhip effect. The first is the demand characteristics. In a single-location inventory model with a positively correlated demand process, they show that the optimal policy that minimizes, say, the retailer's costs leads to variance amplification. There is an intuitive explanation. When the retailer observes a low demand, he takes that as a signal of low future demands as well and places an order that reflects that lowered forecast. Conversely, a high demand suggests to him that the demands in the future periods are likely to be high as well. He then places a large order based on that outlook. In sum, due to the positive correlation between the demands in different periods, the orders placed by the retailer exhibit larger swings than the demands he observes. The second cause is the possibility of supply shortage. To see the intuition, consider a supply chain with one supplier, whose capacity fluctuates over time, and multiple retailers. In periods when the supplier's capacity is likely to be insufficient, the retailers—engaged in a rationing game to secure an adequate supply for themselves—place large orders (larger than what they would order if there were no capacity shortage). Suppose a retailer faces a deterministic demand stream, say, 4 units per period. So he would order 4 units when no capacity shortage is expected, and order more otherwise. Clearly, when the capacity fluctuates over time, the order stream has a larger variance than the demand stream (which has zero variance). The third cause is economies of scale in placing orders. When there is a fixed cost in placing an order, it makes sense to order once every few periods. This order batching leads to the bullwhip effect. The fourth cause is fluctuating purchase costs. In periods when the supply is cheap, you want to buy a lot and stockpile, whereas in periods when supply is expensive, you wait. It is easy to see that the high-low prices encourage extreme orders. In order to dampen the bullwhip effect, one has to attack the root causes. Lee et al. have described several industry initiatives that do just that.

Below, we describe a supply chain model that has been used to show that a quasi-optimal operating policy amplifies the order variance.[17] Graves (1999) considers a supply chain model with a nonstationary demand process. The demand process is an autoregressive integrated moving average (ARIMA) process:

$$\begin{aligned} d_1 &= \mu + \epsilon_1 \\ d_t &= d_{t-1} - (1-\alpha)\epsilon_{t-1} + \epsilon_t, \quad t = 2, 3, \cdots \end{aligned}$$

where d_t is the demand in period t, α and μ are known constants, and ϵ_t are i.i.d. normal random variables with mean zero and variance σ^2. It is assumed that $0 \leq \alpha \leq 1$. [18] From the above description of the demand process, one can write

$$d_t = \epsilon_t + \alpha\epsilon_{t-1} + \cdots + \alpha\epsilon_1 + \mu.$$

Note that each period, there is a shift in the mean of the demand process: the random shock ϵ_t shifts the mean of the demand process by $\alpha\epsilon_t$, starting from period $t+1$. Therefore, each random shock has a permanent effect on the demand process. Note that $\alpha = 0$ corresponds to an i.i.d. demand process, whereas $\alpha = 1$ is a random walk. In general, a larger α means the process depends more on the most recent demand realization.

For the above demand process, a first-order exponential-weighted moving average provides a minimum mean square forecast. Define

$$\begin{aligned} F_1 &= \mu \\ F_{t+1} &= \alpha d_t + (1-\alpha)F_t, \quad t = 1, 2, \cdots \end{aligned}$$

where F_{t+1} is the forecast for the demand in period $t+1$, after observing the demand in period t. It is easy to verify that

$$d_t - F_t = \epsilon_t, \quad t = 1, 2, \cdots.$$

Therefore, the exponential-weighted moving average is an unbiased forecast with a minimum mean square error. Note that

$$F_{t+1} = d_{t+1} - \epsilon_{t+1} = \alpha\epsilon_t + \alpha\epsilon_{t-1} + \cdots + \alpha\epsilon_1 + \mu.$$

At the end of period t (after the realization of d_t or ϵ_t), the forecast for d_{t+i} for any $i \geq 1$ is equal to F_{t+1}.

[17] For other such studies, see, e.g., Chen et al. (2000), Ryan (1997), and Watson and Zheng (2001).

[18] The above demand process is also known as an integrated moving average (IMA) process of order (0,1,1), see Box et al. (1994).

Consider, for a moment, a single-location, single-item inventory system with the above demand process. Assume that when demand exceeds on-hand inventory, the excess demand is completely backlogged. Moreover, the replenishment leadtime is L periods, where L is a known integer. The events in each period are sequenced as follows: demand is realized, an order is placed, the order from L periods ago is received, and demand and backorders (if any) are filled from inventory. Consider a base-stock policy, with the order-up-to level for period t being

$$S_t = S_0 + LF_{t+1}$$

where S_0 is some constant. Note that LF_{t+1} is the forecast for the leadtime demand from period $t+1$ to period $t+L$. (And recall that the order-up-to level should cover the leadtime demand.) There is no optimality proof for this policy; but if orders are allowed to be negative, the policy is optimal. Let us assume that orders can be negative, and thus the order-up-to level for each period is reached exactly. We have the order quantity in period t,

$$q_t = d_t + (S_t - S_{t-1}) = d_t + L(F_{t+1} - F_t).$$

Note that the order quantity q_t reflects the most recent demand d_t as well as an update on the forecast for the total demand during the next L periods.[19]

Let x_t be the inventory level (on-hand inventory minus backorders) at the end of period t. Graves shows that under the above policy,

$$E[x_t] = S_0 + \mu, \quad Std[x_t] = \sigma\sqrt{\sum_{i=0}^{L-1}(1+i\alpha)^2}.$$

Note that when the leadtime demand is normally distributed, the minimum costs of the system are proportional to $Std[x_t]$, which can sometimes, especially when α is large, be a convex function of L. This is in sharp contrast with the traditional setting with i.i.d. demands, where the minimum costs are proportional to the square root of L. Our intuition is challenged, and it is because of the nonstationary demand process!

Another observation is that the variance of q_t is larger than the variance of d_t, given F_t. To see this, first recall that $d_t = F_t + \epsilon_t$. Thus, $Var[d_t|F_t] = \sigma^2$. On the other hand, since $F_{t+1} - F_t = \alpha\epsilon_t$, we have

$$q_t = d_t + L(F_{t+1} - F_t) = (F_t + \epsilon_t) + L\alpha\epsilon_t = F_t + (1 + L\alpha)\epsilon_t.$$

Therefore, $Var[q_t|F_t] = (1 + L\alpha)^2\sigma^2$. This shows that the variance of the order process exceeds the variance of the demand process, and this amplification increases with leadtime and α (larger α means a less stable demand process). Graves shows that the order process $\{q_t\}$ has the same characteristics as the demand process $\{d_t\}$. Thus one can easily extend the analysis to a multi-stage serial system and show that the order variance is further amplified upstream.

Although attention to the bullwhip effect can sometimes help us identify opportunities to improve supply chain performance, it is dangerous if we take as our goal the reduction or elimination of the bullwhip effect. This point is illustrated in a paper by Chen and Samroengraja (1999). They consider a supply chain model with one supplier and N identical retailers. The perspective is that

[19] We are not going to make a big deal out of negative orders here. If you continue to feel that negative orders are annoying, first consult Graves (1999) for further discussions on this and if that is still not enough, then you have a challenging, and potentially rewarding, task ahead of you.

of a central planner whose goal is to minimize the total cost in the supply chain. The supplier's production facility is subject to a capacity constraint, and transportation from the supplier to the retailers incurs fixed costs as well as variable costs. They consider two classes of replenishment strategies at the retail level. One is the staggered policy, whereby each retailer places an order to increase its inventory position to a constant base-stock level Y every T periods, and the reorder intervals of different retailers are staggered so as to smooth the aggregate demand process at the supply site. The other strategy is the (R, Q) policy, whereby each retailer orders Q units from the supplier as soon as its inventory position decreases to R. These two types of replenishment strategies are commonly used in practice when there are fixed ordering costs. The supplier replenishes its inventory through production; the production policy is a base-stock policy modified by a capacity constraint. Numerical examples show that although the (T, Y) policy gives a smoother demand process at the supply site, the (R, Q) policy often provides a lower system-wide cost. [20]

It is also interesting to note that discussions on the bullwhip effect can sometimes become confusing and pointless. Consider a supply chain with a manufacturer and a retailer. The retailer agrees to share its point-of-sale information with the manufacturer. For some unexplained reasons (historical?), the manufacturer has a quantity discount policy in place that charges a lower per-unit price for a larger order. Finally, the manufacturer can ship a retailer order in any way it desires so long as a certain service level is achieved at the retail site. In this decentralized model with information sharing and a specific contractual relationship, the manufacturer can plan its production based on the true demand information at the retail site. From the standpoint of the supply chain, what matters is the manufacturer's production quantities and the shipment quantities to the retailer. The retailer's orders don't matter very much; they exist largely for accounting purposes. There is nothing to worry about even if the retailer's orders are more volatile than the customer demands.

In summary, the existence of the bullwhip effect is only a characteristic of an operating policy, which reflects the economic forces underlying the supply chain and the experience and knowledge of the people who manage it. It is a symptom, not a problem. [21]

[20] Cachon (1999) also studies the impact of staggered ordering policies, which he calls scheduled or balanced ordering policies, on the supply chain performance. The setup is still the one-warehouse N-identical-retailer supply chain. The class of policies considered is that of (T, R, Q) policies: each retailer orders every T periods according to an (R, nQ) policy based on its own inventory position, and the reorder intervals of different retailers are staggered. Cachon provides an exact method to evaluate the supply chain costs under a (T, R, Q) policy as well as numerical examples that illustrate how the supply chain costs respond to changes in the parameters T and Q. Primary conclusions are that the staggering of retailer reorder intervals generally reduces the demand variance at the warehouse and that the combination of increasing T and decreasing Q is an effective way to decrease the total supply chain costs in systems with a small number of retailers and low customer demand variability. Although the general objective of Cachon (1999) coincides with that of Chen and Samroengraja (1999), i.e., to study the impact of variance-reduction policies on supply chain performance, the models are different (whether or not there is a capacity constraint at the warehouse), so are the approaches (the former focuses on a sensitivity analysis whereas the latter compares the optimal solutions from two classes of policies that offer different degrees of variance reduction).

[21] Any discussion of the bullwhip effect would be incomplete without mentioning the beer game, which is described in Sterman (1989) and some of the references therein. The game simulates a four-stage supply chain, consisting of a manufacturer, a distributor, a wholesaler, and a retailer. The demand at the retail site is 4 kegs of beer per period for the first several periods, and then jumps to 8 kegs per period for the rest of the game. The players, who manage the four supply-chain stages, do not know the demand process a priori. For several decades, the beer game has been a very effective tool to illustrate the bullwhip effect to an uncountable number of students in many countries. But it has a shortcoming: it merely demonstrates a phenomenon without offering any solutions. How should we play the game? Nobody knows the answer, a quite awkward situation especially in a classroom setting. It is easy to say what we should have done in hindsight, but that is not helpful to the supply chain's managers. In fact, it is quite possible that most strategies could be explained with a belief system that uses the past to predict the future in a particular

2.2 Upstream Information

So far, our discussions have been confined to the sharing of information coming from the demand side, i.e., an upstream supply-chain member's access to downstream information. We now turn to supply-side information. Interestingly, upstream information has received much less attention in the literature.

Cost Information

Chen (2001b) considers a procurement problem facing an industrial buyer. Given Q units of input, the buyer can generate profits $R(Q)$, an increasing and concave function. The buyer's net profit is therefore $R(Q)$ minus the purchase cost incurred for the input. For convenience, let us call $R(\cdot)$ the buyer's revenue function. The buyer seeks a procurement strategy to maximize its expected (net) profit.

There are n (> 1) potential suppliers for the buyer's input. For supplier i, $i = 1, \cdots, n$, the cost of producing Q units of the buyer's input is $c_i Q$, for any Q. It is common knowledge that the suppliers' unit costs, c_i's, are independent draws from a common probability distribution $F(\cdot)$ over $[\underline{c}, \overline{c}]$. Supplier i privately observes the value of c_i, but not the costs of other suppliers, $i = 1, \cdots, n$.

Here is an optimal solution to the buyer's procurement problem. The buyer announces a quantity-payment schedule, $P(\cdot)$, which is basically a commitment that says that the buyer will pay $P(Q)$ for Q units of input, for any Q. A supplier, if chosen by the buyer, is free to choose any quantity to deliver to the buyer and be paid according to the pre-announced plan. Therefore, the buyer has effectively proposed a business proposition to the potential suppliers. Of course, different suppliers will value this business deal differently, with the lowest-cost supplier deriving the highest value. In an English auction, the suppliers openly bid up the price they are willing to pay for the buyer's proposed contract, with the winner being the supplier willing to pay the highest price.[22] With a little bit of thinking, the lowest-cost supplier always wins the contract and pays a price equal to the value that the second-lowest cost supplier derives from the contract.

To better understand the above solution, suppose the buyer is a retailer, who buys a product from a supplier and re-sells it to customers. The selling price to the customers is p per unit, which is exogenously given. The total customer demand is D, a random variable with cumulative distribution function $G(\cdot)$. If demand exceeds supply, the excess demand is lost. Otherwise, the excess supply is useless and can be disposed of at no cost. The total quantity sold to customers is thus $\min\{Q, D\}$. The buyer's expected revenue is

$$R(Q) = pE[\min\{Q, D\}] = pE[Q - (Q - D)^+] = pQ - p \int_0^Q G(y)dy.$$

way. Frankly, there is little we can teach our students about how to manage a supply chain that resembles the beer game setup (at least, not yet). Interestingly, if we replace the 4-8 demand stream with a stream of i.i.d. random variables, and suppose the players all know the demand distribution, then we know how the game should be played. (A diligent reader would realize that this game was discussed in §2.3 of this chapter.) Under the optimal strategy, the bullwhip effect does not exist. But it may still occur (and it has) depending on the strategies used by the managers. So here is a game that can be used to illustrate the bullwhip effect, which we can say with confidence is bad. For a description of the i.i.d. version of the beer game and some teaching experience with it, see Chen and Samroengraja (2000).

[22] The theory of auctions is huge and well developed. Vickrey (1961) is seminal. Myerson (1981) and Riley and Samuelson (1981) are important milestones for their contributions to optimal auction design. McAfee and McMillan (1987) and Klemperer (1999) provide comprehensive reviews.

Note that this revenue function is concave and increasing in Q. To make things even simpler, suppose the suppliers' costs are drawn from the uniform distribution over $[0,1]$. Under this condition, the optimal quantity-payment schedule is

$$P(Q) = \frac{1}{2}R(Q).$$

Note that this payment schedule is independent of the number of potential suppliers. Moreover, it is a revenue-sharing contract: the business deal the buyer proposes calls for a 50-50 split of the buyer's revenue. It is also a returns contract, which says that the buyer pays the winning supplier a wholesale price of $w = p/2$ for each unit of input delivered (before demand realization), and in case there is excess supply after demand is realized, the buyer can return the excess inventory to the supplier for a full refund. Under this contract, and assuming the returned inventory has no value to any supplier, a supplier, if he wins, earns the following expected revenue (as a function of the production quantity Q):

$$E[\frac{p}{2}Q - \frac{p}{2}(Q-D)^+] = P(Q).$$

To put our hands around the inefficiencies caused by the asymmetric cost information, let us further assume that $G(x) = x$ for $x \in [0,1]$ and $p = 2$. In this case, $R(Q) = 2Q - Q^2$, and the optimal quantity-payment schedule becomes $P(Q) = Q - Q^2/2$. As mentioned earlier, the lowest-cost supplier wins the contract. Let C_1 be the cost of the winning supplier. (Thus $C_1 = \min\{c_1, \cdots, c_n\}$.) The quantity delivered by the winning supplier solves the following problem

$$Q(C_1) = \text{argmax}_Q P(Q) - C_1 Q = \text{argmax}_Q (1 - C_1)Q - Q^2/2.$$

Therefore $Q(C_1) = 1 - C_1$. The total profit for the supply chain (the buyer plus the winning supplier) is $R(Q(C_1)) - C_1 Q(C_1) = 1 - C_1$, with an expected value

$$\pi = 1 - E[C_1].$$

On the other hand, the efficient input quantity, one that maximizes the supply chain profit, is

$$Q^*(C_1) = \text{argmax}_Q R(Q) - C_1 Q = \text{argmax}_Q (2 - C_1)Q - Q^2.$$

Therefore, $Q^*(C_1) = (2 - C_1)/2$. Note that $Q^*(C_1) > Q(C_1)$; asymmetric cost information reduces the input quantity. The maximum expected supply chain profit under full information is

$$\pi^* = \pi + \frac{1}{4}E[C_1^2].$$

Therefore the supply chain inefficiency due to asymmetric information is

$$\pi^* - \pi = \frac{1}{4}E[C_1^2] = \frac{1}{2(n+1)(n+2)}$$

which is decreasing in n. This is the value created if the suppliers disclose their cost information. But why should they?

Leadtime Information

Another important piece of information coming from the supply side is the status of a replenishment order. Chen and Yu (2001a) address the value of leadtime information in the following inventory model. A retailer buys a single product from an outside supplier, stores it in a single location, and sells it to her customers. Customer demand arises periodically, with demands in different periods being i.i.d. random variables. If demand exceeds the on-hand inventory in a period, the excess demand is backlogged. On-hand inventories incur holding costs, and customer backorders incur penalty costs. The analysis is done from the retailer's standpoint: how to make replenishment decisions so as to minimize the retailer's long-run average holding and backorder costs.

Here is the supply process. Let L_t be the leadtime for an order placed in period t. And $\{L_t\}$ is a Markov chain with a finite state space. The one-step transition matrix of the Markov chain is chosen so as to prevent order crossovers (so orders are received in the sequence in which they were placed). The supply process is exogenous, i.e., the evolution of the Markov chain is independent of the operations of the retailer's inventory system.[23] The supplier observes the state of the Markov chain $\{L_t\}$, and he may or may not share this information with the retailer.

Two scenarios are considered. First, suppose the retailer knows the value of L_t for each period t before her replenishment decision. In this case, the optimal policy is to place an order in period t so as to increase the retailer inventory position up to a base-stock level that is a function of L_t, for all t. That is, a state-dependent, base-stock policy is optimal. On the other hand, suppose now that the supplier does not share with the retailer the leadtime information. In this case, the retailer has to rely on the history of order arrivals to infer something about the current leadtime and make her replenishment decisions accordingly. By comparing these two solutions, one sees the value of leadtime information. Numerical evidence indicates that the value of leadtime information is small for small-volume items, but significant for high-volume items where the percentage cost savings due to leadtime information can be as high as 35%.

Capacity Information

Our third example deals with the value of capacity information. Chen and Yu (2001b) consider a model with one retailer and one supplier. There is a single selling season. The retailer has two opportunities to place orders with the supplier before the season starts, one at time 0 and one at time 1. At time 0, the supplier has unlimited capacity, i.e., whatever the retailer orders will be ready for the selling season. At time 1, the supplier's capacity is uncertain, and it can be written as $C - \epsilon$, where C is the "forward capacity," i.e., the supplier's capacity at time 1 perceived at time 0, and ϵ, which can be positive or negative, is an external random shock reflecting uncertainties between time 0 and time 1. At time 0, there are two possible states, high or low, for the total demand in the selling season. A cumulative distribution is given for each demand state. At time 1, the true demand state is revealed; the retailer now has better demand information. This suggests that there is a benefit for the retailer to postpone the ordering decision to time 1. But the cost of doing this is that the retailer may not get what he orders (at time 1), due to the supplier's capacity constraint. Let q_0 be the quantity ordered by the retailer at time 0, and q_1^s the quantity ordered at time 1 if the demand state is s, $s =$ high or low. The optimal values of these quantities balance the benefit from demand information with the cost of capacity risk.

