Sterman, J. 2015. Booms, Busts, and Beer: Understanding the Dynamics of Supply Chains. Handbook of Behavioral Operations Management: Social and Psychological Dynamics in Production and Service Settings. E. Bendoly, van Wezel, W., Bachrach, D., eds. New York, Oxford University Press: 203-237.

Booms, Busts, and Beer: Understanding the Dynamics of Supply Chains

John Sterman

MIT Sloan School of Management

> The central core of many industrial companies is the process of production and distribution. A recurring problem is to match the production rate to the rate of final consumer sales. It is well known that factory production rate often fluctuates more widely than does the actual consumer purchase rate. It has often been observed that a distribution system of cascaded inventories and ordering procedures seems to amplify small disturbances that occur at the retail level... How does the system create amplification of small retail sales changes?... [W]e shall see that typical manufacturing and distribution practices can generate the types of business disturbances which are often blamed on conditions outside the company.
>
> —Jay W. Forrester (*Industrial Dynamics*, 1961, p. 22)

A. Overview

The purpose of a supply chain is to provide the right output at the right time. As customer requirements change, the managers of the supply chain respond by adjusting the rate at which resources are ordered and used. Supply chains are thus governed primarily by negative feedback. Because supply chains typically involve substantial time delays, they are prone to oscillation—production and inventories chronically overshoot and undershoot the appropriate levels. Figure 1 shows industrial production in the US for consumer goods and materials since 1950. The data are detrended (the long-run growth rate of manufacturing output since 1950 is about 3.2%/year). The data reveal three important features:

1. ***Oscillation***: Production fluctuates significantly around the growth trend. The dominant periodicity is the business cycle, a cycle of prosperity and recession averaging about 4.5 years in duration, but exhibiting considerable variability.[1]
2. ***Amplification***: The amplitude of the fluctuations in materials production (upstream in

[1] The NBER, official arbiter of business cycle timing in the US, reports an average peak-to-peak cycle duration of 56.4 months over 33 cycles from 1854 through 2009, with a standard deviation of 28.5 months and a range from 17 to 128 months (http://www.nber.org/cycles.html).

the supply chain compared to consumer goods) is significantly greater than that in consumer goods production. For example, the standard deviation of the fractional rates of change in monthly output is 11.7%/year for consumer goods, but 17.6%/year for materials, some 150% greater.[2]

3. *Phase lag*: The peaks and troughs of the cycle in materials production tend to lag behind those in production of consumer goods.

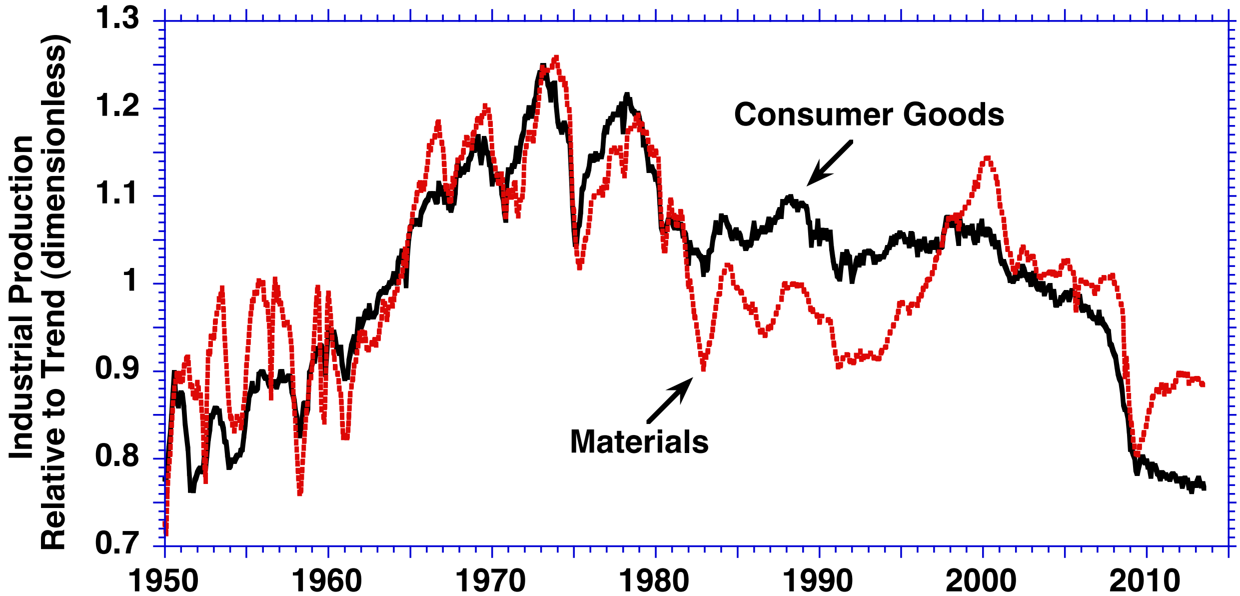

Figure 1 Oscillation, amplification, and phase lag in the aggregate supply chain

Source: US Federal Reserve, series B51000 and B53000, Jan 1950-August 2013. The trend is the best exponential growth fit to each series.

These three features, *oscillation*, *amplification,* and *phase lag,* are pervasive in supply chains. Typically, the amplitude of fluctuations increases as they propagate from the customer to the supplier, with each upstream stage in a supply chain tending to lag behind its immediate customer.

The amplification of fluctuations from consumption to production is even greater in specific industries. The top panel in Figure 2 shows the petroleum supply chain (the figure shows the annualized growth rate; the graph shows 12-month centered moving averages to filter out the high-frequency month-to-month noise).

[2] Standard deviation in annualized rate of change in the seasonally adjusted monthly data.