Before time 0, it is common knowledge that C comes from a given probability distribution. At time 0, before the retailer decides on the value of q_0, the supplier privately observes the realized

[23] Song and Zipkin (1996b) have provided several concrete examples to motivate such a leadtime process. Other models of random leadtimes have been provided by Kaplan (1970), Nahmias (1979), Ehrhardt (1984) and Zipkin (1986).

value of C. The retailer then offers a menu of contracts: a mapping from the supplier's reported capacity (which can be different from the true value of C) to q_0. The supplier then reports a value of C, effectively choosing a value for q_0. (This is the screening idea, to be discussed in detail in §3.1.) The solution to this asymmetric information case is then compared with the full-information scenario where the retailer also sees the value of C (before deciding q_0). The comparison gives the value to the retailer of knowing the supplier's forward capacity.

2.3 Information Transmission

Chen (1999a) considers a supply-chain model where information transmission is subject to delays. A firm has N divisions arranged in series. Customer demand arises at division 1, division 1 replenishes its inventory from division 2, 2 from 3, etc., and division N orders from an outside supplier. The demands in different periods are independent draws from the same probability distribution. Each division is managed by a division manger. Information in the form of replenishment orders flows from downstream to upstream, triggering material flow in the opposite direction. Both flows are subject to delays.[24]

An important feature of the model is that the division managers only have access to local inventory information. That is, each manager knows 1) his on-hand inventory, 2) the orders he has placed with the upstream division, 3) the shipments he has received from the upstream division, 4) the orders he has received from the downstream division, and 5) the shipments he has sent to the downstream division. However, he does not exactly know the shipments that are in transit from the upstream division that may be unreliable, and neither does he know the orders from the downstream division that are currently being processed. The decisions made by each manager can only be based on what he knows.

The first model considered by Chen assumes that the division managers behave as a *team*, i.e., they have a common goal to minimize the system-wide costs.[25] This is reasonable when, e.g., the owner of the firm has implemented a cost-sharing plan whereby each manager's objective function is a fixed, positive proportion of the overall cost of the system. It is shown that the optimal decision rule for each division manager is to follow an installation, base-stock policy. Division i's installation stock is equal to its net inventory (on-hand inventory minus backorders) plus its outstanding orders. Recall that manager i knows the orders he has placed (with the upstream division) as well as the shipments he has received (from the upstream division). The difference between the two is the outstanding orders. Therefore, installation stock is *local information*. The optimal decision rule for each division manager is to place an order in each period to restore the division's installation stock to a constant target level, which may be division-specific.

The solution to the team model reveals the role played by the information leadtimes (delays in the information flow). In terms of division i's safety stock, the information leadtime from division i to $i+1$ plays exactly the same role as the production/transportation leadtime from division $i+1$ to i; the safety stock level only depends on the sum of the two leadtimes. (This is intuitive at first glance. But the optimality proof requires some finessing.)

[24] This may remind you of the beer game, which is described in Sterman (1989). A key difference is the i.i.d. demand process assumed here.

[25] For the economic theory of teams, see Marschak and Radner (1972).

An alternative to the team model is the cost-centers model, where each manager is evaluated based on his division's performance. But how should local performance be determined? Chen suggests using the so-called accounting inventory level. The accounting inventory level at a division is its net inventory under the hypothetical scenario where no orders by the division will ever be backlogged at the upstream division. Note that the accounting inventory level may differ from the actual inventory level, because the upstream division is not always reliable. A division is charged a holding cost if its accounting inventory level is positive and a penalty cost otherwise.[26] It has been shown that the owner of the firm can choose the cost parameters so that when the individual division managers minimize their own (accounting) costs, the system-wide costs are also minimized.[27]

Firms decentralize the control of their operations for many reasons. One key reason is that the local managers are better informed about the local operating environments than the owner is. Therefore, it makes sense to let the local managers make local decisions. In the above supply chain model, let us suppose the division managers all know the true demand distribution, but the owner of the firm does not. Consider the following two scenarios. In one, the owner solves the team model based on her (erroneous) knowledge of the demand distribution and tells her employees to implement the installation, base-stock policies she found. (The owner only provides the decision rule, leaving the division managers to implement it. Since the division managers only have access to local information, the decision rule must be based on local information.) Call this the dictator scenario. In the other scenario, the owner organizes the divisions as cost centers. After the owner has specified a measurement scheme, each division manager chooses a replenishment strategy to minimize his accounting costs by using the true demand distribution. In both cases, the system-wide performance will be suboptimal, because the owner's inaccurate knowledge about the demand distribution has been used in one way or another. Numerical examples provided by Chen show that the system-wide performance under cost centers is nearly optimal, whereas the dictator scenario can be far from optimal. The benefit of decentralization is clear. Moreover, the measurement scheme for the cost centers is rather robust with respect to shifts in the demand distribution; a scheme based on an outdated demand distribution works very well for a new demand distribution so long as the division managers update their replenishment strategies based on the new information.

Finally, what if managers make mistakes? To explore this issue, consider the following specific irrational behavior. Manager i strives to maintain his *net inventory* at a constant level Y: if it is below Y, order the difference; otherwise, do nothing. It is a mistake because the decision maker forgets about the outstanding orders. (Recall that the optimal strategy is to maintain the installation stock at a constant level.) This mistake corresponds to the "misperceptions of feedback" Sterman (1989) found in the beer game. A simulation study shows that such mistakes can be very costly, especially those committed at the downstream part of the supply chain.[28]

When a downstream manager follows an erroneous strategy, the upstream managers receive distorted (and delayed) demand information. This is at the heart of the problem. Now consider the following alternative design of information flow in the supply chain. When division 1 places an

[26] The accounting and management literature advocates that individuals should only be evaluated on controllable performance, see, e.g., Horngren and Foster (1991). For this reason, the actual local inventory level (on-hand inventory minus backorders) at a division is inadequate as a basis for measuring local performance, since it is also affected by decisions made at the other divisions. The accounting inventory level removes the impact of the upstream division, but is still affected by the downstream division's orders.

[27] For other coordination mechanisms for serial inventory systems, see Lee and Whang (1999) and Porteus (2000). Gérard P. Cachon discusses these papers in Chapter 6 of this volume.

[28] Watson and Zheng (2001) provide a more recent attempt to address supply chain mismanagement due to irrational managerial behavior.

order, he is also required to report the demand in the previous period. This demand information is then relayed to the upstream managers along with the orders. Assume that the rational managers place their orders according to the accurate demand information, whereas the irrational ones follow the above "forgetful" strategy. In this way, a downstream ordering mistake can no longer corrupt the upstream order decisions. Simulation results indicate that by making the accurate demand information accessible to the upstream members of the supply chain, the system becomes much more robust. This is another reason for sharing demand information.

Most of the supply chain models on information sharing assume that the transmission of information is instantaneous and reliable. (We just saw one exception.) Moreover, they assume (implicitly) that information/knowledge is always transmittable. (We will see an exception soon.) However, managers are sometimes endowed with knowledge that is so specific to the local operating environment that it is very difficult to share such knowledge. This is perhaps what people mean by "experience," the sharing (or rather the acquisition) of which may take years of apprenticeship. So a more realistic view of organizations is that there are two kinds of knowledge: one can be readily shared (e.g., sales data) and the other is difficult to share, i.e., information sharing takes time and effort, is imperfect with noise, or is just impossible. When the local, specific knowledge plays a dominant role, it is important to give the manager possessing the knowledge the authority to make decisions that have the most use of the knowledge. In other words, decisions rights should reflect the dispersion of knowledge in an organization. This is, however, not the only challenge in designing an organization because the distribution of knowledge is, to some extent, manageable. This points to another aspect of organizational design, i.e., an organization's information structure (or "who knows what"). We refer the reader to Hayek (1945) and Jensen and Meckling (1976, 1992) for further discussions on specific knowledge and the design of organizations. Below, we review a paper from the operations literature that studies the above issues in a supply chain context.

Anand and Mendelson (1997) consider a firm that produces and sells a product in n markets. Production takes place in one location, and the total cost of producing Q units is assumed to be

$$TC(Q) = cQ + \frac{1}{2}\gamma Q^2.$$

A common reason cited for assuming increasing marginal costs is capacity constraints, e.g., overtime is used when production exceeds a certain threshold level and the overtime wage is higher than the regular wage. The n markets each face an independent, linear demand curve. Consider market i, $i = 1, \cdots, n$. There are only two possible market states, high or low. If the market is high, the (inverse) demand curve is $P(q_i) = a_H - bq_i$, where q_i is the quantity of the product allocated to market i, and $P(q_i)$ is the corresponding market clearing price. On the other hand, if the market is low, the demand curve is $P(q_i) = a_L - bq_i$, with $a_L < a_H$. (No transshipments are allowed among the markets after the initial allocation.) The market state is denoted by a binary random variable, s_i: if $s_i = 1$ (0) the market is high (low).

A key feature of the model is that the n markets are managed by branch managers who possess two types of information: one is specific knowledge that is not transmittable to anyone else, and the other is transferable data. This is modeled by assuming that $s_i = x_i y_i$, where both x_i and y_i are binary random variables, with x_i representing transferable market-i data and y_i the unobservable market-i condition, $i = 1, \cdots, n$. It is assumed that $\{x_i, y_i, i = 1, \cdots, n\}$ are independent random variables with $Pr(x_i = 1) = t$ and $Pr(x_i = 0) = 1-t$, $0 < t < 1$, and $Pr(y_i = 0) = Pr(y_i = 1) = 1/2$, $i = 1, \cdots, n$. The value of t is common knowledge. The branch manager at market i observes the value of x_i as well as a binary signal L_i that may contain information about y_i with $y_i = L_i$ with probability $1-\alpha$ and $y_i = 1 - L_i$ with probability α. If $\alpha = 0$ or 1, then the signal is perfect; if $\alpha = 1/2$, then the signal does not provide any new information beyond the prior on y_i. It is clear that we can restrict to $0 \leq \alpha \leq 1/2$ without loss of generality. Under this restriction, the value of $(1-\alpha)$ represents the precision of the signal. The local signal L_i is branch manager i's specific knowledge that is not transferable to anyone else, $i = 1, \cdots, n$. (Refer to the original paper for motivating stories behind this elaborate design. An alternative view one may take is that information is always transmittable, but some kinds of information are very costly to transmit.)

The decision variables are the supply quantities q_i. The objective of any decision maker (to be specified below) is to maximize the firm's expected profits, which are equal to the revenues generated by the branches minus the production cost that depends on the total quantity.

Anand and Mendelson then consider three different organizational designs, depending on where decision rights reside and how information is distributed (through the design of the firm's information system, e.g.). The first design is a centralized one, where a "center" makes all the decisions by using all the transferable data but none of the specific knowledge. The firm thus has in place an information system that allows the branches to report their transferable data, x_i for $i = 1, \cdots, n$, to the center. The second design is decentralized, where each branch manager i makes his own quantity decision q_i based on his own specific knowledge (L_i) and transferable data (x_i). Therefore, in this case, there is no information sharing so that all local knowledge (transferable or not) remains local. The third design is in between the previous two, with the branches making their own quantity decisions based on their specific knowledge and all the transferable data (again, enabled by an intra-firm information system). That is, branch manager i determines the value of q_i with knowledge of $(x_1, x_2, \cdots, x_n)$ and L_i, $i = 1, \cdots, n$. This design is referred to as the "distributed" structure. The analysis of the second and the third organizational structures follows that of a team model, where the team members (i.e., branch managers) share a common goal but have access to different sets of information. (The team model thus assumes away all potential incentive problems. Anand and Mendelson also consider transfer-pricing schemes when incentive issues can't be ignored.)

It is intuitive (and true) that the distributed design dominates the decentralized design in terms of the firm's expected profits. The difference represents the value of information sharing, which is shown to increase in the number of branches at first and then decrease. In other words, the distributed structure adds more value to firms that operate in a moderate number of markets. On the other hand, the difference between the centralized and decentralized systems captures the tradeoff between coordination, information sharing, and local knowledge. The centralized system benefits from better coordination of quantity decisions (due to centralized decision making) and the pooling of all the transferable data. However, the decentralized system sometimes performs better than the centralized system, indicating the usefulness of the local knowledge.

In their concluding remarks, Anand and Mendelson said that "the design of organizations requires an analysis of what kinds of information the firm needs to acquire, alternative ways of distributing this informational endowment and ways of structuring the organization (i.e., allocation of decision rights) to match its information structure." (The parenthetical explanation is added.) What they have done in their paper is to treat the allocation of information and decision rights as design variables to be jointly determined, leaving out information acquisition.

3 Incentives for Sharing Information

Information sharing in supply chains with independent players is tricky. When a player has superior information, two things may happen. He may withhold it to gain strategic advantage, or he may reveal it to gain cooperation from others. If the former, the other (less informed) players may try to provide incentives for him to reveal his private information; this is called *screening*. If the latter, we have *signaling*, i.e., revealing information in a credible way. Sometimes it is impossible to say who has more or less information; players simply have different information about something they all care about. For example, different retailers may obtain different signals about the market demand for a product. In this case, a player's willingness to share his information depends on if the others are going to share their information and how the revealed information will be put to use. This section reviews papers that deal with information exchanges in decentralized supply chains. [29]

3.1 Screening

This subsection presents several examples where a firm tries to "smoke out" either consumer preferences or private information held by an employee or by a supply chain partner.[30]

A Simple Example

A product line usually refers to a range of goods of the same generic type, but differentiated along some attributes. For example, Dell offers two lines of notebook computers, Inspiron and Latitude, and each product line consists of models with different speeds, storage capacities, etc. An important question is how a firm can optimally design and price a product line.

Suppose a monopolist is offering a line of goods differentiated along a quality dimension. There are two consumers, with different demand intensities for quality. (This is thus a toy problem. But the basic ideas are here.) Consumer i values the variety with quality level q at $\theta_i q$, $i = 1, 2$, with $0 < \theta_1 < \theta_2$. Therefore, both consumers prefer more quality, but differ in their willingness to pay for any given quality level. Each consumer buys one unit of the good, or nothing at all. There are constant marginal costs of production at any given quality level, and the marginal cost of producing variety q is q^2.

[29] A branch of economics (sometimes called information economics) addresses issues arising from various information asymmetries. One type of information asymmetry is often studied under the heading "moral hazard," which refers to situations where one party (called agent) performs a task on behalf of another (called principal), and the agent's effort level is unobservable to the principal. A conflict arises because the principal prefers the agent to work hard while the agent dislikes exerting effort. The solution is an incentive contract that pays the agent for his output. The principal-agent theory, originating from economics, has been used/developed in the accounting, marketing, and lately operations literatures. Kreps (1990) provides an excellent introduction to the principal-agent theory; some of the seminal papers in this area are cited later in this chapter. For principal-agent models in the operations literature, see, e.g., Porteus and Whang (1991), Chen (2000b), and Plambeck and Zenios (2000a,b). We choose not to review this part of the literature here because the goal of providing incentives in principal-agent models is to induce a certain level (or pattern) of effort by an agent (not to facilitate information sharing). By the way, many of the ideas behind the papers reviewed in this section originated from economics often under rubrics such as adverse selection, mechanism design, or signaling.

[30] There is a large body of research on screening in queueing contexts. For example, a service provider can charge different prices for different priority levels. An arriving customer decides which priority class to join based on his/her (private) cost of waiting. For a comprehensive survey of this literature, see Hassin and Haviv (2001).

Consider first the case of a perfectly discriminating monopolist who is able to sell to each consumer individually (i.e., deny one consumer access to the product offered to the other consumer) and prevent any re-sales. The monopolist will thus charge each consumer his reservation price and the only remaining problem is what variety to offer to each consumer. To solve this problem, simply maximize $\theta_i q - q^2$ over q for each i. Therefore, the optimal strategy is to offer quality level $\theta_i/2$ to consumer i and charge him $\theta_i^2/2$ for it. Both consumers will buy the products offered them and derive zero surplus. The firm may be able to achieve this ideal solution in some cases. For example, a telephone company is routinely charging different rates to business users and residential users (it is relatively easy for the company to verify if a user is business or residential and it is very difficult to trade phone calls between business and residential users). In other cases, it is impossible to deny one consumer access to the products offered to other consumers, for technical or legal reasons. In other words, a product line, whatever it may be, must be made available to all types of consumers. What should the monopolist do then?

First, note that the above solution will not work when consumers self-select. When given the above two variety-price combinations: $(\theta_i/2, \theta_i^2/2)$ for $i = 1, 2$, consumer 1 will continue to choose $(\theta_1/2, \theta_1^2/2)$, but consumer 2 will switch from $(\theta_2/2, \theta_2^2/2)$ to $(\theta_1/2, \theta_1^2/2)$ and earns a positive surplus (before, he earned zero surplus). In order to prevent consumer 2 from switching, the firm must lower the price for variety $\theta_2/2$, if everything else stays unchanged. Finding the monopolist's optimal strategy involves a systematic tradeoff among multiple dimensions. Let (q_i, p_i), $i = 1, 2$, be the quality-price pairs offered to the market (i.e., the two consumers). Suppose consumer i chooses (q_i, p_i). (If the two pairs are identical, then the product line consists of only one good.) We first consider the case where both consumers are served (i.e., they each buy a unit). The monopolist's problem can be written as:

$$
\begin{array}{lll}
\max\limits_{q_1,p_1,q_2,p_2} & (p_1 - q_1^2) + (p_2 - q_2^2) & \\
s.t. & \theta_1 q_1 \geq p_1 & (P1) \\
 & \theta_2 q_2 \geq p_2 & (P2) \\
 & \theta_1 q_1 - p_1 \geq \theta_1 q_2 - p_2 & (SL1) \\
 & \theta_2 q_2 - p_2 \geq \theta_2 q_1 - p_1 & (SL2)
\end{array}
$$

where the objective function represents the firm's total profits, the first two constraints ((P1) and (P2)) are necessary in order for the consumers to participate, and the last two constraints ((SL1) and (SL2)) ensure that the consumers choose the right bundle. This problem is easy. First note that $\theta_2 q_1 - p_1 \geq \theta_1 q_1 - p_1$ and the right side is nonnegative from (P1). This, together with (SL2), implies that $\theta_2 q_2 \geq p_2$. Therefore, (P2) is redundant and is thus deleted. Moreover, (P1) must bind, for otherwise, one can simultaneously increase p_1 and p_2 by the same amount without violating any constraints. Note that (SL1) and (SL2) can be combined to produce:

$$\theta_2(q_2 - q_1) \geq p_2 - p_1 \geq \theta_1(q_2 - q_1).$$

It then follows that $q_2 \geq q_1$, which then implies $p_2 \geq p_1$. On the other hand, (SL2) must bind, because if not, one can increase p_2 to get a better solution for the monopolist. Therefore, $p_2 - p_1 = \theta_2(q_2 - q_1) \geq \theta_1(q_2 - q_1)$ because $q_2 \geq q_1$. Consequently, (SL1) is implied by the binding version of (SL2) plus $q_2 \geq q_1$. In sum, the above optimization is equivalent to

$$
\begin{array}{ll}
\max\limits_{q_1,p_1,q_2,p_2} & (p_1 - q_1^2) + (p_2 - q_2^2) \\
s.t. & p_1 = \theta_1 q_1
\end{array}
$$

$$\theta_2 q_2 - p_2 = \theta_2 q_1 - p_1$$
$$q_2 \geq q_1.$$

This has a closed-form solution:

$$q_1^* = \max\{\theta_1 - \frac{\theta_2}{2}, 0\}, \quad q_2^* = \frac{\theta_2}{2}.$$

Note that compared with the previous solution without consumer self-selection, the lower demand-intensity consumer (i.e., consumer 1) gets a lower quality product, a lower price, and the same (zero) surplus, whereas the higher demand-intensity consumer gets the same quality product, a lower price, and a positive surplus. Moreover, if we interpret $q_2^* - q_1^*$ as the breadth of the product line, self-selection leads to a broader range of goods. Finally, if $\theta_1 \leq \theta_2/2$, consumer 1 is not served.

The above simple example captures the basic idea of screening. For more sophisticated models involving consumer self-selection, see, e.g., Mussa and Rosen (1978), Maskin and Riley (1984), and Moorthy (1984).

Quality, broadly interpreted, represents product attributes for which the consumer preference is of the more-is-better kind. Other product attributes include color, size, taste, etc., and each consumer is likely to have a unique ideal point in the attribute space.[31] It is possible to differentiate products along these dimensions as well. For this part of the literature, we refer the reader to Shocker and Srinivasan (1979), Green and Krieger (1985), Lancaster (1979, 1990), de Groote (1994), Dobson and Kalish (1988, 1993), Nanda (1995), Chen et al. (1998), and Yano and Dobson (1998). A recent book edited by Ho and Tang (1998) contains further references on this topic.

Market Segmentation and Product Delivery

Different customers may exhibit different degrees of aversion to waiting: some want to have their orders delivered right away, while others can tolerate a delay. Therefore, the delivery schedule of a product can be a useful tool for segmenting the market. One benefit of such a segmentation strategy is that when a customer places an order that does not have to be shipped immediately, the firm obtains advance demand information that can be used for better production-distribution planning. A potential cost of this strategy occurs when a price discount must be offered in order for a customer to accept a delay. How can a firm design an optimal price-delivery schedule?

Chen (2001a) provides a model to address the above question. A firm sells a single product to consumers. The firm announces a price-delay schedule $\{(p_k, \tau_k)\}_{k=0}^K$, for some nonnegative integer K, where p_k is the price a customer pays if he agrees to have his order shipped τ_k units of time after order placement, with $0 = \tau_0 < \tau_1 < \cdots < \tau_K$ and $p_0 > p_1 > \cdots > p_K$. The maximum price p_0 is paid only if a consumer wants immediate shipment.

The consumers divide into M segments. Let $u_m(\tau)$ be the maximum (or reservation) price that the customers in segment m are willing to pay for one unit of the product if their orders are shipped τ units of time after order placement, $m = 1, \cdots, M$. It is assumed that $u_m(\cdot)$ is decreasing,

[31] The *ideal-point model* is often used to describe consumer preferences along dimensions that are not quality-like. For example, a consumer's utility of buying a product with level x of certain attribute can be written as $A - (x - a)^2$, where A is the maximum possible utility level and a is the ideal attribute level for the consumer. Different consumers can have different ideal attribute levels.

differentiable, convex with $u'_m(\tau) < u'_{m+1}(\tau)$ (< 0) for all τ.[32] Thus, segment 1 is uniformly more sensitive to waiting than segment 2, segment 2 more so than segment 3, and so on. The surplus that a type-m customer derives from (p_k, τ_k) is $u_m(\tau_k) - p_k$. His objective is to choose a pair from the schedule that maximizes his surplus, i.e., consumer self-selection.

The product is replenished by an N-stage supply chain. Stage 1 is the final stocking point from which product is shipped to customers, stage 1 is replenished by stage 2, stage 2 by stage 3, etc., and stage N by an outside supplier with ample stock. The transit time from one stage to the next is constant. Customer orders are satisfied first-come, first-served according to the sequence of their shipping dates, not the dates in which the orders are placed. If the firm cannot ship an order on the date chosen by the customer because the product is out of stock (at stage 1), the order is backlogged. The backlogged orders are shipped as soon as inventory becomes available. In this case, the firm incurs a goodwill loss (or backorder cost). In addition, the firm incurs holding costs for inventories held in the supply chain and variable costs for every unit sold.

Determining an optimal price-delivery schedule turns out to be a hard problem. Below is a brief description of the solution.

The optimal schedule always has the segments bundled in a sequential manner, with lower segments choosing higher prices and shorter delays. For example, suppose there are five segments in the market, and the firm offers $\{(p_0, 0), (p_1, \tau_1), (p_2, \tau_2)\}$. Sequential bundling means something like the following. Suppose segments 1 and 2 choose $(p_0, 0)$, segment 3 chooses (p_1, τ_1), and segments 4 and 5 choose (p_2, τ_2). Given this, a property of the optimal schedule is that segment 3 is indifferent between (p_1, τ_1) and $(p_0, 0)$, and segment 4 is indifferent between (p_2, τ_2) and (p_1, τ_1). (It is assumed that when indifferent, a consumer will choose the lower price.) These indifference relationships imply that a price-delay schedule is fully specified by the delays and the marginal segments, i.e., (τ_1, τ_2) and segments 3 and 4 in the above example.

Now fix the price-delay schedule and consider the firm's supply chain. Suppose segment m chooses a shipping delay equal to l_m, $m = 1, \cdots, M$. Therefore, if a segment-m customer places an order at time t, the corresponding demand occurs at time $t + l_m$. Suppose $l_m > 0$ for some m. Therefore, some orders serve as warnings of future demand, and the question is how this information can be incorporated into the firm's replenishment strategy. The optimal strategy, and the corresponding minimum supply chain costs, can be obtained by carefully separating the known demand information from the unknown and following the approach of Chen and Zheng (1994). The optimal policy is an echelon base-stock policy with floating order-up-to levels (one for each stage).

To gain some intuition on how the shipping delays translate into cost savings, suppose $N = 2$. Let the leadtimes at stages 1 and 2 be 4 and 2 periods respectively. To make matters really simple, assume there is only one segment choosing a shipping delay of l periods. It is easy to see that if $l = 6$, the supply chain faces no demand uncertainty at all and as a result, no inventories need to be carried at any stage. Now suppose $l = 5$. In this case, there is no need to carry any inventory at stage 1. Consider the problem facing stage 2. Say a customer order arrives at time t (at stage 1). This order needs to be shipped out of stage 1 at time $t + l = t + 5$. This means that the order needs to be shipped out of stage 2 at time $t + 1$. Therefore, stage 2's problem is basically a single-location

[32] Note that $u_m(0) - u_m(\tau)$ is the cost of waiting for a segment-m customer. The assumption that $u_m(\cdot)$ is convex implies that the cost of waiting is concave, i.e., the marginal cost of waiting is decreasing. This is true if, for example, the excitement about the product decreases over time after the order is placed. In this case, the marginal cost of waiting is very high in the first few days, while the excitement still lingers, and decreases as time goes by.

inventory problem where customers choose a shipping delay of 1 period. This, together with the fact that the leadtime at stage 2 is 2 periods, implies that the demand uncertainty facing stage 2 is just one-period worth of demand, not the usual two-periods worth of demand. Finally, if, say, $l = 3$, then both stages will need to carry some safety stock. With some thinking, the reader will see that the demand uncertainty facing stage 1 is one-period ($= 4 - 3$) worth of demand and the demand uncertainty facing stage 2 remains to be two-periods worth of demand.

An optimal price-delay schedule can be obtained by solving an optimization problem that captures both the costs and benefits of the segmentation strategy. Numerical results show that the net benefit of this strategy can be substantial.

Screening and Moral Hazard

Sometimes the party being screened also takes hidden actions. Chen (2000c) studies such a model. Suppose a firm sells a single product through a single sales agent. The market demand is the sum of the agent's selling effort (a), the market condition (θ), and a random shock (ϵ), i.e.

$$X = a + \theta + \epsilon$$

where θ and ϵ are independent random variables, $Pr(\theta = \theta_H) = \rho$ and $Pr(\theta = \theta_L) = 1 - \rho$ for $0 < \rho < 1$ and $\theta_H > \theta_L > 0$, and $\epsilon \sim N(0, \sigma^2)$. The agent privately observes the value of θ, and the agent's effort level is not observable to the firm. The firm's decisions are how to compensate the agent for his work (i.e., a wage contract) and how much to produce before demand realization. Given a contract, the agent decides whether or not to accept it and if so, how much effort to expend. [33]

The model assumes the following sequence of events: 1) the firm (or principal) offers a menu of wage contracts (for screening); 2) the agent privately observes the value of θ; 3) the agent decides whether or not to participate (work for the firm) and if so, which contract to sign; 4) under a signed contract, the firm determines the production quantity, and the agent makes the effort decision; and 5) ϵ is realized.

Consider the agent's decisions when offered a menu of contracts. First, he considers each contract on the menu and determines the maximum expected utility that can be obtained under the contract. Suppose $s(\cdot)$ is the contract being considered, i.e., $s(x)$ is the wage paid to the agent if the total sales is x. Assume the agent's utility for net income z is $U(z) = -e^{-rz}$ with $r > 0$. [34] Note that $U(\cdot)$ is increasing and concave, implying that the agent is risk averse. The net income is the wage received, $s(X)$, minus the cost of effort, $V(a) = a^2/2$.[35] To determine the maximum expected utility

[33] Coughlan (1993) reviews the salesforce compensation literature. A common assumption is that the total sales is a function of selling effort and a random shock and that effort is unobservable to the firm. This is the moral hazard problem, which has been widely studied in the economics/agency-theory literature, see, e.g., Shavell (1979), Harris and Raviv (1978,1979), Holmstrom (1979, 1982), and Grossman and Hart (1983). If, in addition to the moral hazard problem, the firm is in an informational disadvantage in terms of the sales environment, i.e., the sales people have superior information about the sales response function (the productivity of selling effort, the sensitivity of customers to price changes, the sales prospects, etc.), then the firm also faces an adverse selection problem. The typical solution is a menu of contracts. The salesforce compensation literature in marketing includes Basu et al. (1985), Lal (1986), Lal and Staelin (1986), Rao (1990), and Raju and Srinivasan (1996).

[34] The negative exponential utility function is widely used in agency models.

[35] The quadratic form is not critical for the analysis. An often-assumed feature of the cost-of-effort function is increasing marginal cost of effort.

achievable under $s(\cdot)$, the agent solves the following optimization problem

$$\max_a E[-e^{-r(s(X)-V(a))}].$$

Recall that the agent has already observed the value of θ when evaluating the contract. Therefore, the above expectation is with respect to ϵ given the observed value of θ. If the maximum expected utility is greater than or equal to U_0, the agent's reservation utility representing the best outside opportunity for the agent, then s is said to be acceptable to the agent. [36] Among all the contracts on the menu, the agent chooses the one with the highest achievable expected utility and participates if this utility level exceeds U_0.

We now turn to the principal's problem. Recall that the firm must make its production decision before observing the total sales. This is reasonable when the customers demand fast delivery of their orders and the production leadtime is relatively long. (It is thus impossible to follow make-to-order.) Let Q be the production quantity. Let c be the cost per unit produced. If $X \leq Q$, the excess supply is salvaged at p per unit. On the other hand, if $X > Q$, the excess demand must be satisfied by a special production run at a cost of c' per unit. Let the unit selling price be $1 + c$ (the profit margin is thus normalized to 1). To avoid trivial cases, assume $p < c < c' < 1 + c$. As mentioned before, the firm makes contracting as well as production decisions with the objective of maximizing its expected profit (the principal is thus risk neutral). If $s(\cdot)$ is the contract signed by the agent, the firm's profit is

$$\begin{aligned}(1+c)X - s(X) - cQ + p(Q-X)^+ - c'(Q-X)^- \\ = X - s(X) - [(c-p)(Q-X)^+ + (c'-c)(Q-X)^-]\end{aligned}$$

where $w^+ = \max\{w, 0\}$ and $w^- = \max\{-w, 0\}$. Note that the optimal production quantity minimizes

$$E[(c-p)(Q-X)^+ + (c'-c)(Q-X)^-]$$

where the expectation is with respect to X given the principal's knowledge about the market condition and the agent's selling effort (inferred not observed) after a contract is signed.

Since there are only two possible market conditions, the firm needs to offer at most two contracts. Let $s_H(\cdot)$ be the contract chosen by the high-type agent, and $s_L(\cdot)$ chosen by the low type. The principal, by putting herself in the shoes of the agent, can anticipate the amount of selling effort under each type. Let a_H be the selling effort of the high-type agent, and a_L the effort of the low type. Assume that $s_H(\cdot) \neq s_L(\cdot)$. In this case, the principal discovers the market condition after observing the contract choice made by the agent. If $\theta = \theta_H$ then $X \sim N(a_H + \theta_H, \sigma^2)$; otherwise, if $\theta = \theta_L$, then $X \sim N(a_L + \theta_L, \sigma^2)$. And the principal can make her quantity decision accordingly. This is a benefit the principal obtains from screening.

One way to achieve screening is by offering a menu of (two) linear contracts. Let $s_H(x) = \alpha_H x + \beta_H$ be the contract intended for the high-type agent, and $s_L(x) = \alpha_L x + \beta_L$ the contract intended for the low-type agent, with $\alpha_H, \alpha_L \geq 0$. It can be shown that the optimal values of the contract parameters are

$$\alpha_H = \frac{1}{1 + r\sigma^2}$$

[36] It is reasonable to assume that the reservation utility does not depend on the agent's type, because what distinguishes the high type from the low type is the market condition, something unrelated to the agent's intrinsic quality.

$$
\begin{aligned}
\alpha_L &= \frac{1}{1+r\sigma^2}\max\left\{1-\frac{\rho}{1-\rho}(\theta_H-\theta_L),0\right\}\\
\beta_L &= -\frac{\ln U_0}{r}-\alpha_L\theta_L-\frac{1-r\sigma^2}{2}\alpha_L^2\\
\beta_H &= -\frac{\ln U_0}{r}+\alpha_L(\theta_H-\theta_L)-\alpha_H\theta_H-\frac{1-r\sigma^2}{2}\alpha_H^2.
\end{aligned}
$$

Another way to achieve screening is suggested by Gonik (1978). Under his scheme, the firm asks the salesperson to submit a forecast of the total sales. If the forecast is F, then $s(x|F)$—a given function of the actual total sales x parameterized by F—is the compensation for the agent. Therefore, the firm is effectively offering a menu of contracts; by submitting a forecast, the agent chooses a particular contract from the menu. Gonik's original proposal uses the following functional forms: $s(x|x) = \alpha x + \beta$ for all x, and for any x and F,

$$
s(x|F) = \begin{cases} s(F|F) - u(F-x) & x \le F \\ s(F|F) + v(x-F) & x > F \end{cases}
$$

where α, β, u, and v are contract parameters chosen by the firm with $u > \alpha > v > 0$. Note that $s(x|x) \ge s(x|F)$ for all F and x. Therefore, if the agent expects to sell x units, it is in his best interest to submit a forecast that is equal to x. Also, for any given F, $s(x|F)$ is increasing in x, providing the agent with incentives to generate more sales.

It can be shown that the agent's optimal effort level is $a^* = \alpha$, which is entirely determined by only one contract parameter, α, and it is independent of the agent's type. Moreover, the optimal forecast decision is $F^* = z^* + a^* + \theta = z^* + \alpha + \theta$ for some value z^*, which depends on α, u, v but is independent of β and the agent's type. Therefore, the high-type agent forecasts $F_H = z^* + \alpha + \theta_H$ and the low-type forecasts $F_L = z^* + \alpha + \theta_L$. The agent is screened!

Numerical examples comparing the menu of linear contracts with the Gonik scheme show that the former dominates the latter in terms of the firm's expected profits.

Screening in Supply Chains

An important type of information asymmetry in supply chains is about cost structures. A supplier may only have imperfect knowledge about a buyer's cost structure, and vice versa. Here again the less informed may try to screen the more informed with a menu of contracts. Below, we describe a few papers that deal with screening in supply chains with asymmetric cost information.[37]

Ha (2001) provides a screening model where the supplier does not know the buyer's marginal cost. The setting is that of the newsvendor model (with pricing): the buyer faces a demand that is stochastic and price-sensitive, and before demand realization, an order quantity must be determined together with the selling price. The demand model is the additive kind, i.e., $D = \mu(p) + Y$ where p is the selling price, μ is a decreasing and concave function, and Y is a random variable independent of p. Let s be the supplier's marginal cost of production, and c the buyer's marginal cost (of selling and maybe additional processing). A key feature of the model is that c is known only to the buyer, with the supplier endowed with a prior distribution of c over a finite interval. Everything else is assumed to be common knowledge. The analysis is from the standpoint of the supplier: how to offer a menu of contracts to the buyer so as to maximize the supplier's expected profit.

[37] A careful reader would realize that some of the models discussed in §2.2 fall under this category.

The contract menu is restricted to be of the form: $\{p(\hat{c}), q(\hat{c}), R(\hat{c})\}$, where $\hat{c}$ is the buyer's announced marginal cost (which can be different from the true cost c), and works as follows: if the buyer announces $\hat{c}$, then the supplier will deliver $q(\hat{c})$ units to the buyer for a total payment of $R(\hat{c})$, and the buyer is to set the selling price at $p(\hat{c})$. The functions $p(\cdot)$, $q(\cdot)$ and $R(\cdot)$ are chosen by the supplier. Given this menu, the buyer then decides whether or not to sign a contract and if so, which one (by choosing $\hat{c}$). Ha has solved this mechanism design problem.

As mentioned in Ha (2001), the above menu of contracts is a nonlinear contract with price fixing, and this may run into problem with commercial laws such as Resale Price Maintenance (RPM), see, e.g., Tirole (1988). This would not be a problem if the contract menu is changed to $\{q(\hat{c}), R(\hat{c})\}$ and the retailer is free to choose any selling price after contract signing. Ha has not solved this problem, except for a special case where the selling price is exogenously given.

Corbett and de Groote (2000) consider a model with one supplier and one retailer, where the retailer's holding cost parameter h_b is unknown to the supplier. The basic setup is a two-stage economic lot-sizing problem with deterministic demand and no backlogging, with an additional restriction that the supplier's lot size is equal to the retailer's (i.e., the lot-for-lot replenishment). The supplier, however, is endowed with a prior distribution of h_b. The problem facing the supplier can be formulated as a direct revelation game, whereby the supplier asks the retailer to announce the value of h_b: if the announced value is $\hat{h}_b$, then the lot size is $Q(\hat{h}_b)$ and the discount is $P(\hat{h}_b)$ given as a lump-sum payment per unit of time. The task is to determine the pair of functions $Q(\cdot)$ and $P(\cdot)$ so as to minimize the supplier's expected costs subject to the incentive compatibility constraint that the retailer always wants to announce his true holding cost. Corbett and de Groote show that the optimal $Q(\cdot)$ and $P(\cdot)$ are both decreasing functions, which can thus be interpreted as a quantity discount scheme because larger quantities are associated with larger discounts.