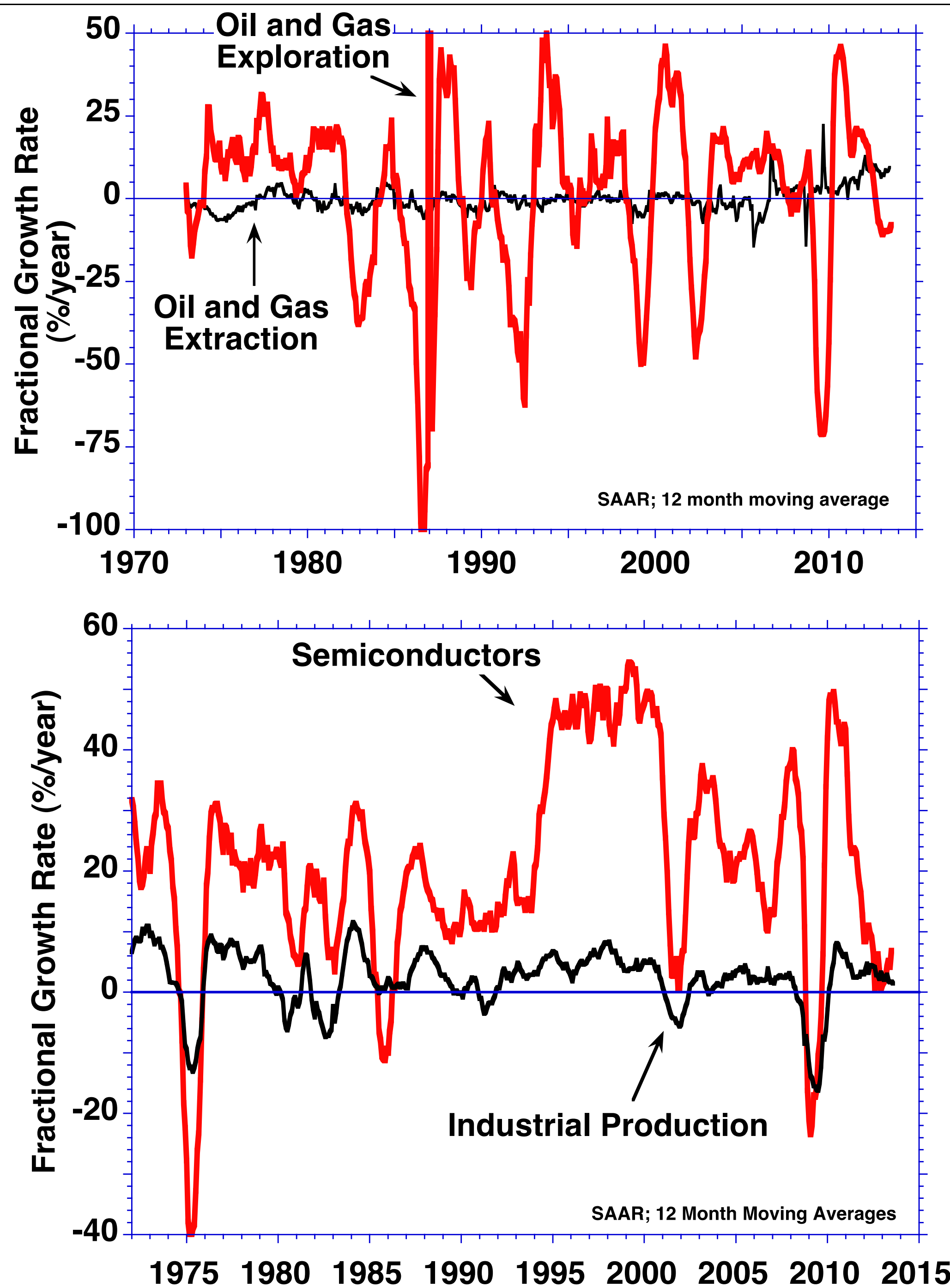

Figure 2 Amplification in supply chains

Top: Oil and gas production and drilling activity, US.
Bottom: Semiconductor production compared to industrial production, US
Graphs show 12-month centered moving averages of the annualized fractional growth rates calculated from the seasonally adjusted monthly data.

The amplification is substantial: drilling activity fluctuates about three times more than production, imposing large boom and bust cycles on the suppliers of drill rigs and equipment. The bottom panel shows the semiconductor industry. Semiconductor production is at the upstream end of the supply chain for electronic equipment and fluctuates far more than industrial production as a whole. Other industries show similar amplification within their supply chains, including machine tools (Anderson, Fine, & Parker 2000, Sterman 2000).

A central question in operations management is whether the oscillations, amplification and phase lag observed in supply chains arise as the result of *operational* or *behavioral* causes.

Operational theories assume that decision makers are rational agents who make optimal decisions given their local incentives and information. Supply chain instability must then result from the interaction of rational actors with the physical and institutional structure of the system. Physical structure includes the network linking customers and suppliers and the placement of inventories and buffers within it, along with capacity constraints and time delays in production, order fulfillment, transportation, and so on. Institutional structure includes the degree of horizontal and vertical coordination and competition among firms, the availability of information to decision makers in each organization, and the incentives faced by each decision maker.

Behavioral explanations also capture the physical and institutional structure of supply chains, but view decision makers as boundedly rational actors with imperfect mental models, actors who use heuristics to make ordering, production, capacity acquisition, pricing and other decisions (Morecroft 1985, Sterman 2000, Boudreau *et al.* 2003, Gino & Pisano 2008, Bendoly *et al.* 2010, Croson *et al.* 2013). These heuristics may yield excellent or suboptimal results depending on how well they capture the complexity of the situation (Simon 1969, 1982). Behavioral theories also recognize the errors and biases that often arise in judgment and decision making (e.g., Kahneman, Slovic, & Tversky 1982). Behavioral explanations also recognize that situational factors such as time pressure and poverty consume scarce cognitive resources that can lead to poor decisions (Shah, Mullainathan, & Shafir 2012) and that decisions made in conditions of stress can be strongly conditioned by fear, anger, and other psychophysiological reactions (Lo & Repin 2002, Rudolph & Repenning 2002).

To illustrate the difference between operational and behavioral theories, consider a simple supply chain with a single producer servicing two competing retailers. If there is an unexpected increase in final demand, both retailers will place additional orders with the supplier. If those orders exceed the supplier's capacity, then the product will be placed on allocation—each retailer receives only a fraction of what they desire. In that case, rational retailers might respond strategically by ordering more than they actually want in hopes of gaining a larger share of the total available shipments from the supplier, leading to what Sterman (2000) calls "phantom orders." The result would be amplification of the change in final demand, or even a demand bubble (Lee *et al.* 1997, Cachon & Lariviere 1999, Armony & Plambeck 2005). Alternatively, retailers might use behavioral decision rules such as "order more whenever there is scarcity" or even suffer from emotional overreactions leading to hoarding inventory as deliveries fall. Such heuristics and emotional reactions would also lead to amplification of final demand even though the retailers are not behaving rationally (e.g., Sterman & Dogan 2014).

The difference matters: if supply chain instability arises from operational factors and rational behavior, then policies must be directed at changing the physical and institutional structure of the system, including incentives. If, however, instability arises from bounded rationality and emotional arousal such policies may not be sufficient. To illustrate, shortages of gasoline have sometimes caused retail service stations to run out, leading to "Sorry—No Gas" signs; episodes include the 1979 gas crisis in the US, transport strikes in Europe in 2000, and the aftermath of Superstorm Sandy in 2012 on the east coast of the US. In each case, gas shortages led to long lines as people queued, often for hours, in an attempt to top off their tanks. For example, after Superstorm Sandy,

> "… drivers waited in lines that ran hundreds of vehicles deep, requiring state troopers and local police to protect against exploding tempers.
>
> … The lines themselves only exacerbated the problem; reports in the local media provoked drivers to buy gasoline before stations ran out. Some spent what fuel they had searching for more and could be seen pushing vehicles toward relief.
>
> 'I just want to have it, because you don't know how long this is going to last,' said Richard Bianchi, waiting in the half-mile line at the Sunoco in Union [New Jersey] with a tank that was three-quarters full.
>
> 'People are panicking,' said Jimmy Qawasmi, the owner of a Mobil in the Westchester County town of Mamaroneck."[3]

If such behavior is rational, then policies that alter the institutional structure and incentives such as maximum purchases or odd-even rules (limiting people to purchases every other day based on the last digit of their license plates) should reduce demand and ease the shortage. If hoarding is a behavioral and emotional response to scarcity, then these actions may worsen the situation by reinforcing people's belief that there really is a shortage and increasing the number of people who queue even when their tanks are nearly full. Of course any situation may involve

[3] New York Times, 2 November 2012, http://www.nytimes.com/2012/11/02/nyregion/gasoline-shortages-disrupting-recovery-from-hurricane.html.

a mix of strategic, rational action and behavioral, emotional responses.

In this chapter I show how supply chain instability, including oscillation, amplification and phase lag, arise from the interaction of the basic physics of supply chains with behavioral decision processes. Amplification and phase lag arise from the presence of basic physical structures including stocks of inventory and delays in adjusting production or deliveries to changes in incoming orders. Oscillations, however, are not inevitable. They arise from boundedly rational, behavioral decision processes. Experimental studies show, furthermore, that supply chain instability, including oscillation, amplification and phase lag, along with demand bubble, hoarding, and phantom ordering, arise even in experimental settings in which there are no operational factors that might make such behavior rational. I also present two learning activities that can be used effectively to teach principles of supply chains. The *Manufacturing Case* (Booth Sweeney & Sterman 2000) is a simple paper-and-pencil exercise that tests participants' understanding of time delay in a simple inventory management setting, and explores the origin of amplification and phase lag. The *Beer Distribution Game* (Sterman 1989a) is a role-play simulation of a simple supply chain and is widely used to teach principles of operations management, system dynamics and systems thinking.

B. Theoretical Perspective

Supply chains consist of cascades of firms, each receiving orders and adjusting production and production capacity to meet changes in demand. Each link in a supply chain maintains and controls inventories of materials and finished product. To understand the behavior of a supply chain and the causes of oscillation, amplification, and phase lag, it is first necessary to understand the structure and dynamics of a single link; that is, how an individual firm manages its inventories and resources as it attempts to balance production with orders. Such balancing processes always involve negative feedbacks

The Stock Management Problem: Structure

All negative feedback processes involve comparing the state of the system to the desired state, then initiating a corrective action to eliminate any discrepancy. In such a *stock management* task, the manager seeks to maintain a stock (the state of the system) at a particular target level, or at least within an acceptable range. Stocks are altered only by changes in their inflow and outflow rates. Typically, the manager must set the inflow rate to compensate for losses and usage, and to counteract disturbances that push the stock away from its desired value.

Often there are lags between the initiation of a control action and its effect and lags between a change in the stock and the perception of that change by the decision maker. The duration of these lags may vary and may be influenced by the manager's own actions.

Stock management problems occur at many levels of aggregation. At the level of a firm, managers must order parts and raw materials to maintain inventories sufficient for production to proceed at the desired rate. They must adjust for variations in the usage of these materials and for changes in their delivery delays. At the individual level, you regulate the temperature of the water in your morning shower, guide your car down the highway, and manage your checking account balances. At the macroeconomic level, central banks like the US Federal Reserve seek to manage the stock of money to stimulate economic growth and avoid inflation, while compensating for variations in credit demand, budget deficits, and international capital flows.