Corbett (2001) considers a supplier-retailer model with stochastic demand. The setup is basically the same as the classic (Q,r) model: whenever the retailer inventory position falls to the reorder point r, it orders (and the supplier produces) a batch of Q units. The twist here is that the supplier makes the lot-sizing decision (i.e., the value of Q) and incurs a fixed cost for each batch produced, and that the retailer determines the reorder point r and is responsible for the holding and backorder costs incurred at the retail site. [38] The inefficiency in this supply chain is evident if there is no coordination: the supplier will set the batch size to be infinity! Corbett derives screening solutions to the following scenarios: 1) the supplier privately observes the value of the fixed cost, and 2) the retailer privately observes the backorder penalty cost. Also discussed is how consignment—the practice of giving the ownership of retailer inventory to the supplier—affects supply chain coordination. We omit the details. For other studies on supply chain models with asymmetric cost information, see Corbett et al. (2001) and the references therein.

Another type of information asymmetry in supply chains is about demand information. For example, a retailer, due to its proximity to the market, may possess better information about the demand than the supplier. Cachon and Lariviere (1999) consider a one-period model with one supplier and N retailers. The retailers are local monopolists, each of which receives a private signal about its own market, which in turn determines its desired stocking level. The supplier has a finite

[38] Here lies a critical assumption: the supplier sets the retailer's order quantity. It is worth thinking about an alternative model where there are two quantity decisions: the supplier sets its production quantity, and the retailer sets its order quantity. Under the current cost structure, it is reasonable to assume that the production quantity is larger than the order quantity. Consequently, the supplier will also incur some inventory holding costs. How would supply chain coordination come about in this case? The same comment applies to Corbett and de Groote (2000).

capacity, and must determine an allocation mechanism in the event the sum of the retailer orders exceeds the capacity. (An allocation mechanism is therefore a mapping from a vector of retailer orders to a vector of capacity allocations.) They consider various allocation mechanisms and their impact on the supply chain. They found that some mechanisms induce the retailers to truthfully order their desired quantities, but the supply chain often fares better with a mechanism that induces order inflation (i.e., the retailers order more than they need hoping to get a higher allocation in the event of capacity shortage). In other words, the truthful sharing of retailer order information is not necessarily the appropriate goal for the supply chain. A recent paper by Deshpande and Schwarz (2002) considers a similar problem and derives an optimal mechanism from the supplier's standpoint.

3.2 Signaling

We begin with a simple example to illustrate the basics of signaling. We then review several supply chain models where an informed party likes to convey a piece of private information to the uninformed. These models are, interestingly, all set forth in the context of new product introductions.

A Simple Example

Consider a supply chain with one manufacturer and one retailer. The manufacturer (she) produces one product and sells it through the retailer (he). The retailer faces a linear demand function $D = a - p$. The demand intercept, a, has two possible values: $a = 8$ or $a = 4$. The manufacturer observes the value of a, while the retailer assesses a probability of ρ that $a = 8$ and a probability of $1 - \rho$ that $a = 4$, $0 < \rho < 1$. And the manufacturer is aware of this assessment by the retailer. To make numbers simple, say the manufacturer's marginal cost of production is zero. The one-period game begins with the manufacturer offering a wholesale price w. Then, the retailer sets the retail price p. Finally, the market demand is realized and profits accrue to the two players.

To start, let us consider the full information case, i.e., the value of a is also known to the retailer. In the high-type case (i.e., $a = 8$), the retailer chooses p to maximize $(p - w)(8 - p)$ when the wholesale price is w. The optimal solution is $p = (8 + w)/2$. Given this, the manufacturer maximizes $w(8 - (8 + w)/2)$. The solution is $w_H = 4$, which leads to a retail price of $p_H = 6$. The profits for the manufacturer and the retailer are $\pi_H^M = 8$ and $\pi_H^R = 4$, respectively. On the other hand, if $a = 4$, i.e., the low-type case, we have $w_L = 2$, $p_L = 3$, $\pi_L^M = 2$, and $\pi_L^R = 1$.

We can already see that the high-type manufacturer has an incentive to "pretend" to be low type. For example, suppose, miraculously, the high-type manufacturer charges $w = 2$ and the retailer believes she is actually low type. The retailer then chooses a retail price p to maximize $(p - w)(4 - p)$, leading to $p = 3$. In this case, the high-type manufacturer's profit is $w(8 - p) = 10$, which is higher than $\pi_H^M = 8$ under full information. The intuition is clear: the manufacturer "prefers" the retailer to think the demand intercept is low and hence, to charge a lower retail price leading to a higher demand. Interestingly, the retailer's actual profit in this case is $(p-w)(8-p) = 5$, which is also higher than the profit under full information $\pi_H^R = 4$. On the other hand, the low-type manufacturer does not want the retailer to think otherwise. To verify this is a good exercise.

Now back to the case with asymmetric information. Suppose the manufacturer sets the wholesale price at w. Let $\mu(w)$ be the probability the retailer attributes to the event that $a = 8$. If, e.g., $\mu(w) = \rho$, then the retailer obtains no new information about the demand intercept after observing

w. The other extreme is $\mu(w) = 0$ or 1, in which case the retailer learns the exact value of the demand intercept. With w and $\mu(w)$, the retailer chooses a retail price to maximize his expected profits:

$$(p - w)[8\mu(w) + 4(1 - \mu(w)) - p] = (p - w)[4 + 4\mu(w) - p].$$

Therefore, the optimal solution is

$$p(w, \mu(\cdot)) = \frac{4 + 4\mu(w) + w}{2} = 2(1 + \mu(w)) + w/2.$$

Given $\mu(\cdot)$, the manufacturer can anticipate what the retailer is going to do through the above equation for each possible value of w. The optimal wholesale price for the high-type manufacturer is the solution to

$$\max_w \pi_H^M(w) \overset{def}{=} w(8 - p(w, \mu(\cdot))) = w(6 - 2\mu(w) - \frac{w}{2}).$$

Let the solution be w_H^*. Similarly, the low-type's optimal wholesale price w_L^* solves

$$\max_w \pi_L^M(w) \overset{def}{=} w(4 - p(w, \mu(\cdot))) = w(2 - 2\mu(w) - \frac{w}{2}).$$

In sum, given $\mu(\cdot)$, the two players simply play a Stackelberg game with the manufacturer as the leader and the retailer as the follower. But the story does not end here. Where does $\mu(\cdot)$ come from? It must be consistent with the pricing strategies that prevail in the Stackelberg game. For example, if $w_H^* \neq w_L^*$, then a wholesale price equal to w_H^* signals to the retailer that $a = 8$, and a wholesale price of w_L^* signals $a = 4$. Therefore, to be consistent, we must have $\mu(w_H^*) = 1$ and $\mu(w_L^*) = 0$. On the other hand, if $w_H^* = w_L^*$, then the retailer is going to see only one wholesale price no matter what the manufacturer type is. In this case, consistency calls for $\mu(w_H^*) = \rho$. In the former case, we have a separating equilibrium because the retailer is able to separate the two manufacturer types; in the latter, a pooling equilibrium because the two manufacturer types do the same thing. A belief structure that is consistent is referred to as an equilibrium belief.

Let us see if there exists any equilibrium, separating or pooling, in the above game.

Suppose a pooling equilibrium exists. Let w^0 be the wholesale price chosen by both types of the manufacturer. Let $\mu^0(\cdot)$ be the retailer belief. Recall that consistency calls for $\mu^0(w^0) = \rho$. We claim that if $\rho < 3 - 2\sqrt{2} \approx 0.1716$, then the following strategy profile and belief form a pooling equilibrium: $w^0 = 2 - 2\rho$, $\mu^0(w) = \rho$ for all $w \leq 2 - 2\rho$ and $\mu^0(w) = 1$ for all other w. To verify, all we need to do is check that under the given belief, $w^0 = 2 - 2\rho$ is indeed the optimal choice for both manufacturer types. Consider first the high type. If the manufacturer offers a wholesale price w greater than w^0, then $\pi_H^M(w) = 4w - w^2/2$, which is maximized at $w = 4$ with $\pi_H^M(4) = 8$. If $w = w^0$, $\pi_H^M(w^0) = 6w^0 - (w^0)^2/2 - 2\rho w^0$. It is easy to verify that $\pi_H^M(w^0) > 8$ when $\rho < 3 - 2\sqrt{2}$. Moreover, for all $w < w^0$, $\pi_H^M(w) = 6w - w^2/2 - 2\rho w$ is increasing in w. This shows that the high-type manufacturer's optimal choice of a wholesale price is w^0. Now consider the low-type manufacturer. If the low-type manufacturer offers a wholesale price $w > w^0$, the retailer thinks the manufacturer is of high type. Under this scenario, the low-type manufacturer's profit function is $\pi_L^M(w) = -w^2/2$. Therefore, it is not in the interest of the low-type manufacturer to "pretend" to be of high type. For all $w \leq w^0$, $\pi_L^M(w) = 2w - w^2/2 - 2\rho w$, which is maximized at $w = w^0$. This establishes that the above strategy-belief combination is a pooling equilibrium. [39]

[39] Various equilibrium refinements are possible. We consider one here, i.e., the test of equilibrium domination that is also known in the literature as the "intuitive criterion." The reader is referred to Kreps (1990) for discussions on the intuitive criterion and references to other refinements. In the above pooling equilibrium, the high-type manufacturer

The following is a separating equilibrium: $w_H = 4$, $w_L = 6 - \sqrt{20} \approx 1.5279$, $\mu(w) = 0$ for all $w \le w_L$, and $\mu(w) = 1$ for all $w > w_L$. To verify, first consider the high-type manufacturer. For any wholesale price w greater than w_L, the retailer thinks $a = 8$ and the manufacturer faces the profit function $\pi_H^M(w) = 4w - w^2/2$, which is maximized at $w = 4$ with $\pi_H^M(4) = 8$. If the manufacturer wants to pretend to be of low type, he must offer a wholesale price $w \le w_L$. Over this range, $\pi_H^M(w) = 6w - w^2/2$, which is an increasing function in w for $w \le w_L$. Therefore, misleading the retailer gives the high-type manufacturer a profit lower than or equal to $6w_L - w_L^2/2 = 8$, which is not better than "truth-telling." This verifies that the optimal choice for the high-type manufacturer is $w_H = 4$. On the other hand, recall from our earlier discussions that the low-type manufacturer will never want to pretend to be of high-type (doing so would give him a negative profit). For wholesale prices $w \le w_L$, the low-type manufacturer's profit function is $\pi_L^M(w) = 2w - w^2/2$. It is easy to see that this function is increasing over $w \le w_L$. Thus the low-type manufacturer's optimal wholesale price is w_L. Moreover, the choices by the two manufacturer types confirm the above retailer belief. We thus have a separating equilibrium.

You may wonder if there exists any other pooling or separating equilibrium. It is a good exercise to try to find one. To read more about signaling games, see, e.g., Kreps (1990), Fudenberg and

---

obtains an expected profit $(8 - p^0)w^0$, where p^0 is the retailer's selling price in the equilibrium, i.e.

$$p^0 = 2(1+\rho) + w^0/2 = 3 + \rho.$$

Likewise, the low-type manufacturer's equilibrium expected profit is $(4-p^0)w^0$. The first step in the test is to identify all signals (i.e., wholesale prices) that are "equilibrium dominated." A wholesale price w is equilibrium dominated if the maximum achievable profit for the manufacturer under w is less than what she gets in equilibrium. Consider the high-type manufacturer. We know her expected profit in equilibrium is $(8 - p^0)w^0$. If she charges wholesale price w and the retailer sets the retail price at p, her profit is $(8 - p)w$. The maximum achievable profit is $8w$, which is obtained when $p = 0$. Thus w is equilibrium dominated at the high type if

$$8w < (8 - p^0)w^0 \;\; or \;\; w < (1 - p^0/8)w^0 = \frac{(5-\rho)(1-\rho)}{4} \stackrel{def}{=} \tilde{w}_H.$$

Similarly, w is equilibrium dominated at the low type if

$$4w < (4 - p^0)w^0 \;\; or \;\; w < (1 - p^0/4)w^0 = \frac{(1-\rho)^2}{2} \stackrel{def}{=} \tilde{w}_L.$$

The central idea of the intuitive criterion is that it should be obvious what the retailer's beliefs should be at wholesale prices that are equilibrium dominated. For example, if the retailer observes $w < \tilde{w}_H$, then the signal must not come from the high-type manufacturer and thus $\mu(w) = 0$. Similarly, a signal $w < \tilde{w}_L$ tells the retailer that the manufacturer cannot be of low type. But can she be of high type in this case? No, because $\tilde{w}_L < \tilde{w}_H$ and thus $w < \tilde{w}_H$. We are in a quandary here; a reasonable assumption is that such a wholesale price will never be observed. Under this assumption, the intuitive criterion suggests that $\mu(w) = 0$ for all $w < \tilde{w}_H$. Notice that $\mu^0(w) = \rho$ for the same range. (Check that $\tilde{w}_H < w^0$.) Let us see if this change in retailer belief will change the manufacturer's signaling strategy, assuming the retailer continues with his optimal response $p(w, \mu(\cdot))$ where $\mu(\cdot)$ is the updated belief. First, consider the high-type manufacturer. If she offers $w < \tilde{w}_H$, then $\mu(w) = 0$ and

$$\pi_H^M(w) = 6w - w^2/2 < 6\tilde{w}_H - (\tilde{w}_H)^2/2$$

which can be shown to be less than 8, which is the manufacturer's expected profit if the wholesale price is 4 (and thus the retailer thinks she is high type), which in turn is less than what she gets by charging w^0. Therefore, the high-type's choice remains intact. Now consider the low-type manufacturer. For any $w < \tilde{w}_H$, we have

$$\pi_L^M(w) = 2w - w^2/2 < 2\tilde{w}_H - (\tilde{w}_H)^2/2.$$

It can be shown that if $\rho < 0.11$ then

$$2\tilde{w}_H - (\tilde{w}_H)^2/2 < (w^0)^2/2$$

where the right-hand side is the low-type manufacturer's expected profit if she chooses w^0 as the wholesale price. Consequently, for $\rho < 0.11$, the pooling equilibrium is sustained by the intuitive criterion.

Tirole (1992), and Kreps and Wilson (1982).

Demand Signaling in New Product Introductions

When a manufacturer introduces a new product to the market, it often possesses some private information about the potential market demand for the product. This information is critical for a retailer who is deciding whether or not to carry the product because the retailer may not be able to recoup the overhead for a low-demand product. Similarly, the information is valuable to a supplier who is considering how much capacity to build for the manufacturer's product; building a lot of capacity for a low-demand product is wasteful. In both cases, the manufacturer has an incentive to report high demand, whether the actual demand is high or low. (This is different than the scenario considered in the above example, where the manufacturer benefits if the retailer thinks the market is low.) As a result, a simple announcement by the manufacturer will not be believed. To be credible, the manufacturer needs to put money where its mouth is, i.e., signaling. Below are several papers that deal with this issue.

Chu (1992) considers a distribution channel consisting of a manufacturer and a retailer. The product, produced by the manufacturer and sold by the retailer, draws a demand that depends on the market condition, the retail price P, and the manufacturer's advertising expenditure A:

$$Q^i = a - b^i P + f(A), \quad i = H, L$$

where Q^i is the demand for the product under market condition i, a is a constant, b^i is the demand sensitivity to price, and $f(\cdot)$ is a concave, increasing function. Assume $b^H < b^L$. Thus, for any given P and A, the demand is higher when the market condition is 'H'. For convenience, we say the market condition is either high ($i = H$) or low ($i = L$). The manufacturer knows the true market condition, whereas the retailer assesses a probability ρ that the market is high (and the manufacturer knows of this assessment). The manufacturer incurs a constant marginal cost of production C.

A signaling game is where the manufacturer (with superior information) moves first by offering a wholesale price P_w and spending A on advertising. Given P_w and A, the retailer updates his belief about the market condition from ρ to $\hat{\rho}$, decides whether or not to carry the manufacturer's product, and if the latter, sets the retail price. The retailer incurs a fixed cost F for carrying the product. The retailer will accept the manufacturer's offer if his expected profit (excluding the carrying cost) exceeds F, and will reject it otherwise.

Chu makes an additional assumption that as soon as the retailer accepts the manufacturer's offer, he sees the true market condition. Therefore, the retail price can be made contingent upon the value of the slope b^i.

An equilibrium for the signaling game consists of a manufacturer strategy $\{P_w^i, A^i\}$, $i = H, L$, a retailer accept/reject strategy $R(x, y)$ that is a binary function, and a retailer belief $\mu(x, y)$, which is the posterior probability that the market is high, for any possible offer $(P_w, A) = (x, y)$ from the manufacturer. Recall that the retailer's pricing decision is made after learning the market condition. Thus if the retailer accepts an offer (x, y) from the manufacturer, the optimal retail price is

$$P^i(x, y) = \text{argmax}_P (P - x)(a - b^i P + f(y)), \quad i = H, L.$$

Let the retailer's profit (excluding the fixed cost F) be $\pi_R^i(x, y)$, i.e., $\pi_R^i(x, y) = (P^i(x, y) - x)(a - b^i P^i(x, y) + f(y))$. Before the accept/reject decision, the retailer's expected profit is $\pi_R(x, y) \stackrel{def}{=}$

$\mu(x,y)\pi_R^H(x,y) + (1-\mu(x,y))\pi_R^L(x,y)$. Therefore, if $\pi_R(x,y) \geq F$, the retailer accepts the manufacturer's offer (x,y), i.e., $R(x,y) = 1$; otherwise, if $\pi_R(x,y) < F$, the retailer rejects the offer (x,y), i.e., $R(x,y) = 0$. Anticipating all this, the manufacturer, knowing her own type, maximizes her profits:

$$(P_w^i, A^i) = \text{argmax}_{P_w,A} R(P_w,A)(P_w - C)(a - b^i P^i(P_w,A) + f(A)), \quad i = H, L.$$

An equilibrium with $(P_w^H, A^H) \neq (P_w^L, A^L)$ is a separating equilibrium; otherwise, if $(P_w^H, A^H) = (P_w^L, A^L) \stackrel{def}{=} (\hat{P}_w, \hat{A})$, we have a pooling equilibrium. For a separating equilibrium, the consistency requirement for the retailer belief is $\mu(P_w^H, A^H) = 1$ and $\mu(P_w^L, A^L) = 0$. For a pooling equilibrium, consistency requires $\mu(\hat{P}_w, \hat{A}) = \rho$.

Chu has identified a separating equilibrium, where the high-type manufacturer advertises and prices above its complete information levels. He has also identified a pooling equilibrium, where both the high-type and the low-type manufacturers advertise and price at or above the complete information levels of the high-type manufacturer. (The complete information case is simply one where the retailer knows the true market condition, and the two players carry out a Stackelberg game with the manufacturer as the leader, setting the wholesale price and the advertising expenditure, and the retailer as the follower, setting the retail price, with an option to reject the manufacturer's offer.)

Chu proceeds to consider the case where the retailer moves first to screen the manufacturer. This is achieved through a slotting allowance, which is a lump-sum payment from the manufacturer to the retailer in order for the latter to carry the product. The game proceeds in the following sequence. The retailer specifies a slotting allowance, which the manufacturer can either accept or reject. If rejected, the game ends with zero profit for both parties. If the manufacturer agrees to pay the slotting allowance, she gets to set the wholesale price and an advertising expenditure. Given these, the retailer then sets the retail price after observing the market condition (as in the signaling case).

Notice that once the manufacturer has accepted to pay the slotting allowance, the rest of the game is the same as in the complete information case, because the retailer sees the market condition (due to screening) before his pricing decision. It is possible to choose a slotting allowance such that only the high-type manufacturer finds it acceptable. For example, let the slotting allowance be the high-type manufacturer's maximum profits in the complete information case. (Recall that this is what the manufacturer can achieve in the Stackelberg game with complete information, where the manufacturer moves first by announcing the wholesale price P_w and the advertising expenditure A, and the retailer follows by setting the retail price P.) It is clear that such a slotting fee is unacceptable to the low-type manufacturer. In this case, only high-type products will be carried by the retailer, who takes all the channel profits.

So, signaling or screening? This is, of course, determined by the balance of power in the channel. Clearly, the manufacturer prefers to move first, by signaling, whereas the retailer prefers to move first too, by screening. From the channel's perspective, the result depends on the effectiveness of advertising.

Consider, for a moment, the complete information case. Compared with the channel-optimal solution, the Stackelberg solution leads to a wholesale price greater than the manufacturer's marginal cost, which in turn leads to a retail price greater than the channel-optimal retail price. This is the well-known double-marginalization phenomenon.[40] On the advertising side, because the manufacturer reaps only part of the benefits from advertising (because the retailer also makes a margin), the advertising level in the Stackelberg solution tends to be lower than the channel-optimal advertising level. In sum, inefficiencies result because the wholesale price is too high and the advertising expenditure is too low.

Now consider the asymmetric information case with a high-type manufacturer. (It is reasonable to ignore the low-type manufacturer if a low-type product is not sustainable.) As mentioned earlier, the signaling game leads to a wholesale price and an advertising level both higher than those in the Stackelberg solution (with complete information). Therefore, signaling increases the wholesale-price distortion (further away from the channel optimum) but decreases the advertising distortion. When advertising has low effectiveness, the former effect dominates the latter, leading to a channel profit even lower than in the Stackelberg solution. In contrast, screening with a slotting allowance restores the channel profit to the Stackelberg-game level, because the slotting allowance, being a fixed fee, does not alter the pricing and advertising decisions in the channel. Therefore, one can say that signaling involves wasteful expenditures, whereas screening keeps the money in the channel. On the other hand, if advertising is highly effective, then the channel may be better off with signaling.

Lariviere and Padmanabhan (1997) further investigate the role of slotting allowances in new product introductions. Suppose, as before, a manufacturer introduces a new product through an independent retailer. The manufacturer begins by offering the terms of trade, consisting of a wholesale price w and a slotting allowance A. The retailer then either accepts or rejects the terms. If the former, the retailer agrees to carry the product and proceeds to set a retail price p and exert merchandising effort e. The quantity sold can be expressed as

$$D(e,p) = \tau + f(e) - \beta p$$

where τ is a market-size parameter, β measures demand sensitivity to price, and $f(\cdot)$ is an increasing, concave function. The cost of merchandising effort is assumed to be e as well, i.e., a linear effort-cost model. The retailer incurs a fixed cost K for carrying the product. He accepts the contract offered by the manufacturer if and only if his profit is nonnegative. A key feature of the model is that the two players possess asymmetric information about the market size. It is assumed that τ takes one of two possible values H and L with $H > L$. The manufacturer knows the value of τ, whereas the retailer assesses a probability θ that $\tau = H$ (and the manufacturer knows about this assessment). As in all signaling games, the retailer may infer something about the value of τ from the terms of trade offered (w, A), i.e., forming a posterior belief $\mu(w, A)$ that $\tau = H$.

Lariviere and Padmanabhan have characterized a separating equilibrium in the above signaling game. The equilibrium consists of a contract offered by the manufacturer, $(\hat{w}, \hat{A})$, and a supporting retailer belief $\mu(\cdot,\cdot)$ such that $\mu(\hat{w}, \hat{A}) = 1$ and $\mu(w, A) = 0$ for all $(w, A) \neq (\hat{w}, \hat{A})$. The parameters of the model are such that if the market is low, it is impossible for the manufacturer and the retailer to make nonnegative profits at the same time. Therefore, in any separating equilibrium where the retailer learns the true type of the manufacturer, only the high-type product may be accepted by the retailer.

Let us identify the constraints that a separating equilibrium must satisfy. Suppose the manufacturer offers (w, A) that leads the retailer to believe that the market is high. The retailer's profit

[40] Spengler (1950) is the first to discuss this phenomenon.

function (excluding fixed costs) is thus

$$\pi_R(e,p) = (p-w)(H+f(e)-\beta p) + A - e.$$

The retailer's optimal response is thus $(\tilde{e},\tilde{p}) \stackrel{def}{=} \text{argmax}_{(e,p)}\pi_R(e,p)$. Note that the slotting allowance, since it is fixed, does not affect the retailer's pricing and effort decisions. But it certainly affects whether or not the retailer will accept the manufacturer's offer. Acceptance results only if

$$\pi_R(\tilde{e},\tilde{p}) \geq K.$$

In order for the belief to be correct in equilibrium, it must be unprofitable for the low-type manufacturer to mimic the high-type. Suppose the low-type manufacturer offers (w, A), the contract offered by the high-type. As a result, the retailer believes $\tau = H$ and thus responds by choosing $(\tilde{e}, \tilde{A})$. The manufacturer's profit is thus:

$$\pi_M^L(w,A) = (w-c)(L+f(\tilde{e})-\beta\tilde{p}) - A$$

where c is the manufacturer's marginal production cost. Doing so must be unprofitable for the low-type manufacturer, i.e. $\pi_M^L(w,A) \leq 0$. On the other hand, the high-type manufacturer's profit is

$$\pi_M^H(w,A) = (w-c)(H+f(\tilde{e})-\beta\tilde{p}) - A.$$

The high-type manufacturer seeks a separating equilibrium that maximizes her profit by solving the following optimization problem:

$$\begin{aligned} \max_{(w,A)} \quad & \pi_M^H(w,A) \\ s.t. \quad & (\tilde{e},\tilde{p}) = \text{argmax}_{(e,p)}\pi_R(e,p) \\ & \pi_R(\tilde{e},\tilde{p}) \geq K \\ & \pi_M^L(w,A) \leq 0. \end{aligned}$$

The solution $(\hat{w}, \hat{A})$ is the contract the high-type manufacturer will offer in equilibrium.

The following results have been obtained. First, when the fixed cost K is lower than a threshold level K^*, the separating equilibrium does not involve any slotting allowance, i.e., $\hat{A} = 0$, and the wholesale price $\hat{w}$ is greater than the full information wholesale price (for high market condition).[41] Second, when $K \geq K^*$, the separating equilibrium requires a positive slotting allowance, i.e., $\hat{A} > 0$, and a wholesale price $\hat{w}$ that is still greater than the full information price. The main message here is that slotting allowances can be a useful signaling device to show to the retailers that the products the manufacturers are introducing are promising. This is in contrast with Chu (1992), who models slotting allowances only as a screening device.

Desai and Srinivasan (1995) study demand signaling in the presence of moral hazard. A principal sells a product through an independent agent. The product is new, and the principal has private information about the demand for the product.[42] The agent can also influence the demand by expending selling effort, which is not observable to the principal. Both the principal and the agent are risk neutral. The issue is the contract emerging between the two parties in this two-sided information asymmetry model.

[41] The full information case is where the retailer also observes the value of τ. In this case, due to the assumptions made about the model parameters, the only relevant problem is what faces the high-type manufacturer. This problem is solved in a Stackelberg fashion with the manufacturer being the leader and the retailer the follower.

[42] The model can be interpreted in other ways. We choose this principal-agent, new-product story for convenience.

More specifically, the demand function is either Q^H or Q^L with

$$Q^J = T^J - p + f(a) + \epsilon, \quad J = H, L$$

where $T^H > T^L$. The principal knows the true demand function, whereas the agent remains uncertain. For convenience, if the demand function is Q^H, we say the principal is of high-type; otherwise, the principal is of low-type. The agent's prior belief is that the principal is of high-type with probability ρ, with $0 < \rho < 1$. It is assumed that both types of principals have the same marginal cost c. The agent determines the selling price p and selling effort a, whose impact on sales is captured by an increasing concave function $f(\cdot)$. The cost of effort (to the agent) is $w(a)$, a convex increasing function. The demand is subject to a random shock ϵ with mean zero. Although the realized demand is observable and contractible, the selling effort is not.

The sequence of events is as follows. The principal offers a contract to the agent; the contract specifies a payment from the agent to the principal that can be a function of the realized demand. Given a contract, the agent then updates his belief about the principal's type, from his prior belief ρ to a posterior belief $\hat{\rho}$. Based on this updated belief, the agent chooses a selling price and an effort level to maximize his expected profit. The principal, anticipating the agent's behavior, selects a contract that maximizes her expected profit, under the constraint that the expected profit for the agent is at least $\bar{u}$, a reservation utility level reflecting the agent's outside opportunities.

Desai and Srinivasan, citing Arrow (1985) that simple linear contracts are prevalent in practice, start with a simple, two-part contract with a fixed fee F and a variable fee r, so that the agent pays the principal $F + rq$ if the realized demand is q. They then consider a nonlinear contract with a quadratic component.

First, notice that although the model involves private information on both sides, it is crucial to have the private demand information (held by the principal). To see this, let us first consider the first-best scenario where both parties have full information, i.e., the agent also knows the true demand function, and the principal can observe the agent's selling effort. In this case, the principal can demand a specific level of effort. Given a contract (r, F, a) of variable fee, fixed fee, and effort level, the agent chooses a retail price to maximize his expected profit. This is an easy problem. In the solution, the principal sets the variable fee r equal to her marginal cost c (to avoid double marginalization), the agent earns exactly his reservation utility, and the principal extracts all the surplus (via the fixed fee). It can be verified that the first-best effort level and fixed fee are higher with the high-type principal than with the low-type principal.

Now suppose the agent's selling effort is unobservable to the principal (but the agent still knows the true demand function). The solution to this problem is also well known. From agency theory, since the agent is risk neutral, the moral hazard problem can be easily overcome by making the agent a residual claimant. For either type of principal, simply set the variable fee to the marginal cost and the fixed fee to make the agent's participation constraint binding. The first-best solution prevails.

The above discussions make it clear that if the principal did not hold private demand information, the problem would be trivial. From now on, we focus on the case with asymmetric demand information. To understand the impact of moral hazard on demand signaling, Desai and Srinivasan first consider the signaling game without moral hazard, i.e., assuming the principal can observe the agent's selling effort. The signaling instruments are therefore (r, F, a). They have identified a separating equilibrium in which the high-type principal offers (r^S, F^S, a^S), whereas the low-type principal offers her corresponding first-best contract. It is shown that $r^S > r^H_{fb}$, $F^S < F^H_{fb}$, and $a^S = a^H_{fb}$, where r^H_{fb}, F^H_{fb} and a^H_{fb} are the variable fee, fixed fee, and effort level in the high-type principal's first-best solution. The somewhat unintuitive result is that it is unnecessary for the high-type principal to distort the effort level (from the first-best level) in order to credibly convey its type; a distortion in the variable fee suffices. (Since the variable fee is raised, the fixed fee must be reduced to meet the agent's participation constraint.)

If the agent's selling effort is unobservable to the principal (thus not contractible)—as in the original case with two-sided information asymmetry—the principal's signaling instruments are reduced to (r, F). A separating equilibrium for this game has also been identified, where the high-type principal offers (r^{SM}, F^{SM}), and the low-type principal continues to offer her first-best contract. It has been shown that $r^H_{fb} < r^{SM} < r^S$ and $F^S < F^{SM} < F^H_{fb}$. This leads to a main conclusion of the paper that moral hazard dampens the signaling distortions of the fixed and variable fees. Moreover, the high-type principal makes less profit when the agent's effort is unobservable.

Interestingly, when we allow nonlinear contracts, the first-best solution can be achieved for both types of principals even in the presence of moral hazard. This is established under a three-part contract where the agent's payment to the principal is $F + r_1 q + r_2 q^2$, with q being the realized demand. A separating equilibrium has been identified where both types of principals earn their first-best profits. The above conclusion is obtained under the following additional assumptions: $\bar{u} = 0$, $c = 0$, $f(a) = a$, and $w(a) = a^2$. It is worth noting that in equilibrium, the two variable fees are such that $r_1 > 0$ and $r_2 < 0$. In other words, the first-best is achieved not by making the agent a residual claimant. This is in contrast with the way in which the first-best is achieved when the principal does not hold private information about demand.

The above three papers on demand signaling focus on a manufacturer's task of conveying her private demand information to a retailer, a downstream supply-chain partner. We next consider the problem of signaling to an upstream supply-chain partner.

In the relationship between a manufacturer and her supplier, it is often the case that the manufacturer knows more about the demand (for the end product) than the supplier does. Suppose the supplier is always the bottleneck of the supply chain, i.e., production is constrained only by the supplier's capacity. From the manufacturer's standpoint, the more capacity the supplier installs the better (a higher capacity relaxes the constraint on production). Therefore, the manufacturer has an incentive to inflate her demand forecast, hoping the supplier will increase his capacity. But, of course, the supplier knows this and will not take the manufacturer's demand forecast at face value. The result is a communication breakdown leading to an efficiency loss. Is there a way for the manufacturer to reveal her demand forecast in a credible way?

Cachon and Lariviere (2001) provide a model to address the above question. A manufacturer sells a single product that has uncertain demand, D. The manufacturer relies on a single supplier for a critical component of the product. Let K be the supplier's capacity, which must be built before demand realization. Some demand may be lost due to the capacity constraint, with $\min\{D, K\}$ being the demand fulfilled and the rest lost. The supplier incurs a cost for every unit of capacity built. A key feature of the model is that the manufacturer possesses private information about the demand. It is assumed that $D = \theta X$, where θ and X are independent random variables. Moreover, θ has only two possible values, H and L with $H > L$. The parameter θ may be interpreted as an indicator of the market condition. The manufacturer observes the value of θ, whereas the supplier assesses a probability ρ that $\theta = H$ and $1 - \rho$ that $\theta = L$. (Both players have the same information about X, i.e., its distribution.) For convenience, the manufacturer observing $\theta = H$ is said to be of high-type, whereas if $\theta = L$, low-type. The manufacturer moves first by offering a contract to the supplier, the supplier builds capacity given the contract and his belief about demand, demand is then realized, and finally, production takes place (within the capacity set by the supplier). This is a signaling game because the informed party (i.e., the manufacturer) takes a more active role by designing the contract.

Let Z be the set of all admissible contracts. Both the supplier and the manufacturer agree that only contracts in Z are considered. (Cachon and Lariviere consider linear contracts with commitments and options.) The supplier's capacity decision depends on the contract offered by the manufacturer and his belief about the market condition. Let $z \in Z$ be the contract offered, and b the supplier's belief, i.e., b is a probability distribution for the values of θ. Note that b might be a function of z. Let $\pi(K, z, b)$ be the supplier's expected profit if he builds capacity K. The optimal capacity level, from the supplier's standpoint, is then $K(z, b) = \text{argmax}_K \pi(K, z, b)$. Consequently, the manufacturer's expected profit is a function of her type t, the contract offered to the supplier z, and the supplier's belief b about the market condition, i.e., $\Pi_t(z, b)$, $t = H, L$.

We focus on separating equilibria, whereby the supplier learns the manufacturer's type upon seeing the contract offered by the manufacturer. That is, given a contract z, the supplier assigns probability one to either $\theta = H$ or $\theta = L$. For convenience, denote the former by $b = H$ and the latter $b = L$. Therefore, the supplier's belief can be characterized by a partition of Z, i.e., Z_H and Z_L with $Z = Z_H \cup Z_L$ and $Z_H \cap Z_L = \emptyset$. If a contract $z \in Z_H$ is offered, $b = H$; otherwise if $z \in Z_L$ is offered, $b = L$. In equilibrium, the supplier's belief must be correct, i.e.

$$\max_{z \in Z_L} \Pi_L(z, L) \geq \max_{z \in Z_H} \Pi_L(z, H) \quad \text{and} \quad \max_{z \in Z_H} \Pi_H(z, H) \geq \max_{z \in Z_L} \Pi_H(z, L).$$

In words, given the supplier's belief (i.e., a partition of Z), a type-t manufacturer has no incentives to mislead the supplier into believing that she is of type-t', $t \neq t'$. A partition or belief with the above properties is called an equilibrium partition or equilibrium belief. Corresponding to an equilibrium partition, Z_H and Z_L, is a pair of contracts, z_H and z_L, that are offered by the high- and low-type manufacturer, respectively. Clearly,

$$z_t = \text{argmax}_{z \in Z_t} \Pi_t(z, t), \quad t = H, L.$$

A separating equilibrium, then, consists of an equilibrium partition, Z_H and Z_L, and a corresponding pair of contracts, z_H and z_L.

Now assume that

$$\Pi_t(z, H) \geq \Pi_t(z, L), \quad \forall z \in Z, \quad t = H, L. \tag{4}$$

In words, for any contract, the manufacturer, regardless of her type, is better off if the supplier believes that the market condition is high rather than low. This seems plausible because a belief of high market condition leads the supplier to build more capacity, to the benefit of the manufacturer. Below, we characterize a separating equilibrium under the above condition.

Let z_L^* be the optimal contract for the low-type manufacturer in the full-information case, i.e.,

$$z_L^* = \text{argmax}_{z \in Z} \Pi_L(z, L).$$

Assume that the above optimization problem has a unique solution; thus $\Pi_L(z_L^*, L) > \Pi_L(z, L)$ for all $z \neq z_L^*$. It is easy to establish that any equilibrium partition (Z_H, Z_L) has $z_L^* \in Z_L$. The proof is by contradiction. Suppose $z_L^* \in Z_H$. Then $\Pi_L(z_L^*, H) \geq \Pi_L(z_L^*, L) > \Pi_L(z, L)$ for all $z \in Z_L$, where the first inequality follows from (4). Thus the low-type wants to pretend to be the high-type. Thus (Z_H, Z_L) is not an equilibrium partition, a contradiction.

Now take any equilibrium partition (Z_H, Z_L). Let (z_H, z_L) be the corresponding pair of contracts offered by the two manufacturer types. From the above discussion, $z_L = z_L^*$. Also, by the definition of equilibrium partitions, we have

$$\Pi_L(z_H, H) \leq \Pi_L(z_L^*, L) \tag{5}$$

and

$$\Pi_H(z_H, H) \geq \max_{z \in Z_L} \Pi_H(z, L). \tag{6}$$

Also by the definition of z_H and (4), for any $z \in Z_H$, $\Pi_H(z_H, H) \geq \Pi_H(z, H) \geq \Pi_H(z, L)$. This inequality, together with (6), implies

$$\Pi_H(z_H, H) \geq \max_{z \in Z} \Pi_H(z, L). \tag{7}$$

Therefore, the high-type manufacturer's expected profit, i.e., $\Pi_H(z_H, H)$, is no greater than the maximum value of the objective function in

$$\begin{aligned} \max_{z \in Z} \quad & \Pi_H(z, H) \\ s.t. \quad & \Pi_L(z, H) \leq \Pi_L(z_L^*, L) \\ & \Pi_H(z, H) \geq \max_{z' \in Z} \Pi_H(z', L) \end{aligned}$$

where the first and second constraints are from (5) and (7) respectively. Suppose this problem has a unique solution $z = z_H^*$. Define $Z_H^* = \{z_H^*\}$ and $Z_L^* = Z \setminus Z_H^*$. It is easy to verify that (Z_H^*, Z_L^*) is an equilibrium partition. The contract offered by the high-type (resp., low-type) manufacturer is z_H^* (resp., z_L^*) with the corresponding expected profit $\Pi_H(z_H^*, H)$ (resp., $\Pi_L(z_L^*, L)$). Thus $(Z_H^*, Z_L^*, z_H^*, z_L^*)$ is a separating equilibrium. Note that from the above arguments, both manufacturer types can do no better than this with any other equilibrium partition.