The stock management control problem can be divided into two parts: (1) the stock and flow structure of the system and (2) the decision rules managers use to order and produce new units (Figure 3).

The stock to be controlled, S, accumulates the acquisition rate AR less the loss rate LR:

$$S_t = \int_{t_0}^{t} (AR - LR)ds + S_{t_0} \tag{1}$$

Losses include any outflow from the stock. Losses may arise from usage (as in a raw material inventory) or decay (as in the depreciation of plant and equipment). The loss rate must depend on the stock itself—losses must approach zero as the stock is depleted—and may also depend on other endogenous variables, **X,** and exogenous variables, **U**. Losses may be nonlinear and may depend on the age distribution of the stock, captured by a loss rate function:

$$LR = f_{LR}(S, X, U) \tag{2}$$

In general, managers cannot add new units to a stock simply because they desire to do so. Typically there are delays in acquiring new units, creating a *supply line* of orders that have been placed but not yet received: A firm seeking to increase its capital stock cannot acquire new units immediately but must await construction or delivery. New workers cannot be hired and trained instantly. It takes time for your car to stop after you step on the brakes, and it takes time for the economy to respond after the Federal Reserve changes interest rates. The supply line accumulates orders less the rate at which units are completed and enter the stock:

$$SL_t = \int_{t_0}^{t} (OR - AR)ds + SL_{t_0} \tag{3}$$

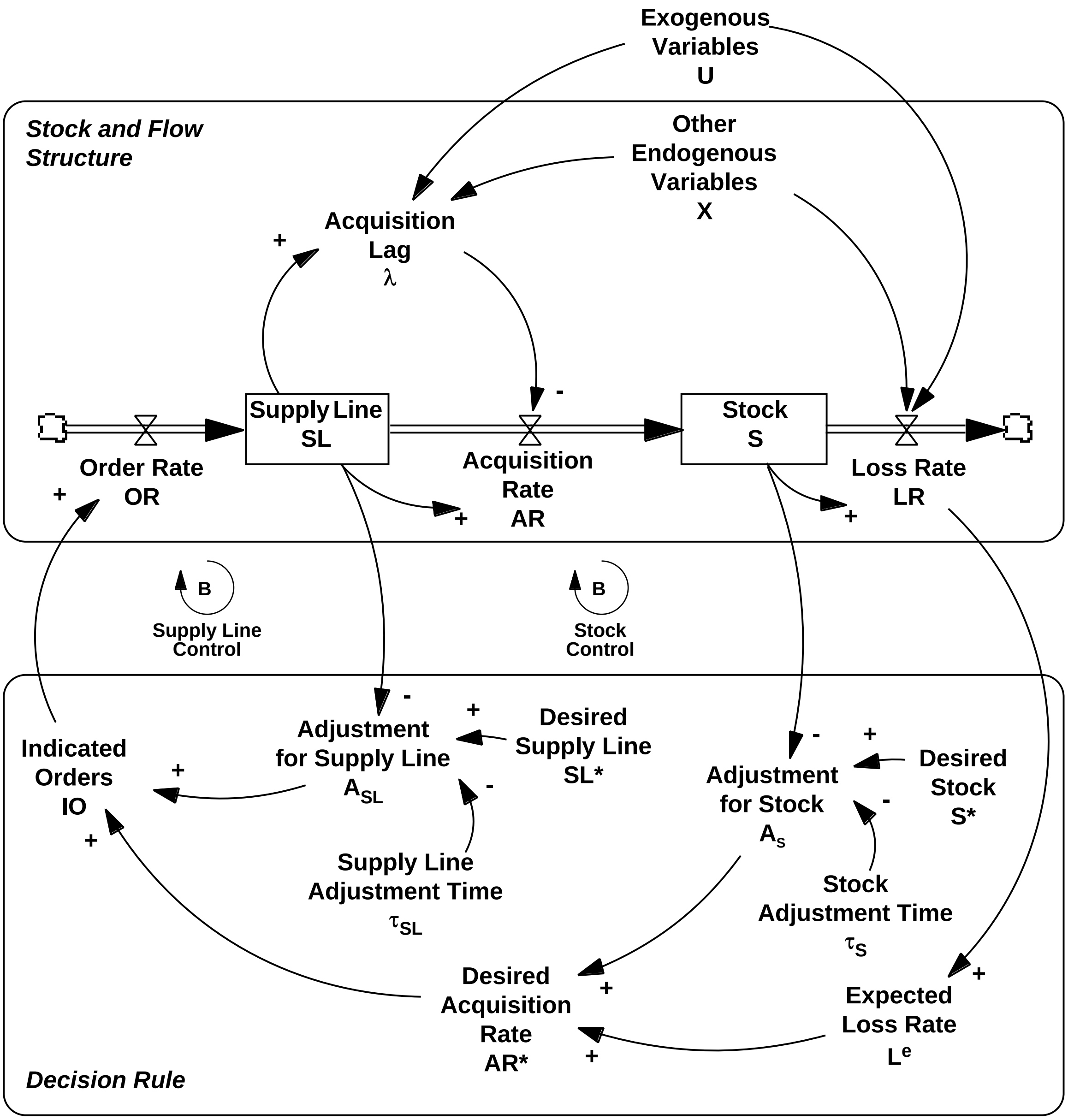

Figure 3 The generic stock management structure

The determinants of the desired supply line are not shown (see text).

The acquisition rate depends on the supply line SL of units that have been ordered but not yet received and the average acquisition lag, λ:

$$AR = \mathcal{L}(SL, \lambda) \tag{4}$$

$$\lambda = f_{\lambda}(SL, \boldsymbol{X}, \boldsymbol{U}) \tag{5}$$

where the lag operator $\mathcal{L}(SL, \lambda)$ denotes a material delay or distributed lag in which acquisitions lag orders with a mean delay of λ time units (Sterman 2000, Ch. 11).

In general, the acquisition lag is a variable that depends on the supply line itself and on the other endogenous and exogenous variables. For example, the acquisition rate is typically capacitated: production depends on the plant, equipment, labor and other resources of the firm; deliveries from a supplier depend on the supplier's inventory and transportation capacity; construction of new buildings depends on the capacity of the construction industry in the region, and so on. Consider the recovery of the housing industry from the great recession of 2007. During the downturn, housing starts and the supply line of homes under construction were very small. Homes could be built quickly because labor and equipment were abundant. The acquisition lag would be at or below normal. As housing starts recover, the acquisition lag would remain near the normal value until construction activity nears the capacity of the construction industry. Once capacity is fully utilized, the acquisition lag increases as the supply line rises relative to the acquisition rate. The acquisition lag can also be influenced by managerial decisions, as when a firm chooses to expedite delivery of materials by paying premium freight or speeding production by use of overtime and extra shifts.

The structure represented by Figure 3 is quite general. The system may be nonlinear. There may be arbitrarily complex feedbacks among the endogenous variables, and the system may be influenced by a number of exogenous forces, both systematic and stochastic. The delay in acquiring new units is often variable and may be constrained by the capacity of the supplier. Table 1 maps common examples into the generic form. In each case, the manager must choose the order rate over time to keep the stock close to a target. Note that most of these systems tend to generate oscillation and instability. The structure can be applied to systems from management to medicine and beyond. As an example, McCarthy *et al.* (2014) develop a system dynamics model to improve treatment of long-term hemodialysis patients suffering from anemia. The model captures the time delays between treatments with erythropoeisis-stimulating agents and changes in hemoglobin levels, showing how a new treatment protocol stabilizes hemoglobin levels, leading to better patient outcomes and lower treatment costs.