While the supplier in Cachon and Lariviere (2001) is the only source of supply for the manufacturer, Van Mieghem (1999) considers a setting with two sources of supply: the manufacturer's in-house production facility and an outside supplier (i.e., subcontractor). At the beginning of the game, the manufacturer and the subcontractor simultaneously and independently make their capacity investment decisions, i.e., the manufacturer decides how much in-house capacity K_M to build and the subcontractor decides on its own capacity K_S. The manufacturer faces market demand D_M, which can be served by in-house as well as the subcontractor's production. Moreover, the subcontractor can also sell its product to a separate market with demand D_S. At the time of the capacity decisions, only a joint probability distribution of D_M and D_S is known (to both players). After the capacity decisions, the demands are realized and the firms decide on their production/sales quantities. More specifically, the manufacturer determines how much to produce in-house (i.e., x_M) and how much to order from the subcontractor (i.e., x_t^M); the subcontractor decides how much to sell to its own market (i.e., x_S) and how much of the manufacturer's order to fill (i.e., x_t^S). Of course, $x_M \leq K_M$, $x_M + x_t^M \leq D_M$, $x_t^S \leq x_t^M$, $x_t^S + x_S \leq K_S$, and $x_S \leq D_S$. These production/sales decisions are governed by various contractual arrangements between the two firms. In one contract (called the price-only contract), a transfer price p_t is specified ex ante for each unit supplied by the subcontractor. (This price is known before the capacity decisions.) In this scenario, the two firms sequentially solve for their production/sales quantities, with the manufacturer as the first mover. Under another contractual arrangement, there are simply no ex-ante contracts and the parties negotiate the transfer quantity and price after the demands are realized. In this scenario, the firms arrive at production/sales quantities that maximize their total profits and the surplus (relative to a scenario with no transactions between the two) is split between them based on their relative bargaining power, which is captured by an exogenous index. Van Mieghem also considers state-dependent contracts: a price-only contract with p_t being a function of the installed capacities and the realized demands, or an incomplete contract/bargaining arrangement where the bargaining-power index is state-dependent. For each of these contractual arrangements, one can solve the resulting two-stage stochastic game and examine a contract's impact on the coordination of both capacity and production/sales decisions. Although information sharing is not the focus of Van Mieghem (1999) (in fact, there is no information asymmetry), the paper does provide an interesting discussion on coordinating capacity and quantity decisions in a manufacturer-subcontractor supply chain. Among the key findings are 1) a higher transfer price can actually increase the manufacturer's profit and 2) only state-dependent contracts (price-only or incomplete contract) can coordinate both the quantity and capacity decisions. It remains an interesting open question as to the impact of information asymmetries (about demands, capacities, costs, etc.) on the manufacturer-subcontractor supply chain.

3.3 Information Sharing in Competitive Environments

We begin by considering papers that deal with information sharing among horizontal competitors, e.g., competing retailers sharing market demand information. These papers are all published in economics journals in the 1980s. (Are economists tired of this problem? Extensions to supply chains may breathe new life.) Recently, there have been several attempts to generalize the horizontal information-sharing literature to vertical information sharing in supply chains, e.g., will competing retailers share their demand information with their common supplier? We will discuss these papers as well. [43]

A Simple Example

We begin with a simple example to illustrate the incentives of sharing demand information

[43] A somewhat related paper is Lee and Whang (2002) who consider a supply chain with one supplier and n retailers. These supply chain members are independent firms. The retailers face a selling season with two periods. The retailers independently make their inventory decisions at the beginning of the first period by ordering from the supplier. The retailers are able to adjust their inventory levels after the first-period demands are realized by trading among themselves through a secondary market. They study the impact of the secondary market on the supply chain. The paper is related if one considers the secondary market as an institution that facilitates the sharing (more precisely, the aggregation) of the information embodied in the first-period demands (about the needs for inventories at the beginning of the second period). In this respect, Mendelson and Tunca (2001) is also related. Another paper that touches upon information sharing in a competitive environment is Anupindi and Bassok (1999), who study a decentralized supply chain with one manufacturer and two retailers. The retailers order inventories from the manufacturer, and their ordering decisions are inter-dependent because of consumer market search, i.e., there is a positive probability that a consumer who experiences a stockout at one retailer will look for inventory at the other retailer. It is clear that market search provides incentives for one retailer to increase its inventory (to capture the benefits of demand spillovers), given the other retailer's ordering decision; and such incentives are stronger when search becomes easier. The consequence is that the manufacturer sees an increase in total retailer order quantities as market search increases, which can be facilitated by the installation of an information system that makes the inventory status at the retailers visible to the consumers.

between two competitors.[44] Consider a duopoly facing stochastic linear demand, $p = a - Q$, where p is the market price, a is a random variable, and Q is the total output of the duopolists. Assume that $a = 50, 150$ with equal probabilities. Also, the marginal production costs are zero for both firms. The two firms are engaged in a Cournot competition, i.e., making quantity decisions independently. Before their quantity decisions, firm 1 observes the true value of a, while firm 2 does not receive any private signal about a. Consider a two-stage game. In the first stage, firm 1 decides whether or not to share its information about a with firm 2. Information sharing, if it occurs, is assumed to be truthful. Firm 1 then observes the true value of a. Information transmission takes place according to the agreement reached in the first stage.[45] In the second stage, the firms make quantity decisions based on their information about a.

First, suppose firm 1 has decided to conceal its information about a. The second-stage game is thus a Bayesian game (with incomplete information). Firm 1's strategy is a decision rule that specifies a quantity for each signal it may receive. Let firm 1's output be q_1^h if it observes $a = 150$ and q_1^l if it observes $a = 50$. Let q_2 be firm 2's output. Firm 2's best response to firm 1's strategy is obtained by solving:

$$\max_{q_2} \frac{1}{2}(50 - q_1^l - q_2)q_2 + \frac{1}{2}(150 - q_1^h - q_2)q_2.$$

That is,

$$q_2 = \frac{200 - q_1^l - q_1^h}{4}. \tag{8}$$

Similarly, firm 1's best strategy against firm 2's quantity is

$$q_1^l = \frac{50 - q_2}{2}, \text{ and } q_1^h = \frac{150 - q_2}{2}. \tag{9}$$

Solving (8) and (9) gives a Bayesian Nash equilibrium: $(q_1^l, q_1^h, q_2) = (\frac{25}{3}, \frac{175}{3}, \frac{100}{3})$. The expected profits for firm 1 and firm 2 are: $(\pi_1^{ns}, \pi_2^{ns}) = (\frac{15{,}625}{9}, \frac{10{,}000}{9})$, where the superscript 'ns' stands for 'no sharing of information.'

Now suppose firm 1 has decided to share its information with firm 2. If $a = 50$, the two firms each produce 50/3; otherwise, if $a = 150$, they each produce 50. The expected profits are $(\pi_1^s, \pi_2^s) = (\frac{12{,}500}{9}, \frac{12{,}500}{9})$.

Comparing the above two scenarios, we see that it is to the interest of the informed firm to conceal its information. Note that the total profits of the two firms are also higher with no information sharing. This is a pretty gloomy picture for information sharing.

A body of literature in economics is devoted to the investigation of information sharing among horizontal competitors. It turns out that whether or not it is optimal to share information depends on many things, including the type of competition (Cournot or Bertrand), the type of information (e.g., common demand information or private cost information), and whether or not the products sold by the competitors are substitutes or complements.

[44] This example was given by Gal-Or (1985).

[45] If firm 1 first observes a and then decides whether or not to share the information with firm 2, we have a very different model. It looks like a signaling problem, with retail competition as a subgame.

Duopoly with Demand Information: Cournot and Bertrand, Substitutes and Complements

Vives (1984) considers the following duopoly model. Two firms, each producing a differentiated good, face the following inverse demand functions:

$$p_i = \alpha - \beta q_i - \gamma q_j, \;\; i, j = 1, 2, \; j \neq i$$

where q_i are the quantities of the goods and p_i their prices, with $|\gamma/\beta| \leq 1$. If $\gamma = 0$, the goods are independent and the firms are local monopolists. If $\gamma > 0$, the goods are substitutes. If $\gamma < 0$, they are complements. Both firms have constant and equal marginal costs, which are normalized to zero. The two firms are engaged in either Cournot competition where they compete in quantities or Bertrand competition where they compete in prices.

Note that profits of firm i are given by $\pi_i = p_i q_i$. Since π_i is symmetric in p_i and q_i and the demand curves are linear, Cournot (resp., Bertrand) competition with substitutes has similar strategic properties as Bertrand (resp., Cournot) competition with complements.

The common demand intercept, α, is a normally distributed random variable with known mean and variance. It is assumed that the firms employ an "independent testing agency" to collect samples of α. Through the agency, firm i has contracted for n_i observations: $(r_{i1}, \cdots, r_{in_i})$, where $r_{ik} = \alpha + u_{ik}$, and the u_{ik}'s are i.i.d. normal random variables with zero mean and variance σ_u^2 and they are independent of α. Moreover, firm i has instructed the agency to put the first m_i observations that it has contracted for in a common pool, available for the other firm. Therefore, firm i's best (minimum variance unbiased) estimate of α based on $n_i + m_j$, $j \neq i$, observations is

$$s_i = \alpha + \frac{\sum_{k=1}^{n_i} u_{ik} + \sum_{k=1}^{m_j} u_{jk}}{n_i + m_j}.$$

If $m_1 = m_2 = 0$, then there is no sharing of information. On the other hand, if $m_1 = n_1$ and $m_2 = n_2$, there is a complete sharing of information.

The firms play a two-stage game. First, they decide how much information to put in the common pool, i.e., choosing m_1 and m_2 independently. (n_1 and n_2 are not decision variables in the model.) The values of n_i and m_i, $i = 1, 2$, are common knowledge. The agency then collects independent observations of α and distributes the information according to the agreement reached in the first stage (i.e., transmitting some information to a firm privately and some to the common pool). At the second stage, the firms independently choose their quantities in Cournot competition or prices in Bertrand competition. Each pair of (m_1, m_2) defines a subgame with incomplete information, which can be solved by using the concept of Bayesian Nash equilibrium.

Here are some results obtained by Vives. In Cournot competition with substitutes (or Bertrand competition with complements), expected profits of firm i decrease with m_i. So no information sharing is the unique equilibrium. In Cournot competition with complements (or Bertrand competition with substitutes), expected profits of firm i increase with m_i and with m_j, $j \neq i$. So complete information sharing is the unique equilibrium. If the goods are independent, expected profits of firm i are increasing in m_j and unaffected by m_i, $j \neq i$. In this case, any pair of (m_1, m_2) is an equilibrium.

Oligopoly with Demand Information: Cournot, Substitutes

Gal-Or (1985) provides an alternative model of information sharing in an oligopoly model with Cournot competition. There are n firms producing a common product at no cost. The industry is facing a linear demand function:

$$p = a - bQ + u, \quad a, b > 0$$

where p is price and Q is the total quantity produced. The prior distribution of u is normal with zero mean and a finite variance. Before making their quantity decisions, each firm receives a noisy signal for u. The signal observed by firm i is x_i. The following assumptions characterize the signals:

$$\begin{aligned}
&x_i = u_i + e_i, \;\; u_i \sim N(0, \sigma), \;\; e_i \sim N(0, m); \\
&Cov(e_i, e_j) = 0, \;\; i \neq j; \\
&Cov(u_i, e_j) = 0, \;\; \forall i, j; \\
&Cov(u_i, u_j) = h, \;\; i \neq j; \\
&u = \sum_{i=1}^{n} u_i/n.
\end{aligned}$$

Furthermore, it is assumed that $h \geq 0$, a parameter that measures the (positive) level of correlation among the signals.

Before making quantity decisions, the firms choose whether or not to reveal their private signals to the other firms, and how complete this revelation will be. This is modeled by assuming that an outside agency is responsible for information transmission. The firms are required to commit themselves to a fixed amount of garbling prior to learning their signals. Upon learning its private signal, each firm i reports its private signal x_i to the agency, who then reports a message $\hat{x}_i$ to the other firms, with

$$\hat{x}_i = x_i + f_i, \;\; f_i \sim N(0, s_i)$$

where the f_i's are independent of each other and of any u_j and e_j, $j = 1, \cdots, n$. The value of s_i represents the level of garbling. If this noise variance is zero for all firms, we have complete information sharing. If it is infinite for all firms, there is no sharing of information. The case with a finite noise variance represents partial information sharing.

Gal-Or characterizes a symmetric equilibrium in the following two-stage game. (A symmetric equilibrium is reasonable because the firms have symmetric cost/information structure.) At the first stage, each firm i chooses s_i independently. Once chosen, this vector of noise variances becomes common knowledge. The firms then receive their private signals, and the outside agency reports messages with levels of garbling determined in the first stage. At the second stage, the firms make their quantity decisions simultaneously. Each firm i's strategy is a decision rule that determines its output level as a function of its private signal x_i and the vector of messages reported by the outside agency. The conclusion is that no information sharing is the unique symmetric Nash equilibrium. Therefore, allowing for partial revelation and various degrees of correlation between the private signals does not change the incentives for sharing demand information. Similar results have been obtained by Clarke (1983).

Duopoly with Cost Information: Cournot and Bertrand, Substitutes

Gal-Or (1986) shows that the incentives for sharing information depend not only on the type of competition (Cournot or Bertrand) and the relationship between the products (substitutes or complements), as Vives (1984) has shown, but also on the type of information under consideration. The private signals received by Gal-Or's firms are about unknown private costs, instead of an unknown demand intercept as in previous works. She considers a duopoly model consisting of two firms each producing a differentiated product. The demand is linear, as in Vives (1984), with an additional assumption that $\gamma > 0$, i.e., the products are substitutes. The production costs are linear, with c_i the unit cost of production for firm i, $i = 1, 2$. The value of c_i is a normal random variable with zero mean and a known variance, with c_1 and c_2 being independent. Each firm receives a signal for its own unit cost. Firm i receives signal z_i, where

$$z_i = c_i + e_i, \quad i = 1, 2$$

where $e_i \sim N(0, m)$, e_i and c_j are independent for any i and j, and e_1 and e_2 are independent.

Information sharing is implemented by an outside agency. Prior to receiving their private signals, each firm commits to a level of garbling that the agency will use in reporting the private information. The reported signal is

$$\hat{z}_i = z_i + f_i, \quad f_i \sim N(0, s_i), \quad i = 1, 2$$

where f_1 and f_2 are independent, and f_i is independent of c_j and e_j for all i and j. As in Gal-Or (1985), the values of the s_i's are determined independently by the firms at the first stage of the game, and they represent the degree of information sharing. At the second stage of the game, the firms choose their output levels (in Cournot competition) or prices (in Bertrand competition). The second-stage strategy is a decision rule based on available information, i.e., z_i and $\hat{z}_j$, $j \neq i$, for firm i.

The main finding is that complete (resp., no) information sharing is a dominant strategy in Cournot (resp., Bertrand) competition. Notice that changing the type of information (from demand to cost) reverses the incentives for information sharing. A similar set of results was obtained by Shapiro (1986).[46]

Vertical Information Sharing in the Presence of Horizontal Competition

Consider a supply chain with multiple parties and dispersed information. Suppose a subset of those parties, called the insiders, have decided to share information among themselves. What is the impact of such an agreement on the insiders, the outsiders (i.e., those who do not engage in information sharing), and the supply chain as a whole? The answer to this question is complicated partly due to the spillover effect of information sharing. That is, an outsider may gain valuable information from the insiders either directly if the insiders fail to keep the shared information confidential or indirectly by observing the actions taken by the insiders. The former kind of information leakage may be prevented by the insiders through a contract protecting the confidentiality of the shared data, while the latter kind is impossible to avoid as long as the shared information affects the behavior of an insider that is observable to the outsiders. The outsiders may change what they do

[46] Li (1985) generalizes the above literature on the incentives for sharing demand or cost information in Cournot oligopolies by making weaker distributional assumptions about the random variables.

upon learning the information, with potentially significant impact on the insiders' profits and their decisions whether or not to share information in the first place.

Li (2002) considers a model with one manufacturer and n symmetric retailers ($n \geq 2$). The retailers sell an identical product, which is produced by the manufacturer. Both production and sales incur constant marginal costs, which are normalized to zero. The consumer market (to which the retailers sell) is characterized by the inverse demand function, $p = a + \theta - Q$, i.e., the prevailing retail price p is determined by a known constant a, a random variable θ, and the total supply Q, which is the sum of the individual quantities given by the retailers. (Thus the retailers engage in a Cournot competition.) The manufacturer determines the wholesale price P. Each retailer i receives a private signal Y_i about θ, with the joint distribution of $(\theta, Y_1, \cdots, Y_n)$ being common knowledge.

The sequence of events is as follows:

1. Each retailer decides whether or not to share his private signal with the manufacturer. If a retailer decides to share, the information revelation is assumed to be truthful. Let K be the set of retailers who decide to share their information. Because the retailers are symmetric, we only need to know the cardinality of K, i.e., $|K| \stackrel{def}{=} k$, $k = 0, 1, \cdots, n$.
2. Each retailer receives his private signal. Information transmission occurs according to the arrangements made in the first step.
3. The manufacturer sets the wholesale price. The wholesale price P is thus a function of the disclosed information $\{Y_j, j \in K\}$.
4. The retailers simultaneously choose their sales quantities and place orders with the manufacturer. Retailer i's strategy thus depends on whether or not i is a member of K, his signal Y_i, the wholesale price P, and the information embedded in the wholesale price P (which is observable to all).
5. The manufacturer produces the retailer orders.

Li shows that there is an equilibrium outcome where P is a monotone function of $\sum_{j \in K} Y_j$. Thus in equilibrium, the retailers $i \notin K$ (i.e., the outsiders) can infer the value of $\sum_{j \in K} Y_j$. Moreover, knowing this sum is as good as knowing the individual signals $Y_j, j \in K$. Therefore, the leakage of the private signals from the insiders to the outsiders is complete. In other words, even though the information sharing is between the retailers in K and the manufacturer, we could as well imagine that the retailers in K announce their private signals in public.

Suppose that k retailers have decided to share information with the manufacturer, $k = 0, 1, \cdots, n$. Let $\Pi_R^S(k)$ be the expected profits for a retailer who shares information, and $\Pi_R^N(k)$ the expected profits for a retailer who does not share information. Let $\Pi_M(k)$ be the manufacturer's expected profits. Li shows that $\Pi_M(k)$ is increasing and concave in k. Therefore, the manufacturer always benefits if a retailer decides to share information. On the other hand, $\Pi_R^N(k-1) > \Pi_R^S(k)$ for all $k = 1, \cdots, n$. In words, a retailer is always better off by switching from sharing information to not sharing. Consequently, no information sharing is the unique equilibrium outcome.

If the manufacturer's gains from information sharing exceed the losses of the retailers, the manufacturer can pay the retailers for their private information. Let $\Pi(k)$ be the supply chain's total profit when k retailers share information, $k = 0, 1, \cdots, n$. Thus $\Pi(k) = \Pi_M(k) + k\Pi_R^S(k) + (n-k)\Pi_R^N(k)$.

Li shows that $\Pi(n) \geq \Pi(0)$ if and only if $(n-2)(n+1) \geq 2s$, where s is an indicator of the informativeness of the retailers' private signals, with a smaller s value meaning more informative. Consequently, there is no guarantee that the supply chain will benefit from information sharing. In cases where the supply chain does benefit from information sharing (when the number of retailers is large or the demand signals are informative), there exists a Pareto improvement if the manufacturer pays the retailers for sharing their information.

Li has also considered a case where the retailers hold private information about their costs. This is done by modifying the above model with demand uncertainty as follows. First, let $\theta \equiv 0$, thus eliminating demand uncertainty. Let C_i be retailer i's marginal cost, $i = 1, \cdots, n$. After making decisions about whether or not to share their cost information with the manufacturer but before making quantity decisions, each retailer i observes his own cost C_i. The retailers' costs are assumed to be positively correlated.

As expected, the manufacturer always benefits if a retailer decides to share his cost information. However, complete information sharing, i.e., all retailers decide to share their cost information with the manufacturer, is now always an equilibrium. And it is the unique equilibrium sometimes. Moreover, complete information sharing increases the supply chain's total profits. [47]

There are two recent extensions of Li (2002). One is Li and Zhang (2001), where the manufacturer can produce before the retailers make their quantity decisions. Production after receiving the retailers' orders is still possible, but incur a higher cost. Moreover, the manufacturer must satisfy all retailers' orders. The manufacturer makes the early-production decision and the wholesale price decision at the same time. By and large, replacing make-to-order with make-to-stock does not alter the qualitative insights. The second extension is Zhang (2001), who focuses on the sharing of demand information but considers both Cournot and Bertrand competition with substitutes and complements. It is thus a direct generalization of Vives (1984) to a supply chain setting.

4 Future Research

The role of information in achieving supply chain coordination will continue to be a fruitful research area. As in the past, research will progress in many directions.

Full Information, Centralized Control

Here we imagine a supply chain controlled by a central planner with all relevant information. The challenge is to determine a strategy that optimizes the supply chain-wide performance. Many people say that this is the traditional way of thinking in operations management/operations research. But that should not be construed to mean that the area is unimportant. In fact, there are many important problems that are begging for solutions.

[47] The fact that the retailers will share their cost information with the manufacturer is striking at first glance. This may have a lot to do with the assumption that the decision whether or not to share information is made before observing the private information. This assumption is particularly strong with the cost information; if the retailers have been in the business for a while, it seems reasonable that they have better information about their own costs than anyone else prior to making information-sharing decisions. Another scenario that may alter the result is when the parties interact repeatedly.

To see how difficult these problems can be and how little we understand them, simply consider a two-level supply chain with one distribution center replenishing multiple local sales offices. (This kind of supply-chain structure is often under the unglamorous name of one-warehouse multi-retailer systems.) The truth of the matter is nobody knows what the optimal policy is. Many have studied "heuristic" policies: the ones that seem to make intuitive sense. One example is the control rules that are based on echelon inventory positions, i.e., the replenishment strategy at the distribution center is based on the total inventory (on-hand and in-transit) in the system. The allocation decision at the distribution center when it runs out of stock is often of the myopic sort or based on some, often arbitrary, priority rule. This heuristic approach sometimes puts us in a very awkward position, where a strategy based on full information is actually inferior to a strategy that is based only on local information (e.g., the so-called installation-stock policies). Information has a negative value!

To be sure, many years of work has suggested to us that the optimal strategy for the above system (or many other multi-echelon systems with common cost or topological structures) is, if it exists, likely to be very complex. So for all practical purposes, we should confine ourselves to heuristic policies that are easy to implement. But the quest for better heuristics will never stop, unless we know that the heuristic we have is already very close to optimality. The most powerful statement one can make about a heuristic's closeness to optimality is the worst-case gap between the two. Since the optimal policy is unknown, the worst-case analysis must rely on bounds on the optimal performance. Discovering heuristics with small worst-case gaps is the holy grail of multi-echelon inventory theory. [48]

Decentralized Information, Shared Incentives

Now replace the central planner with local managers who are each responsible for managing part of the supply chain. These managers only have access to local information. But they share a common goal to optimize the supply chain-wide performance. This is the team-model approach (Marschak and Radner 1972). And it is appropriate when, e.g., the supply chain consists of multiple divisions of the same firm, and the divisions' incentives are aligned. It is an intermediate step to a full-blown decentralized system.

An important feature of team models is that the control rules used by the local managers can only be based on local information. To obtain a solution to a team model, it is convenient to take the view of an analyst, who optimizes the system-wide objective by restricting to strategies that can be implemented by the local managers (i.e., a strategy, when followed by a local manager using his local information, leads to an unambiguous decision). Therefore, it looks like a centralized planning model, but with an added informational constraint. The difference between the team model and its full-information, central-planner counterpart reveals the value of information. Section 2 of this

[48] A lower-bounding methodology has been given by Chen and Zheng (1994) for general stochastic multi-echelon inventory systems. But the use of lower bounds in evaluating the optimality of heuristic policies has been sporadic and is largely numerical. It is hard to resist the temptation to mention the spectacular successes achieved for general multi-echelon inventory systems with deterministic demand, for which a class of heuristic policies—the so-called power-of-two policies—have been guaranteed to be within 2% of optimality. See, e.g., Maxwell and Muckstadt (1985), Roundy (1985, 1986), and Federgruen et al. (1992). So far, unfortunately, the only comparable result for a stochastic system is the one in Chen (1999b) established for a simple two-stage serial system. In Chapter 10 of this volume, Sven Axsäter reviews studies of heuristic policies in one-warehouse multi-retailer systems.

chapter has reviewed many papers in this area. This will continue to be a fruitful research direction.

Decentralized Information, Independent Entities

A full-blown decentralized supply chain consists of independent firms with asymmetric information. Section 3 of this chapter has covered many such models. As mentioned there, a member of the supply chain may take the initiative of "setting the stage" by either screening or signaling, or the supply chain's members come together (cooperatively or noncooperatively) to form some trading rules. This is a relatively new area for many in operations management. Below we describe several promising research directions.

One potentially fruitful research area is the integration of price discovery with a firm's internal optimization. In §2.2, we have already seen a procurement example with one buyer and multiple potential suppliers with private cost information. There, the solution is a marriage between an auction mechanism and a supply contract. Infusing auction theory into operations management research is exciting.

Another interesting research direction is information acquisition. In §3.1, we saw an example where a firm can 'buy' advance demand information from customers. The challenge was to balance the cost of information acquisition with the benefit of the acquired information. It is certainly possible to study information acquisition in other contexts, with other kinds of information and between members of a supply chain.

In §3.3, we have seen papers dealing with information sharing among competing firms. How about competing supply chains? Information sharing between two supply chains can happen in many different ways: 1) same-layer, cross-channel (e.g., retailer to retailer, supplier to supplier), 2) inter-layer, same-channel (retailer to supplier), and 3) inter-layer, cross-channel (retailer in supply chain A to supplier in supply chain B, and vice versa). Opportunities abound. [49]

Bounded Rationality and Robust Supply Chain Design

Real firms (and people) have limitations. First of all, their data may be inaccurate. For example, a retailer may not know exactly how many units of a product are in the store. This occurs even at successful retailers who have invested large sums in information technology, mostly to track sales and automate transactions. [50] In supply chains, inaccurate data may also result from imperfect transmission of information, which can be noisy and laden with delays. On the other hand, managers like to have easy, intuitively appealing control rules. This is simply because people have limited information-processing power. The same holds for modelers/analysts/researchers: real-world problems are complex with multiple facets, and it is impossible to include all the complexities in a model. (Simply put, people—managers or not—are boundedly rational.) As a result, any "optimal solution" obtained from a model is unlikely to be implemented as is; at best, it will inform a manager's "insight" or "intuition," which in turn influences the final decision. Given inaccurate data, modeling limitations, and managers' desire for simplicity, there is a pressing need to develop simple control mechanisms that are robust to such imperfections. This is virtually an uncharted territory. But it is worthwhile to ask the question.

[49] For an example on competing supply chains, see Corbett and Karmarkar (2001) and the references therein.

[50] See Raman et al. (2000) and Raman and Ton (2000) for discussions on the magnitudes and drivers of inaccurate inventory data in retail stores.

References

Anand, K. and H. Mendelson. 1997. Information and organization for horizontal multimarket coordination. *Management Science* 43 (12), 1609-1627.

Anupindi, R. and Y. Bassok. 1999. Centralization of stocks: retailers vs. manufacturer. *Management Science* 45 (2), 178-191.

Arrow, K. 1985. The economics of agency. In *Principals and Agents*, Ed. by J. Pratt and R. Zeckhauser, Harvard Business School Press, Cambridge, MA.

Aviv, Y. 2001. The effect of collaborative forecasting on supply chain performance. *Management Science* 47 (10), 1326-1343.

Aviv, Y. and A. Federgruen. 1997. Stochastic inventory models with limited production capacity and periodically varying parameters. *Probability in the Engineering and Informational Science* 11, 107-135.

Aviv, Y. and A. Federgruen. 1998. The operational benefits of information sharing and vendor managed inventory (VMI) programs. Working paper, Washington University and Columbia University.

Axsäter, S. 1990. Simple solution procedures for a class of two-echelon inventory problems. *Operations Research* 38 (1), 64-69.

Axsäter, S. 1993a. Continuous review policies for multi-level inventory systems with stochastic demand. In *Handbook in Operations Research and Management Science, Vol. 4, Logistics of Production and Inventory,* Ed. by S. Graves, A. Rinnooy Kan and P. Zipkin, North Holland.

Axsäter, S. 1993b. Exact and approximate evaluation of batch ordering policies for two-level inventory systems. *Operations Research* 41 (4), 777-785.

Axsäter, S. 1997. Simple evaluation of echelon stock (R,Q) policies for two-level inventory systems. *IIE Transactions* 29, 661-669.

Axsäter, S. 1998. Evaluation of installation stock based (R, Q) policies for two-level inventory systems with Poisson demand. *Operations Research* 46 (Supp. No.3), S135-S145.

Axsäter, S. 2000. Exact analysis of continuous review (R,Q) policies in two-echelon inventory systems with Compound Poisson demand. *Operations Research* 48 (5), 686-696.

Axsäter, S. and K. Rosling. 1993. Installation vs. echelon stock policies for multi-level inventory control. *Management Science* 39 (10), 1274-1280.

Azoury, K. 1985. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science* 31, 1150-1160.

Azoury, K. and B. Miller. 1984. A comparison of the optimal ordering levels of Bayesian and non-Bayesian inventory models. *Management Science* 30, 993-1003.

Barnes-Schuster, D., Y. Bassok and R. Anupindi. 1998. Coordination and flexibility in supply contracts with options. Working paper, University of Chicago.

Basu, A., R. Lal, V. Srinivasan, and R. Staelin. 1985. Salesforce-compensation plans: An agency theoretic perspective. *Marketing Science* 4(4), 267-291.

Bitran, G., E. Haas and H. Matsuo. 1986. Production planning of style goods with high setup costs and forecast revisions. *Operations Research* 34, 226-236.

Blanchard, O. 1983. The production and inventory behavior of the American automobile industry. *Journal of Political Economy* 91 , 365-400.

Blinder, A. 1982. Inventories and sticky prices. *American Economic Review* 72, 334-349.

Blinder, A. 1986. Can the production smoothing model of inventory behavior be saved? *Quarterly Journal of Economics* 101, 431-454.

Bourland, K., S. Powell and D. Pyke. 1996. Exploiting timely demand information to reduce inventories. *European Journal of Operational Research* 92, 239-253.

Box, G., G. Jenkins and G. Reinsel. 1994. *Time Series Analysis: Forecasting and Control*, 3rd Edition, Holden-Day, San Francisco, CA., 110-114.

Bradford, J. and P. Sugrue. 1990. A Bayesian approach to the two-period style-goods inventory problem with single replenishment and heterogeneous Poisson demands. *Journal of Operational Research Society* 41, 211-218.

Brown, A. and H. Lee. 1998. The win-win nature of options based capacity reservation arrangements. Working paper, Vanderbilt University.

Cachon, G. 1999. Managing supply chain demand variability with scheduled ordering policies. *Management Science* 45 (6), 843-856.

Cachon, G. 2001. Exact evaluation of batch-ordering inventory policies in two-echelon supply chains with periodic review. *Operations Research* 49 (1), 79-98.

Cachon, G. and M. Fisher. 2000. Supply chain inventory management and the value of shared information. *Management Science* 46 (8), 1032-1048.

Cachon, G. and M. Lariviere. 1999. Capacity choice and allocation: strategic behavior and supply chain performance. *Management Science* 45 (8), 1091-1108.

Cachon, G. P. and M. Lariviere. 2001. Contracting to assure supply: how to share demand forecasts in a supply chain. *Management Science* 47 (5), 629-646.

Caplin, A. 1985. The variability of aggregate demand with (s,S) inventory policies. *Econometrica* 53, 1395–1409.

Cattani, K. and W. Hausman. 2000. Why are forecast updates often disappointing? *Manufacturing & Service Operations Management* 2 (2), 119-127.

Chen, F. 1998a. Echelon reorder points, installation reorder points, and the value of centralized demand information. *Management Science* 44 (12, part 2), S221-S234.

Chen, F. 1998b. On (R,nQ) policies in serial inventory systems. In *Quantitative Models for Supply Chain Management*, Ed. by S. Tayur et al., Kluwer Academic Publishers.

Chen, F. 1999a. Decentralized supply chains subject to information delays. *Management Science* 45 (8), 1076-1090.

Chen, F. 1999b. 94%-effective policies for a two-stage serial inventory system with stochastic demand. *Management Science* 45 (12), 1679-1696.

Chen, F. 2000a. Optimal policies for multi-echelon inventory problems with batch ordering. *Operations Research* 48, 376-389.

Chen, F. 2000b. Salesforce incentives and inventory management. *Manufacturing & Service Operations Management* 2 (2), 186-202.

Chen, F. 2000c. Salesforce incentives, market information, and production/inventory planning. To appear in *Management Science.*

Chen, F. 2001a. Market segmentation, advanced demand information, and supply chain performance. *Manufacturing & Service Operations Management* 3 (1), 53-67.

Chen, F. 2001b. Auctioning supply contracts. Working paper, Columbia University.

Chen, F., J. Eliashberg, and P. Zipkin. 1998. Customer preferences, supply-chain costs, and product-line design. In *Product Variety Management: Research Advances*, Ed. by T.-H. Ho and C. S. Tang, Kluwer Academic Publishers.

Chen, F. and R. Samroengraja. 1999. Order volatility and supply chain costs. Working paper, Columbia University.

Chen, F. and R. Samroengraja. 2000. A staggered ordering policy for one-warehouse multi-retailer systems. *Operations Research* 48 (2), 281-293.

Chen, F. and R. Samroengraja. 2000. The stationary beer game. *Production and Operations Management* 9 (1), 19-30.

Chen, F. and J.-S. Song. 2001. Optimal policies for multi-echelon inventory problems with Markov-modulated demand. *Operations Research* 49 (2), 226-234.

Chen, F. and B. Yu. 2001a. Quantifying the value of leadtime information in a single-location inventory system. Working paper, Columbia Business School.

Chen, F. and B. Yu. 2001b. A supply chain model with asymmetric capacity information. Near completion.

Chen, F. and Y.-S. Zheng. 1994. Lower bounds for multi-echelon stochastic inventory systems. *Management Science* 40, 1426-1443.

Chen, F. and Y.-S. Zheng. 1997. One-warehouse multi-retailer systems with centralized stock information. *Operations Research* 45 (2), 275-287.

Chen, Frank, Z. Drezner, J. Ryan, and D. Simchi-Levi. 2000. Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science* 46 (3), 436-443.

Chu, W. 1992. Demand signaling and screening in channels of distribution. *Marketing Science* 11 (4), 327-347.

Clark, A. and H. Scarf. 1960. Optimal policies for a multi-echelon inventory problem. *Management Science* 6, 475-490.

Clarke, R. N. 1983. Collusion and the incentives for information sharing. *Bell Journal of Economics*, Vol. 14, 383-394.

Corbett, C. 2001. Stochastic inventory systems in a supply chain with asymmetric information: Cycle stocks, safety stocks, and consignment stock. *Operations Research* 49 (4), 487-500.

Corbett, C. and X. de Groote. 2000. A supplier's optimal quantity discount policy under asymmetric information. *Management Science* 46 (3), 444-450.

Corbett, C. and U. Karmarkar. 2001. Competition and structure in serial supply chains with deterministic demand. *Management Science* 47 (7), 966-978.

Corbett, C., D. Zhou, and C. Tang. 2001. Designing supply contracts: Contract type and information asymmetry. Working paper, UCLA.

Coughlan, A. 1993. Salesforce compensation: A review of MS/OR advances. In *Handbooks in Operations Research and Management Science: Marketing*, Vol. 5., J. Eliashberg and G. Lilien (Eds.), North-Holland.

de Groote, X. 1994. Flexibility and product variety in lot-sizing models. *European Journal of Operations Research* 75, 264-274.

Desai, P. S. and K. Srinivasan. 1995. Demand signaling under unobservable effort in franchising: Linear and nonlinear price contracts. *Management Science* 41 (10), 1608-1623.

Deshpande, V. and L. Schwarz. 2002. Optimal capacity allocation in decentralized supply chains. Working paper, Purdue University.

Deuermeyer, B. and L. Schwarz. 1981. A model for the analysis of system service level in warehouse/retailer distribution systems: The identical retailer case. In *Studies in the Management Sciences: The Multi-Level Production/Inventory Control Systems*, Vol. 16, Ed. by L. Schwarz, North-Holland, Amsterdam, 163-193.

Diks, E. B. and A. G. de Kok. 1998. Optimal control of a divergent multi-echelon inventory system. *European Journal of Operational Research* 111, 75-97.

Ding, X., M. L. Puterman, and A. Bisi. 2002. The censored newsvendor and the optimal acquisition of information. *Operations Research* 50 (3), 517-527.

Dobson, G. and S. Kalish. 1988. Positioning and pricing a product line. *Marketing Science* 7, 2, 107-125.

Dobson, G. and S. Kalish 1993. Heuristics for positioning and pricing a product line using conjoint and cost data. *Management Science* 39, 160-175.

Dong, L. and H. Lee. 2000. Optimal policies and approximations for a serial multi-echelon inventory system with time-correlated demand. Working paper, Washington University.

Donohue, K. 2000. Efficient supply contracts for fashion goods with forecast updating and two production modes. *Management Science* 46 (11), 1397-1411.

Ehrhardt, R. 1984. (s, S) policies for a dynamic inventory model with stochastic leadtimes. *Operations Research* 32, 121-132.

Eppen, G. and A. Iyer. 1997a. Backup agreements in fashion buying - the value of upstream flexibility. *Management Science* 43 (11), 1469-1484.