System	Stock	Supply Line	Loss Rate	Acquisition Rate	Order Rate	Typical Behavior
Inventory management	Inventory	Goods on order	Shipments to customers	Arrivals from supplier	Orders for goods	Business cycles
Capital investment	Capital plant	Plant under construction	Depreciation	Construction completion	New contracts	Construction cycles
Equipment	Equipment	Equipment on order	Depreciation	Equipment delivery	New equipment orders	Business cycles
Human resources	Employees	Vacancies & trainees	Layoffs and quits	Hiring rate	Vacancy creation	Business cycles
Cash management	Cash balance	Accounts Receivable	Expenditures	Payments received from customers	Bills issues to customers	Fluctuations in cash flow and account balances
Marketing	Customer base	Prospective customers	Defections to competitors	Recruitment of new customers	New customer contacts	Boom and bust in customer base
Hog farming	Hog stock	Immature and gestating hogs	Slaughter rate	Maturation rate	Breeding rate	The “hog cycle”
Agricultural commodities	Inventory	Crops in the field	Consumption	Harvest rate	Planting rate	Commodity cycles
Commercial real estate	Building stock	Buildings under development	Depreciation	Completion rate	Development rate	Real estate booms and busts
Cooking on electric range	Temperature of pot	Heat in coils of range	Diffusion to air	Diffusion from coils to pot	Setting of burner	Overcooked dinner
Driving	Distance to next car	Momentum of car	Friction/drag	Velocity	Gas and brake pedals	Stop-and-go traffic
Showering	Water temperature	Water temp. in pipes	Drain rate	Flow from showerhead	Faucet settings	Freeze-then-burn
Blood sugar regulation	Glucose in bloodstream	Sugar and starch in GI tract	Metabolism	Digestion	Food consumption	Cycles of energy level
Social drinking	Alcohol in blood	Alcohol in stomach	Metabolism of alcohol	Diffusion from stomach to blood	Alcohol consumption rate	Drunkenness

Table 1 Examples of the stock management structure

Turning to the decision rule, the formulation for orders captures the decision-making process of the managers. Following the principles outlined by Sterman (2000, Ch. 13), such formulations must be based only on information actually available to the decision makers, must be robust under extreme conditions, and must be consistent with knowledge of the actual decision-making process, even if the way people actually make decisions is less than optimal. In most stock management situations the complexity of the feedbacks among the variables makes it impossible to determine the optimal strategy. Instead, people use heuristics or rules of thumb to determine the order rate. The ordering decision rule proposed here assumes that managers, unable to optimize, instead exercise control through a locally rational heuristic. The model thus falls firmly in the tradition of bounded rationality and the behavioral theory of the firm pioneered by Simon (1982) and Cyert and March (1963).

Three considerations are fundamental to any decision rule for orders. First, managers should replace expected losses from the stock. Second, managers should reduce the discrepancy between the desired and actual stock by ordering more than expected losses when the stock is less than desired and less than expected losses when there is a surplus. Third, managers should pay attention to the supply line of unfilled orders, and adjust orders to eliminate any discrepancies between the desired and actual supply line.

To formalize this intuition, first note that the order rate in most real life situations must be nonnegative:

$$OR = MAX(0, IO) \tag{6}$$

where IO is the indicated order rate, the rate indicated by other pressures. Order cancellations are sometimes possible and may sometimes exceed new orders. The costs of, and administrative procedures for, cancellations are likely to differ from those for new orders. Cancellations, if possible, should therefore be modeled as a distinct outflow from the supply line, governed by a separate decision rule, rather than as negative orders (Sterman 2000, Ch. 19 provides a suitable formulation).

The indicated order rate is formulated as an anchoring and adjustment process (Tversky & Kahneman 1974). Managers are assumed to base orders on the desired acquisition rate, AR^*, which is the rate at which they would like to add items to the stock. Managers then adjust orders above or below the desired acquisition rate by an amount designed to bring the supply line of unfilled orders in line with its goal (the Adjustment for the Supply Line, A_{SL}):

$$IO = AR^* + A_{SL} \tag{7}$$

The desired acquisition rate is similarly formulated as an anchoring and adjustment process. Managers seek to replace expected losses, L^e, modified by an amount designed to bring the stock in line with its goal (the Adjustment for the Stock, A_S):

$$AR^* = L^e + A_S \tag{8}$$

Why does the desired acquisition rate depend on expected losses rather than the actual loss rate? The current value of a flow represents the instantaneous rate of change. Actual instruments and information systems, however, cannot measure instantaneous rates of change but only average rates over some finite interval. The velocity of an object is calculated by measuring how far it moves over some period of time and taking the ratio of the distance covered to the time interval. The result is the average speed over the interval. The actual speed throughout the interval can vary, and the velocity at the finish line may differ from average. Similarly, the sales rate of a company *right now* cannot be measured. Instead sales rates are estimated by accumulating total sales over some interval of time such as an hour, day, week, month, or quarter. The reported sales rate is the average over the reporting interval, and sales at the end of the period may differ from the average over the interval. No matter how accurate the instruments, the rate of change measured and reported to an observer always differs from the instantaneous rate of change.

While in principle all flows are measured and reported with a delay, in practice the delay is sometimes so short relative to the dynamics of interest that it can safely be omitted. If the loss rate is directly observable by the decision maker with essentially no delay or measurement error it can be acceptable to assume L^e is the actual loss rate. Most often, however, the loss rate is not directly observable and must be estimated, introducing measurement, reporting, and perception delays. Further, even if losses are reported frequently, with little lag, it may be necessary or desirable to filter and smooth those data. For example, most manufacturing firms do not use raw order or shipment data as direct inputs to orders for materials or the production start rate. Orders and shipments are typically quite noisy, while it is costly to change production. Hence firms deliberately filter out high-frequency noise in shipments or customer orders so as to avoid overreacting to temporary variations. Such filtering is often accomplished with exponential smoothing or other forms of moving averages. Finally, expected losses might also include knowledge of seasonal variations or other factors.

The feedback structure of the ordering heuristic is shown in the bottom part of Figure 3. The adjustment for the stock A_S creates the balancing (negative) Stock Control feedback loop. The simplest formulation is to assume the adjustment is linear in the discrepancy between the desired stock S^* and the actual stock:

$$A_S = (S^* - S)/\tau_S \tag{9}$$

where S^* is the desired stock and τ_S is the stock adjustment time (equivalently, $1/\tau_S$ is the fraction of the discrepancy between desired and actual inventory ordered per time unit). The desired stock may be a constant or a variable.

The adjustment for the supply line is formulated analogously to the adjustment for the stock:

$$A_{SL} = (SL^* - SL)/\tau_{SL} \tag{10}$$

where SL^* is the desired supply line and τ_{SL} is the supply line adjustment time. The supply line adjustment forms the negative Supply Line Control loop.

Figure 3 does not show the feedback structure for the desired supply line. In some cases the desired supply line is constant. More often, however, decision makers seek to maintain a sufficient number of units on order to achieve the acquisition rate they desire. By Little's Law the supply line must, in equilibrium, contain λ time units worth of the throughput the decision maker desires to achieve. Several measures for desired throughput are common. The decision maker may set the desired supply line to yield the desired acquisition rate, AR^*:

$$SL^* = \lambda^e * AR^* \tag{10a}$$

where λ^e, the expected acquisition lag, represents the decision maker's current belief about the length of the acquisition delay (which, in general, may differ from the actual acquisition delay).

Equation (11a) assumes a rather high degree of rationality on the part of decision makers. They are assumed to adjust the supply line to achieve the desired acquisition rate, which includes replacement of expected losses and correction of temporary gaps between desired and actual inventory. As described below, experimental evidence shows decision makers are often not so sophisticated. Managers frequently do not adjust the supply line in response to temporary imbalances in the stock but base the desired supply line on their estimate of long-run throughput requirements—the expected loss rate L^e:

$$SL^* = \lambda^e * L^e \tag{11b}$$

The formulation for the desired supply line should be based on empirical investigation of the actual decision-making process (e.g., Senge 1980, Croson *et al.* 2013, Sterman & Dogan 2014).

Whichever formulation for the desired supply line is used, the longer the expected delay in acquiring goods or the larger the desired throughput rate, the larger the supply line must be. If a retailer wishes to receive 1,000 widgets per week from the supplier and delivery requires six weeks, the retailer must have 6,000 widgets on order to ensure an uninterrupted flow of deliveries. The adjustment for the supply line creates a negative feedback loop that adjusts orders to maintain the acquisition rate at the desired value given the (expected) delay between orders and delivery. Without the supply line feedback, orders would be placed even after the supply line contained sufficient units to correct stock shortfalls, producing overshoot and instability. The supply line adjustment also compensates for changes in the acquisition lag. If the acquisition lag doubled, for example, the supply line adjustment would induce sufficient additional orders to restore acquisitions to the desired rate.

There are many possible ways managers may form the expected acquisition lag λ^e, ranging from constants through guesstimates to sophisticated forecasts. Usually, it takes time to detect changes in delivery times. Customers often do not know that goods they ordered will be late until after the promised delivery time has passed. The expected acquisition lag can then be modeled by a perception delay representing the time required to observe and respond to changes in the actual delay. For example, Senge (1980) found expected delivery times for capital plant and equipment lagged the actual delivery times by 1.3 years for firms in the US economy.

The formulation for the order rate conforms to core principles for behavioral models (Sterman 2000, Ch. 13). First, the formulation is robust: Orders remain nonnegative no matter how large a surplus stock there may be, and the supply line and stock therefore never fall below zero. Second, information not available to real decision makers is not utilized (such as the instantaneous value of the loss rate, or the solution to the dynamic programming problem determining the optimal order rate). Finally, the ordering decision rule is grounded in well-established knowledge of decision-making behavior, such as the anchoring and adjustment heuristic. Expected losses form an easily anticipated and relatively stable starting point for the determination of orders. Loss rate information will typically be locally available and highly salient to the decision maker. Replacing losses will keep the stock constant at its current level. Adjustments are then made in response to the adequacy of the stock and supply line. No assumption is made that these adjustments are optimal. Rather, pressures arising from the discrepancies between desired and actual quantities cause managers to adjust the order rate above or below the level that would maintain the status quo.

The Stock Management Problem: Dynamics

To illustrate the behavior of the stock management structure, consider how a manufacturing firm manages its inventory of product. Figure 4 adapts the generic stock management structure to the case of inventory and production control. The firm maintains a stock of finished inventory and fills orders as they arrive. In this simple illustration, assume that customers are delivery sensitive—orders the company cannot fill immediately are lost as customers seek other sources of supply (Sterman 2000, Ch. 18, extends the model to add an explicit backlog of unfilled orders). Production takes time. The supply line is the stock of work in process inventory (WIP), which is increased by production starts and decreased by production.

The key production control and inventory management decisions are order fulfillment (determining the ability to fill customer orders based on the adequacy of inventory) and production scheduling (determining the rate of production starts based on the demand forecast and inventory position of the firm, including the WIP inventory). The model includes three important negative feedbacks. The Stockout loop regulates shipments as inventory varies: If inventory is inadequate, some items will be out of stock and shipments fall below orders. In the extreme, shipments must fall to zero when there is no inventory. The Inventory Control and WIP Control loops adjust production starts to move the levels of inventory and WIP toward their desired levels. In this initial model there are no stocks of materials and no capacity constraints (either from labor or capital). These extensions are treated in Sterman 2000, Ch. 18.

For purposes of illustration, we assume there are no capacity constraints or materials shortages that might limit production starts, so the actual production start rate is equal to the desired production start rate. Following the standard stock management structure, desired production starts are anchored on desired production, then adjusted to bring the stock of WIP in line with the desired WIP level. Desired production, in turn, is anchored on the Expected Order Rate, then adjusted to bring the stock of finished goods inventory in line with the desired level. Because incoming customer orders are typically noisy, the firm, as is common, uses first-order exponential smoothing to filter out the high-frequency random variations in customer orders. The firm seeks to maintain enough finished goods inventory to provide excellent customer service, and so seeks to maintain a certain desired number of weeks of inventory coverage.

Desired inventory coverage consists of the minimum time required to process and ship orders plus safety stock coverage large enough to provide excellent customer service. The model is fully documented in Sterman 2000, Ch. 18.[4]

Note that the model does not include the main operational factors that can make amplification a rational outcome (Lee *et al.* 1997). There are no quantity discounts, so order batching is never rational. Prices are constant, so there is never any incentive to order more (less) in the expectation that prices will rise (fall). Each customer has only one supplier and each supplier only one customer, so there is no incentive to place phantom orders.

Now consider the impact of an unanticipated 20% increase in customer orders, from an initial equilibrium with throughput of 10,000 units/week (Figure 5).

The desired shipment rate rises immediately after the step increase in demand. Inventory coverage immediately drops from its initial value of four weeks to 3.33 weeks. In the instant after the customer order rate jumps, inventory has not yet changed, and the firm is initially able to fill nearly all the incoming orders, despite the increase. However, because production continues at the initial rate of 10,000 widgets/week, inventory falls. As inventory falls, so too does the firm's ability to ship. A maximum of about 5% of orders go unfilled and are lost (along, most likely, with the firm's reputation as a reliable supplier).

The growing gap between desired and actual inventory forces desired production to rise above expected orders. As it does the quantity of work in process required to meet the higher production goal also grows, opening a gap between the desired and actual level of WIP. Thus the desired production start rate rises further above the desired production rate.

As time passes the firm recognizes that the initial increase in demand is not a mere random blip and gradually raises its demand forecast. As expected orders rise, so too does desired inventory, further increasing the gap between desired and actual inventory and boosting desired production still more. Production starts reach a peak more than 42% above the initial level about four weeks after the shock, 210% more than the change in customer orders.

---

[4] The parameters in the simulation in Figure 5 are:

Parameter	Base Case Value (Weeks)
Minimum Order Processing Time	2
Safety Stock Coverage	2
Manufacturing Cycle Time	8
Inventory Adjustment Time	8
WIP Adjustment Time	2
Time to Average Order Rate	8