Eppen, G. and A. Iyer. 1997b. Improved fashion buying with Bayesian updates. *Operations Research* 45 (6), 805-819.

Eppen, G. and L. Schrage. 1981. Centralized ordering policies in a multiwarehouse system with leadtimes and random demand. In *Multi-Level Production/Inventory Control Systems: Theory and Practice*, Ed. by L. Schwarz, North Holland, 51-69.

Federgruen, A. 1993. Centralized planning models for multi-echelon inventory systems under uncertainty. In *Handbook in Operations Research and Management Science, Vol. 4, Logistics of Production and Inventory,* Ed. by S. Graves, A. Rinnooy Kan and P. Zipkin, North-Holland.

Federgruen, A., M. Queyranne and Y.-S. Zheng. 1992. Simple power of two policies are close to optimal in general class of production/distribution networks with general joint setup costs. *Mathematics of Operations Research* 17, 951-963.

Federgruen, A. and P. Zipkin. 1984a. Approximation of dynamic, multi-location production and inventory problems. *Management Science* 30, 69-84.

Federgruen, A. and P. Zipkin. 1984b. Allocation policies and cost approximation for multi-location inventory systems. *Naval Research Logistics Quarterly* 31, 97-131.

Federgruen, A. and P. Zipkin. 1986a. An inventory model with limited production capacity and uncertain demands I. The average cost criterion. *Mathematics of Operations Research* 11, 193-207.

Federgruen, A. and P. Zipkin. 1986b. An inventory model with limited production capacity and uncertain demands II. The discounted-cost criterion. *Mathematics of Operations Research* 11, 208-215.

Fisher, M. and A. Raman. 1996. Reducing the cost of demand uncertainty through accurate response to early sales. *Operations Research* 44 (1), 87-99.

Forrester, J. 1961. *Industrial Dynamics*, John Wiley and Sons Inc., New York.

Forsberg, R. 1995. Optimization of order-up-to S policies for two-level inventory systems with compound Poisson demand. *European Journal of Operational Research* 81, 143-153.

Fudenberg, D. and J. Tirole. 1992. *Game Theory*, The MIT Press, Cambridge, MA.

Gallego, G., Y. Huang, K. Katircioglu, and Y.T. Leung. 2000. When to share demand information in a simple supply chain? Working paper, Columbia University.

Gallego, G. and O. Özer. 2000. Optimal replenishment policies for multi-echelon inventory problems under advance demand information. Working paper, Columbia University and Stanford University.

Gallego, G. and O. Özer. 2001. Integrating replenishment decisions with advance demand information. *Management Science* 47 (10), 1344-1360.

Gallego, G. and B. Toktay. 1999. All-or-nothing ordering under a capacity constraint and forecasts of stationary demand. Working paper, Columbia University.

Gal-Or, E. 1985. Information sharing in oligopoly. *Econometrica*, Vol. 53 (2), 329-343.

Gal-Or, E. 1986. Information transmission – Cournot and Bertrand equilibria. *Review of Economic Studies*, Vol. LIII, 85-92.

Gavirneni, S., R. Kapuscinski and S. Tayur. 1999. Value of information in capacitated supply chains. *Management Science* 45 (1), 16-24.

Gonik, J. 1978. Tie salesmen's bonuses to their forecasts. *Harvard Business Review*, May-June 1978.

Graves, S. 1985. A multi-echelon inventory model for a reparable item with one-for-one replenishment. *Management Science* 31, 1247-1256.

Graves, S. 1996. A multiechelon inventory model with fixed replenishment intervals. *Management Science* 42 (1), 1-18.

Graves, S. 1999. A single-item inventory model for a nonstationary demand process. *Manufacturing & Service Operations Management* 1 (1), 50-61.

Graves, S., D. Kletter, W. Hetzel. 1998. A dynamic model for requirements planning with application to supply chain optimization. *Operations Research* 46, S35-S49.

Graves, S., H. Meal. S. Dasu, Y. Qiu. 1986. Two-stage production planning in a dynamic environment. In *Multi-Stage Production Planning and Control*, by S. Axsäter, C. Schneeweiss, and E. Silver (eds.), Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin. Vol. 266, 9-43.

Green, P. and A. Krieger 1985. Models and heuristics for product-line selection. *Marketing Science* 4 (Winter), 1-19.

Grossman, S. and O. Hart. 1983. An analysis of the principal-agent problem. *Econometrica* 51 (1), 7-45.

Gullu, R. 1996. On the value of information in dynamic production/inventory problems under forecast evolution. *Naval Research Logistics* 43, 289-303.

Gullu, R. 1997. A two-echelon allocation model and the value of information under correlated forecasts and demands. *European Journal of Operational Research* 99, 386-400.

Ha, A. 2001. Supplier-buyer contracting: Asymmetric cost information and cutoff level policy for buyer participation. *Naval Research Logistics* 48, 41-64.

Hariharan, R. and P. Zipkin. 1995. Customer order information, lead times, and inventories. *Management Science* 41 (10), 1599-1607.

Harris, M., and A. Raviv. 1978. Some results on incentive contracts with applications to education and employment, health insurance, and law enforcement. *American Economic Review* 68 (March), 20-30.

Harris, M. and A. Raviv. 1979. Optimal incentive contracts with imperfect information. *Journal of Economic Theory* 20 (April), 231-259.

Hassin, R. and M. Haviv. 2001. *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*, to be published in the International Series in Operations Research and Management Science, Kluwer Academic Publishers, Boston/Dordrecht/London.

Hausman, W. 1969. Sequential decision problems: A model to exploit existing forecasters. *Management Science* 16 (2), B-93-B-111.

Hausman, W. and R. Peterson. 1972. Multiproduct production scheduling for style goods with limited capacity, forecast revisions and terminal delivery. *Management Science* 18, 370-383.

Hayek, F. 1945. The use of knowledge in society. *American Economic Review* 35 (4), 519-530.

Heath, D. and P. Jackson. 1994. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Transactions* 26 (3), 17-30.

Ho, T. and C. Tang (eds.) 1998. *Product Variety Management: Research Advances*, Kluwer Academic Publishers.

Holmstrom, B. 1979. Moral hazard and observability. *Bell Journal of Economics* 10(1), 74-91.

Holmstrom, B. 1982. Moral hazard in teams. *Bell Journal of Economics* 13(2), 324-340.

Horngren, C. T. and G. Foster. 1991. *Cost Accounting: A Managerial Emphasis*, 7th Edition, Prentice Hall, Englewood Cliffs, New Jersey.

Iglehart, D. 1964. The dynamic inventory model with unknown demand distribution. *Management Science* 10, 429-440.

Jackson, P. 1988. Stock allocation in a two-echelon distribution system or 'what to do until your ship comes in.' *Management Science* 34, 880-895.

Jensen, M. and W. Meckling. 1976. Theory of the firm: managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3, 305-360.

Jensen, M. and W. Meckling. 1992. Specific and general knowledge, and organizational structure. In *Contract Economics*, L. Werin and H. Hijkander (eds.), Basil Blackwell, Cambridge, MA.

Johnson, G. and H. Thompson. 1975. Optimality of myopic inventory policies for certain dependent demand processes. *Management Science* 21, 1303-1307.

Kahn, J. 1987. Inventories and the volatility of production. *American Economic Review* 77 (4), 667–679.

Kaplan, R. 1970. A dynamic inventory model with stochastic lead times. *Management Science* 16, 491-507.

Kapuscinski, R. and S. Tayur. 1998. A capacitated production-inventory model with periodic demand. *Operations Research* 46 (6), 899-911.

Karaesmen, F., J. Buzacott, and Y. Dallery. 2001. Integrating advance order information in make-to-stock production systems. Working paper, York University, Canada.

Klemperer, P. 1999. Auction theory: a guide to the literature. *Journal of Economic Surveys* 13 (3), 227-286.

Kopczak, L. and H. Lee. 1994. Hewlett-Packard: Deskjet Printer Supply Chain (A). Stanford University Case.

Kreps, D. 1990. *A Course in Microeconomic Theory*, Princeton University Press, Princeton, New Jersey.

Kreps, D. and R. Wilson. 1982. Sequential equilibrium. *Econometrica* 50, 863-894.

Lal, R. 1986. Delegating pricing responsibility to the salesforce. *Marketing Science* 5(2), 159-168.

Lal, R., and R. Staelin. 1986. Salesforce-compensation plans in environments with asymmetric information. *Marketing Science* 5(3), 179-198.

Lancaster, K. 1979. *Variety, Equity and Efficiency*, Columbia University Press, New York, NY.

Lancaster, K. 1990. The economics of product variety: A survey. *Marketing Science* 9 (3), 189-206.

Lariviere, M. A. and V. Padmanabhan. 1997. Slotting allowances and new product introductions. *Marketing Science* 16 (2), 112-128.

Lariviere, M. A. and E. L. Porteus. 1999. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science* 45 (3), 346-363.

Lee, H. and K. Moinzadeh. 1987a. Two-parameter approximations for multi-echelon reparable inventory models with batch ordering policy. *IIE Transactions* 19, 140-149.

Lee, H. and K. Moinzadeh. 1987b. Operating characteristics of a two-echelon inventory system for reparable and consumable items under batch ordering and shipment policy. *Naval Research Logistics Quarterly* 34, 365-380.

Lee, H., P. Padmanabhan, and S. Whang. 1997a. The bullwhip effect in supply chains. *Sloan Management Review* 38, 93-102.

Lee, H., P. Padmanabhan, and S. Whang. 1997b. Information distortion in a supply chain: The bullwhip effect. *Management Science* 43, 546-558.

Lee, H., K. So and C. Tang. 2000. The value of information sharing in a two-level supply chain. *Management Science* 46 (5), 626-643.

Lee, H. and C. Tang. 1998. Variability reduction through operations reversals. *Management Science* 44 (2), 162-172.

Lee, H. and S. Whang. 1999. Decentralized multi-echelon supply chains: Incentives and Information. *Management Science* 45 (5), 633-640.

Lee, H. and S. Whang. 2000. Information sharing in a supply chain. *Int. J. Technology Management* 20 (3/4), 373-387.

Lee, H. and S. Whang. 2002. The impact of the secondary market on the supply chain. *Management Science* 48 (6), 719-731.

Li, L. 1985. Cournot oligopoly with information sharing. *Rand Journal of Economics* 16 (4), 521-536.

Li, L. 2002. Information sharing in a supply chain with horizontal competition. *Management Science* 48 (9), 1196-1212.

Li, L. and H. Zhang. 2001. Competition, inventory, demand variability, and information sharing in a supply chain. Working paper, Yale School of Management.

Lovejoy, W. 1990. Myopic policies for some inventory models with uncertain demand. *Management Science* 36, 724-738.

Lovejoy, W. 1992. Stopped myopic policies for some inventory models with uncertain demand distributions. *Management Science* 38, 688-707.

Marschak, J. and R. Radner. 1972. *Economic Theory of Teams*, Yale University Press, New Haven.

Maskin, E. and J. Riley. 1984. Monopoly with incomplete information. *Rand Journal of Economics* 15 (2), 171-196.

Matsuo, H. 1990. A stochastic sequencing problem for style goods with forecast revisions and hierarchical structure. *Management Science* 36, 332-347.

Maxwell, W. and J. Muckstadt. 1985. Establishing consistent and realistic reorder intervals in production-distribution systems. *Operations Research* 33, 1316-1341.

McAfee, R. and J. McMillan. 1987. Auctions and bidding. *Journal of Economic Literature* XXV (June), 699-738.

Mendelson, H. and T. Tunca. 2001. Business to business exchanges and supply chain contracting. Working paper, Stanford Business School.

Miller, B. 1986. Scarf's state reduction method, flexibility, and a dependent demand inventory model. *Operations Research* 34, 83-90.

Milner, J. and P. Kouvelis. 2001. More demand information or more supply chain flexibility: what does the answer depend on? Working paper, Washington University, St. Louis.

Moinzadeh, K. and H. Lee. 1986. Batch size and stocking levels in multi-echelon reparable systems. *Management Science* 32, 1567-1581.

Moorthy, K. 1984. Market segmentation, self-selection, and product line design. *Marketing Science* 3, 288-307.

Muharremoglu, A. and J. Tsitsiklis. 2001. Echelon base stock policies in uncapacitated serial inventory systems. Working paper, MIT.

Murray, G., Jr. and E. Silver. 1966. A Bayesian analysis of the style goods inventory problem. *Management Science* 12, 785-797.

Mussa, M. and S. Rosen. 1978. Monopoly and product quality. *Journal of Economic Theory* 18, 301-317.

Myerson, R. 1981. Optimal auction design. *Mathematics of Operations Research* 6 (1), 58-73.

Nahmias, S. 1979. Simple approximations for a variety of dynamic leadtime lost-sales inventory models. *Operations Research* 27, 904-924.

Nanda, D. 1995. Strategic impact of just-in-time manufacturing on product market competitiveness. Working paper, University of Rochester, Rochester, NY.

Novshek, W. and H. Sonnenschein. 1982. Fulfilled expectations Cournot duopoly with information acquisition and release. *Bell Journal of Economics*, Vol.13, 214-218.

Özer, O. 2000. Replenishment strategies for distribution systems under advance demand information. Working paper, Stanford University.

Özer, O. and W. Wei. 2001. Inventory control with limited capacity and advance demand information. Working paper, Stanford University.

Plambeck, E. and S. Zenios. 2000a. Performance-based incentives in a dynamic principal-agent model. *Manufacturing & Service Operations Management* 2 (3), 240-263.

Plambeck, E. and S. Zenios. 2000b. Incentive efficient control of a make-to-stock production system. Working paper, Stanford Business School.

Porteus, E. 2000. Responsibility tokens in supply chain management. *Manufacturing & Service Operations Management* 2 (2), 203-219.

Porteus, E. and S. Whang. 1991. On manufacturing/marketing incentives. *Management Science* 37 (9), 1166-1181.

Raghunathan, S. 2001. Information sharing in a supply chain: A note on its value when demand is nonstationary. *Management Science* 47 (4), 605-610.

Raju, J. and V. Srinivasan. 1996. Quota-based compensation plans for multiterritory heterogeneous salesforces. *Management Science* 42 (10), 1454-1462.

Raman, A., N. DeHoratius and Z. Ton. 2000. Execution: The missing link in retail operations. Working Paper, Harvard Business School.

Raman, A. and Z. Ton. 2000. An empirical analysis of the magnitude and drivers of misplaced SKUs in retail stores. Working Paper, Harvard Business School.

Rao, R. 1990. Compensating heterogeneous salesforces: some explicit solutions. *Marketing Science* 9(4), 319-341.

Riley, J. and W. Samuelson. 1981. Optimal auctions. *American Economic Review* 71 (3), 381-392.

Roundy, R. 1985. 98%-effective integer-ratio lot-sizing for one-warehouse multi-retailer systems. *Management Science* 31, 1416-1430.

Roundy, R. 1986. 98%-effective lot-sizing rule for a multi-product, multi-facility production-inventory systems. *Mathematics of Operations Research* 11, 699-727.

Ryan, J. 1997. Analysis of inventory models with limited demand information. Unpublished Ph. D. dissertation, Northwestern University.

Scarf, H. 1959. Bayes solutions of the statistical inventory problem. *Ann. Math. Statist.* 30, 490-508.

Scarf, H. 1960. Some remarks on Bayes solutions to the inventory problem. *Naval Research Logistics Quarterly* 7, 591-596.

Sethi, S. and F. Cheng. 1997. Optimality of (s,S) policies in inventory models with Markovian demand. *Operations Research* 45, 931-939.

Shapiro, C. 1986. Exchange of cost information in oligopoly. *Review of Economic Studies*, Vol. LIII, 433-446.

Shavell, S. 1979. Risk sharing and incentives in the principal and agent relationship. *Bell Journal of Economics* 10(1), 55-73.

Sherbrooke, C. 1968. METRIC: A multi-echelon technique for recoverable item control. *Operations Research* 16, 122-141.

Shocker, A. and V. Srinivasan. 1979. Multiattribute approaches for product concept evaluation and generations: A critical review. *Journal of Marketing Research* Vol. XVI, 159-180.

Signorelli, S. and J. Heskett. 1984. Benetton (A) and (B). Harvard Business School Case (9-685-014).

Simon, R. 1971. Stationary properties of a two echelon inventory model for low demand items. *Operations Research* 19, 761-777.

Sobel, M. 1997. Lot sizes in production lines with random yield and ARMA demand. Working paper, New York State University at Stony Brook.

Song, J.-S. and P. Zipkin. 1992. Evaluation of base-stock policies in multi-echelon inventory systems with state-dependent demands. Part I. State-independent policies. *Naval Research Logistics* 39, 715-728.

Song, J.-S. and P. Zipkin. 1993. Inventory control in a fluctuating demand environment. *Operations Research* 41, 351-370.

Song, J.-S. and P. Zipkin. 1996a. Evaluation of base-stock policies in multi-echelon inventory systems with state-dependent demands. Part II. State-dependent policies. *Naval Research Logistics* 43, 381-396.

Song, J.-S. and P. Zipkin. 1996b. Inventory control with information about supply conditions. *Management Science* 42, 1409-1419.

Spengler, J. 1950. Vertical integration and antitrust policy. *Journal of Political Economy* 58, 347-352.

Sterman, J. 1989. Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science* 35, 321-339.

Svoronos, A. and P. Zipkin. 1988. Estimating the performance of multi-level inventory systems. *Operations Research* 36, 57-72.

Svoronos, A. and P. Zipkin. 1991. Evaluation of one-for-one replenishment policies for multi-echelon inventory systems. *Management Science* 37, 68-83.

Tirole, J. 1988. *The theory of industrial organization*. The MIT Press, Cambridge, MA.

Toktay, B. and L. Wein. 2001. Analysis of a forecasting-production-inventory system with stationary demand. *Management Science* 47 (9), 1268-1281.

Van Mieghem, J. 1999. Coordinating investment, production, and subcontracting. *Management Science* 45 (7), 954-971.

Van Mieghem, J. and M. Dada. 1999. Price versus production postponement: capacity and competition. *Management Science* 45 (12), 1631-1649.

Vickrey, W. 1961. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance* 16 (1), 8-37.

Vives, X. 1984. Duopoly information equilibrium: Cournot and Bertrand. *Journal of Economic Theory*, Vol. 34, 71-94.

Watson, N. and Y.-S. Zheng. 2001. Adverse effects of over-reaction to demand changes and improper forecasting. Working Paper, University of Pennsylvania.

Yano, C. and G. Dobson. 1998. Profit optimizing product line design, selection and pricing with manufacturing cost considerations: A survey. In *Product Variety Management: Research Advances*, Ed. by T.-H. Ho and C.S. Tang, Kluwer Academic Publisher, 145-176.

Zhang, H. 2001. Vertical information exchange in a supply chain with duopoly retailers. Working paper, Hong Kong University of Science and Technology.

Zheng, Y.-S.. 1992. On properties of stochastic inventory systems. *Management Science* 38, 87-103.

Zheng, Y.-S. and F. Chen. 1992. Inventory policies with quantized ordering. *Naval Research Logistics* 39, 285-305.

Zipkin, P. 1986. Stochastic leadtimes in continuous-time inventory models. *Naval Research Logistics Quarterly* 33, 763-774.