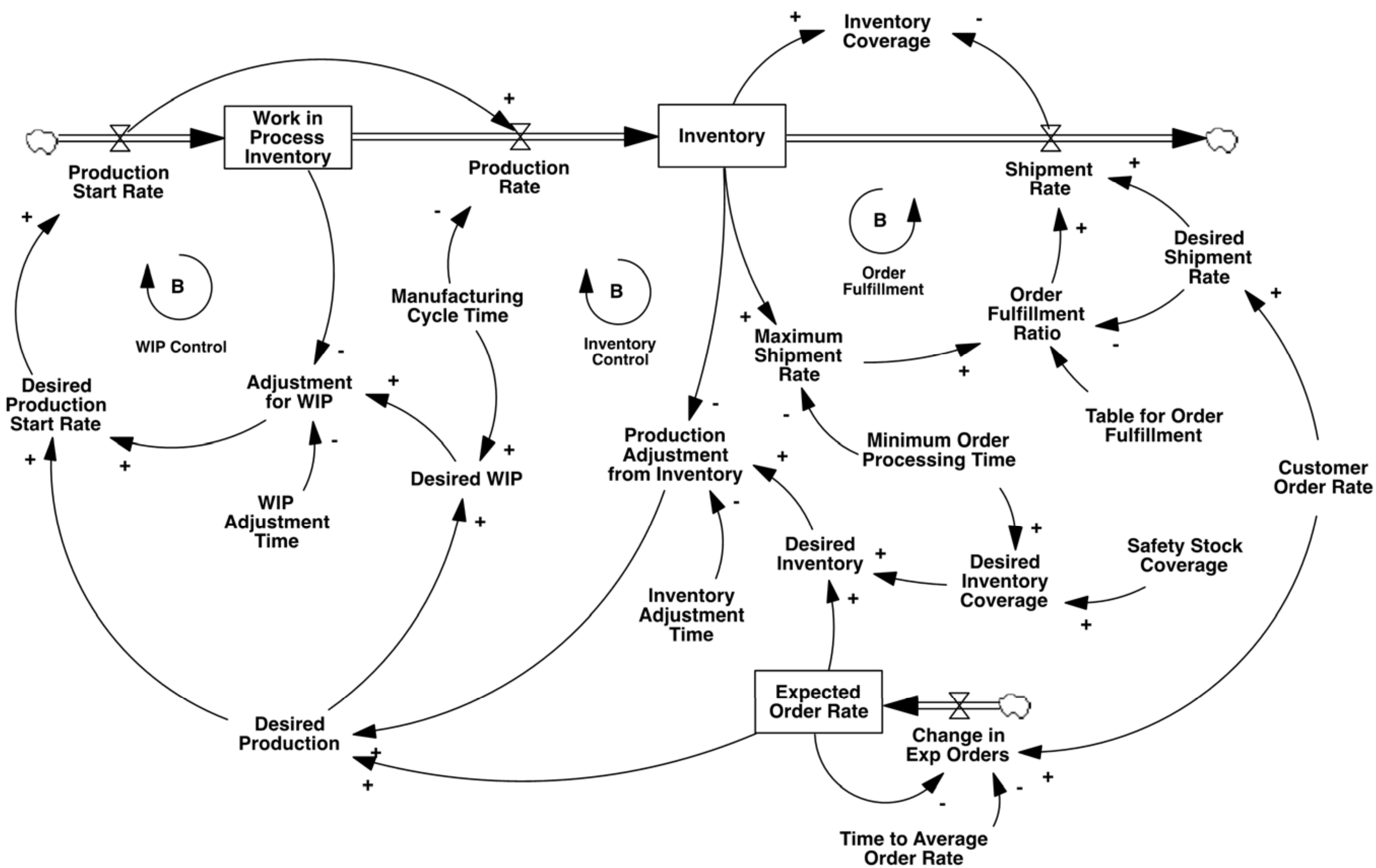

Figure 4 The stock management structure adapted for a manufacturing firm

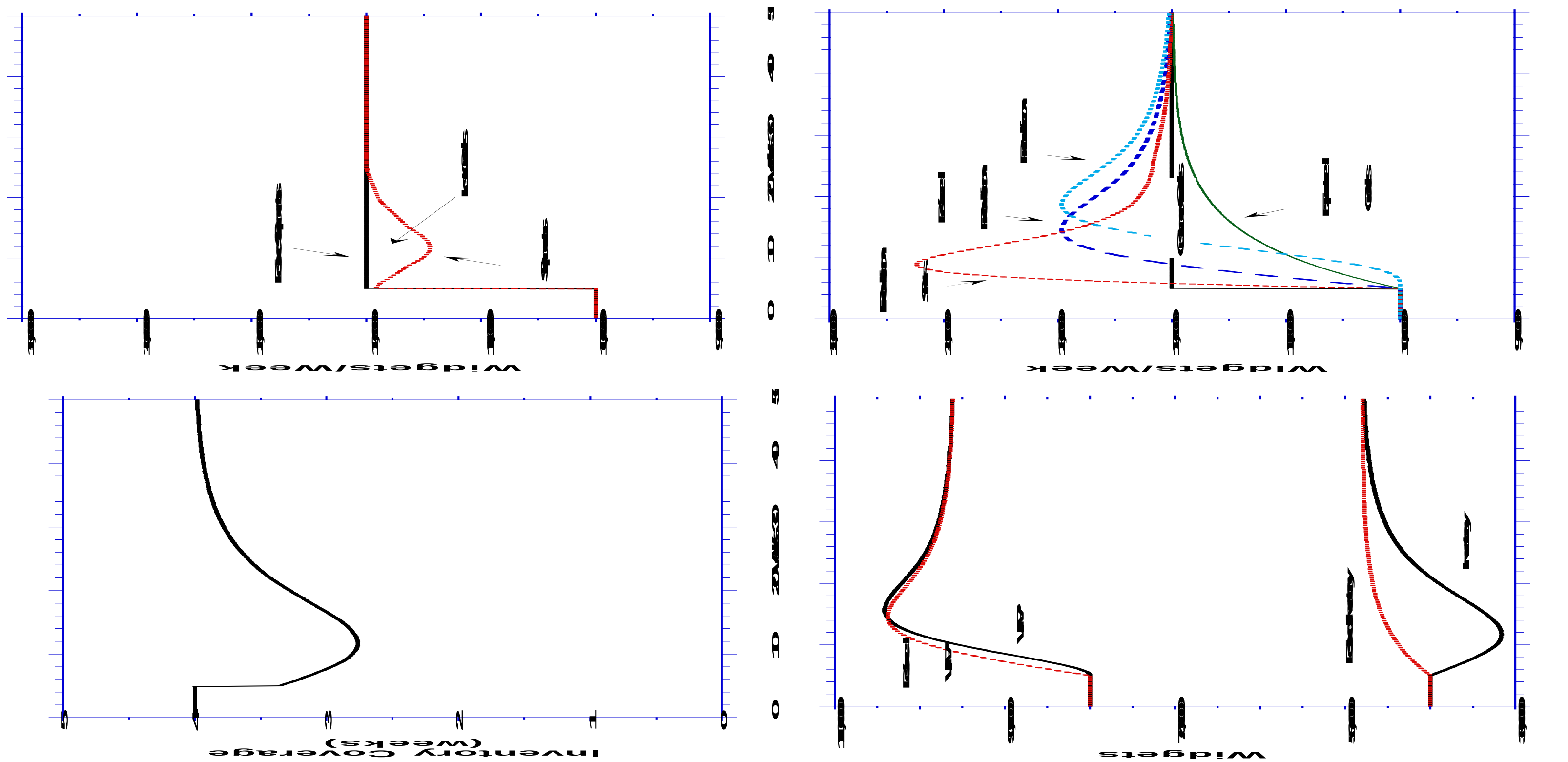

Figure 5 Response of the manufacturing model to a 20% step increase in orders

The rapid increase in production starts soon fills the supply line of WIP, but production lags behind due to the eight week manufacturing delay. Production does not surpass shipments until more than six weeks have passed; throughout this period inventory continues to fall even as the desired inventory level rises. Inventory stops falling when production first equals shipments. The system is not yet in equilibrium, however, because of the large gap between desired and actual inventory and between orders and expected orders. Production eventually rises above shipments, causing inventory to rise until it eventually reaches the new, higher desired level. Note that the peak of production comes about one-quarter year after the change in orders, much longer than the eight-week production delay suggests.

The consequences of the stock management structure for supply chain management are profound.

1. The process of stock adjustment creates significant amplification. The initial response of the firm to an unanticipated increase in demand is a decline in inventory. The production delay means an initial drop in inventory is inevitable—it is a fundamental consequence of the physical structure of the system. The reduction in inventory contrasts sharply with the firm's desire to hold more inventory when demand increases to maintain acceptable inventory coverage and customer service.

2. Amplification of the demand shock is unavoidable. Because inventory must initially fall, the only way to increase it back to its initial level and then raise it to the new, higher desired level is for production to exceed shipments. Production must overshoot the shipment rate long enough and by a large enough margin to build inventory up to the new desired level. Production starts must overshoot orders even more so that the level of WIP can be built up to a level consistent with the higher throughput rate.

3. The peak production start rate must lag the change in customer orders. The adjustment to production from the inventory gap reaches its maximum about when the inventory reaches its minimum. Inventory bottoms out only after production has finally risen enough to equal shipments, an event that must lag the change in orders. Like amplification, phase lag is a fundamental and inevitable consequence of the physical stock and flow structure.

4. Amplification is temporary. In the long run, a 20% increase in customer orders leads to a 20% increase in production starts. But during the disequilibrium adjustment to the new equilibrium, production starts must temporarily rise above orders since that is the only way inventory can be rebuilt to its initial level, and the only way inventory and WIP stocks can rise from their initial levels to the new, higher equilibrium levels consistent with higher customer demand.

The firm's suppliers therefore face much larger changes in demand than the firm itself, and much of that surge in demand is temporary. Upstream firms, such as those supplying plant, equipment and materials, will not face a single, permanent change in orders but a much larger, and temporary, surge in demand. Each supplier will, for the same reasons, necessarily amplify and delay the change in orders they receive. As that signal is passed up the supply chain to their suppliers, and theirs, the result is the characteristic amplification and phase lag observed in commodities, construction, and so many other industries.

The stock management structure thus explains why supply chains generate amplification and phase lag. Given the structure of the system, specifically, production delays and forecast adjustment delays, production and production starts must overshoot, amplify, and lag changes in demand, no matter how smart the managers of the firm may be. Amplification and phase lag arise even though there is no order batching, no price variations, and no horizontal competition among customers for limited supply. Although those factors may indeed contribute to amplification in supply chains, they are not necessary.

Though amplification and phase lag are inevitable, oscillation is not. Even though the actors use boundedly rational and not optimal decision rules, the response to the demand shock shown in Figure 5 is smooth and stable (given the base case parameters).

C. Case Example

The simulation above shows how a single link in a supply chain creates amplification and phase lag in response to changes in customer demand. Do people understand why? Unfortunately, the answer is no. To illustrate, Figure 6 shows a simple learning exercise, the Manufacturing Case (MC), which assesses people's understanding of the stock management structure in an extremely simple context (Booth Sweeney and Sterman 2000).

The manufacturing case is an example of a simple stock management task. Here, a firm seeks to control its inventory in the face of an unanticipated step increase in customer demand and a lag between a change in the production schedule and the actual production rate, analogous to the stock management structure in Figure 4.

Consider a manufacturing firm. The firm maintains an inventory of finished product. The firm uses this inventory to fill customer orders as they come in. Historically, orders have averaged 10,000 units per week. Because customer orders are quite variable, the firm strives to maintain an inventory of 50,000 units to provide excellent customer service (that is, to be able to fill essentially 100% of every order), and they adjust production schedules to close any gap between the desired and actual level. Although the firm has ample capacity to handle variations in demand, it takes time to adjust the production schedule, and to make the product – a total lag of four weeks.

Now imagine that the order rate for the firm's products suddenly and unexpectedly rises by 10%, and remains at the new, higher rate indefinitely, as shown in the graph below. Before the change in demand, production was equal to orders at 10,000 units/week, and inventory was equal to the desired level of 50,000 units.

Sketch the likely path of production and inventory on the graphs below. Provide an appropriate scale for the graph of inventory.

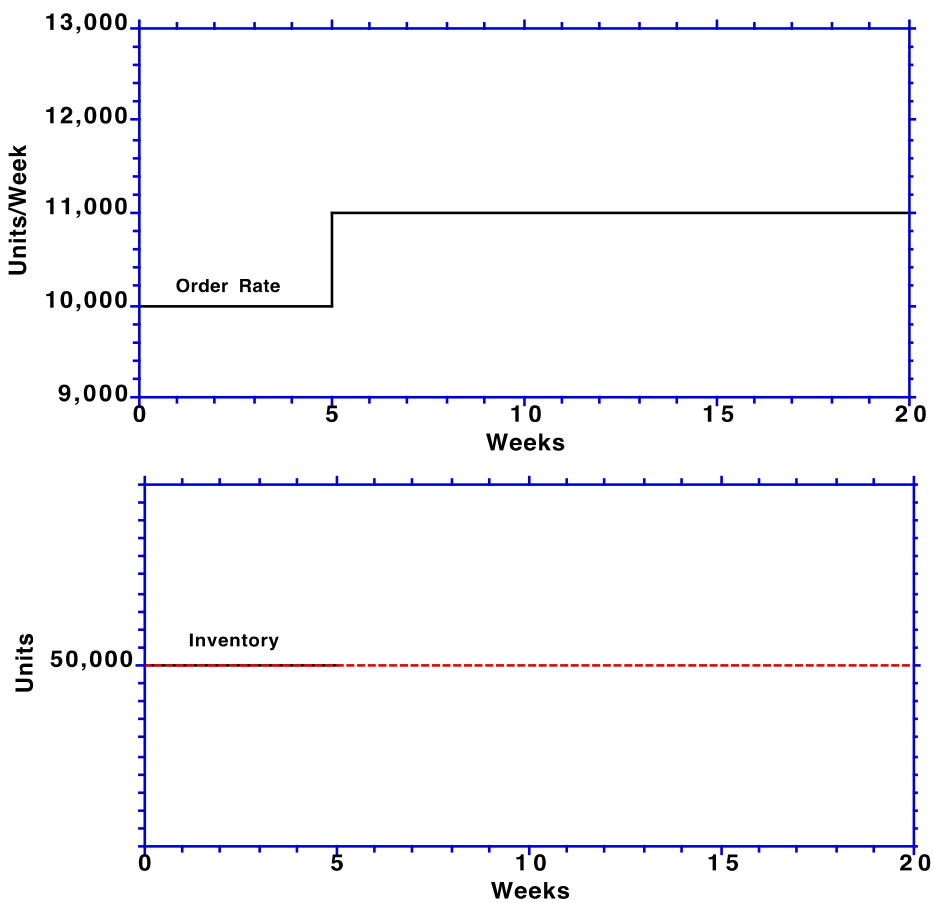

Figure 6 The Manufacturing Case (Booth Sweeney & Sterman 2000)

There is no unique correct answer to the MC task. However, the trajectories of production and inventory must follow certain constraints, and their shapes can be determined without any quantitative analysis. The unanticipated step increase in customer orders means shipments increase, while the production delay means production remains, for a time, constant at the original rate. Inventory therefore declines. The firm must not only boost output to the new rate of orders, but also rebuild its inventory to the desired level. Production must therefore overshoot orders and remain above shipments until inventory reaches the desired level, at which point production can drop back to equilibrium at the customer order rate.

Furthermore, since the task specifies that the desired inventory level is constant, the area bounded by the production overshoot must equal the quantity of inventory lost during the period when orders exceed production, which in turn is the area between orders and production between week five and the point where production rises to the order rate. A few modest assumptions allow the trajectories of production and inventory to be completely specified. When customer orders increase from 10,000 to 11,000 widgets/week, production remains constant at the initial rate, due to the four-week lag. Inventory, therefore, begins to decline at the rate of 1,000 widgets/week. What happens next depends on the distribution of the production lag. The simplest case, and the case most participants assumed, is to assume a pipeline delay, that is,[5]

$$\text{Production}(t) = \text{Desired Production}(t - 4).$$

Assuming production follows desired production with a four week delay means production continues at 10,000 units/week until week nine. During this time, inventory drops by a total of 1,000 units/week * 4 weeks = 4,000 units, thus falling to 46,000 units. Assuming further that the firm understands the delay and realizes that production will remain at its original level for four weeks, management will raise desired production above orders at week five, keep it above orders until an additional 4,000 units are scheduled for production, and then bring desired production back down to orders. Production then traces this pattern four weeks later. Assuming finally that production remains constant during the period of overshoot gives production trajectories such as those shown in Figures 7 and 8. Figure 7, typical of many correct responses, shows production rising in week nine to 12,000 units/week and remaining there for the next four weeks, giving a rectangle equal in shape to that for the period $5 < t \leq 9$ when shipments exceed production. Of

[5] Other patterns for the delay are possible, such as some adjustment before week nine and some after, and were coded as correct as long as production did not begin to increase until after the step increase in orders.

course, the production overshoot can have any shape as long as the area equals 4,000 widgets. Figure 8 shows an unusual correct response in which the participant shows production rising in week nine to 13,000 widgets/week and remaining there for two weeks. This response clearly shows the subject understood the task well. However, that participant was the only one, out of 225, who drew a pattern with the duration of the overshoot ≠ 4 weeks while also maintaining the correct area relationship.

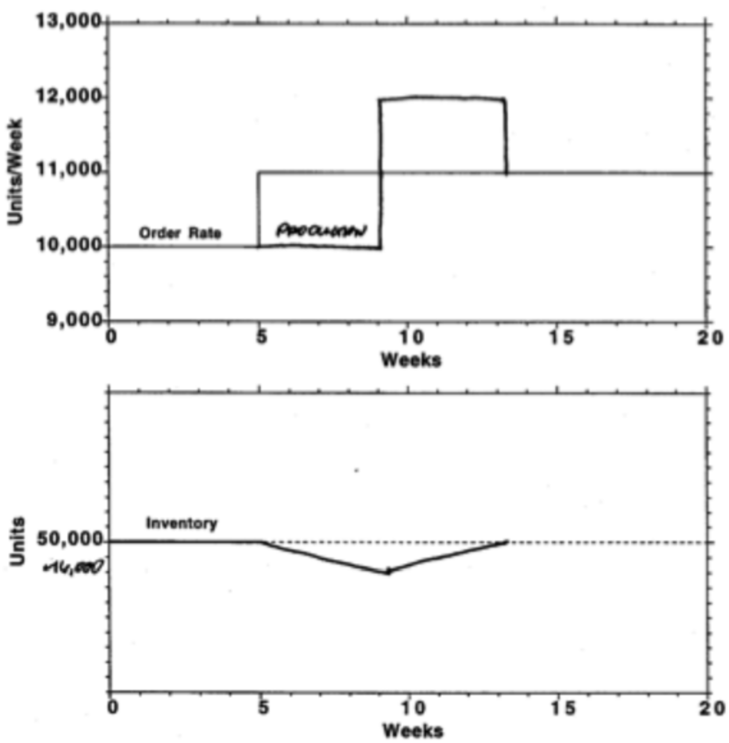

Figure 7 A correct response to the manufacturing case. Note that the path of inventory is consistent with the path of production.

Figure 8 An unusual correct response to the manufacturing case

It is possible that production and inventory could fluctuate around their equilibrium values, but while such fluctuation is not inevitable, the overshoot of production is: the only way inventory can rise is for production to exceed orders, in exactly the same way that the only way the level of water in a bathtub can rise is for the flow in from the tap to exceed the flow out through the drain.

The MC Case is quite simple, involving only one stock, one time delay, and one negative feedback loop. Nevertheless, in a group consisting of MBA students, executive MBA students and other graduate students at the MIT Sloan School of Management (N = 225), performance was poor. Only 44% of the participants showed production overshooting orders. Instead, most showed production adjusting with a lag to the new customer order rate but not overshooting: they failed to understand that building inventory back up to its desired level requires production to exceed orders.

Figure 9 shows typical erroneous responses. The top panel shows the most common error. The participant shows production responding with a lag, but rising up only to the new level of orders. There is no production overshoot. Further, the trajectory of inventory is inconsistent with the production path. The subject shows inventory falling linearly through about week 10 (although given the production path as drawn inventory would actually fall at a diminishing rate). Worse, the subject then shows inventory rising even though production equals orders after week ten.

Booth Sweeney and Sterman (2000) explored several variants of the task. In one, participants were asked to sketch the paths for both production and inventory (as in Figure 7). In another, participants were asked only to sketch the trajectory of production (as in Figure 8). Booth Sweeney and Sterman hypothesized that requiring participants to draw the path for inventory would help them realize the need for production to overshoot customer orders (so as to rebuild inventory to the target level). Overall performance in the inventory graph condition, however, was significantly worse than in the no inventory graph condition ($t = 5.11$, $p < 0.0001$). Only 23% of those in the inventory graph condition correctly showed production overshooting orders, compared to 63% of those in the no inventory graph condition.

Overall, 89% of the participants drew production trajectories that violated conservation of material, showing no production overshoot or an overshoot whose area does not equal the area of the production undershoot they drew. The failure to conform to conservation of mass has been repeatedly demonstrated in similar experiments with different cover stories, including greenhouse gases accumulating in the atmosphere, receipts, and disbursements accumulating in a bank account, and even water filling a bathtub (Sterman & Booth Sweeney 2000, Sterman & Booth Sweeney 2007, Sterman 2008, Cronin *et al.* 2009, Sterman 2010).

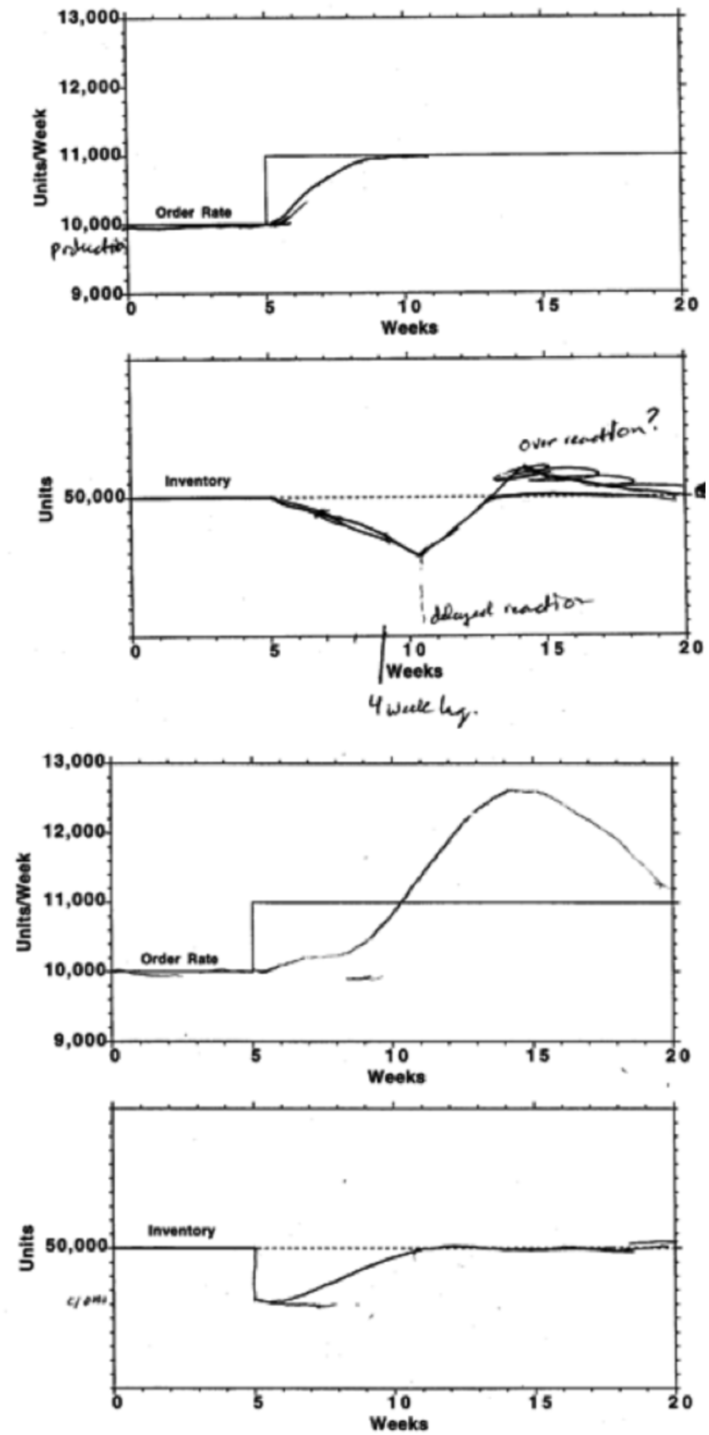

Figure 9 Typical incorrect responses to the Manufacturing Case

The experimental evidence shows that many people, including many with extensive training in Science, Technology, Engineering and Mathematics (STEM), do not understand the most basic principles of accumulation. Inventory control and supply chain management depend fundamentally on accumulations: inventories accumulate production less shipments, backlogs accumulate orders less fulfillment and cancellations, and so on. If people do not understand the basic principles of accumulation it should be no surprise that we continue to observe dysfunctional dynamics in supply chains across a wide variety of industries and products.

D. Learning Activity

The simulation of the stock management structure above explains the origin of amplification and phase lag, but does not exhibit oscillations. How do oscillations arise? Oscillations can arise only when there are time delays in the negative feedbacks controlling the state of the system (Sterman 2000, Ch. 4). The mere existence of a supply line and acquisition lag, however, does not necessarily lead to oscillations. In the manufacturing model above, there is an eight week delay between the start and completion of the manufacturing process, yet the system does not oscillate (with the estimated parameters). In that model, managers fully account for the stock of WIP—the supply line of units in production but not yet received, and reduce orders as soon as they have initiated enough new production to bring inventory up to the desired level even though those units have not yet entered the finished goods inventory.

To oscillate, the time delay must be (at least partially) ignored. The manager must continue to initiate corrective actions in response to the perceived gap between the desired and actual state of the system even after sufficient corrections to close the gap are in the pipeline. But do managers ignore these time delays and the supply line of corrective actions? In many settings, shockingly, the answer is yes.

The Beer Distribution Game

The Beer Distribution Game illustrates how oscillations arise.[6] The game is a role-playing simulation of a supply chain originally developed by Jay Forrester in the late 1950s to introduce students of management to the concepts of system dynamics and computer simulation. Since

---

[6]The game is described in detail in Sterman (1989a, 1992). Information and materials are available from the System Dynamics Society at system.dynamics@albany.edu. There is no real beer in the beer game and it does not promote drinking. When the game is used with, e.g., high school students, it is easily recast as the "apple juice game." Many firms have customized the game to represent their industry.

then the game has been played all over the world by tens of thousands of people ranging from high school students to chief executive officers and senior government officials.

The game is played on a board portraying a typical supply chain (Figure 10). Markets and chips represent orders for and cases of beer. Each brewery consists of four sectors: retailer, wholesaler, distributor, and factory (R, W, D, F). One person manages each sector. A deck of cards represents customer demand. Each week, customers demand beer from the retailer, filling the order out of inventory. The retailer in turn orders beer from the wholesaler, who ships the beer requested from wholesale stocks. Likewise the wholesaler orders and receives beer from the distributor, who in turn orders and receives beer from the factory. The factory produces the beer. At each stage there are order processing and shipping delays. Each link in the supply chain has the same structure.

The players' objective is to minimize total costs for their company. Inventory holding costs are typically set to $0.50 per case per week, and stockout costs (costs for having a backlog of unfilled orders) to $1.00 per case per week. The task facing each player is a clear example of the stock management problem. Players must keep their inventories as low as possible while avoiding backlogs. Incoming orders deplete inventory, so players must place replenishment orders and adjust their inventories to the desired level. There is a delay between placing and receiving orders, creating a supply line of unfilled orders.

The standard game is played with a very simple pattern for customer demand. Starting from equilibrium, there is a small, unannounced one-time increase in customer orders, from four to eight cases per week.

The game is far simpler than any real supply chain. There are no random events—no machine breakdowns, transportation problems, or strikes. There are no capacity constraints or financial limitations. The structure of the game is visible to all. Players can readily inspect the board to see how much inventory is in transit or held by their teammates.

Further, the main operational factors that can make amplification rational do not apply in the beer game: as in the simulation model in Figures 4-5, the operational factors cited by Lee *et al.* (1997) that may lead to amplification of demand are absent. There are no quantity discounts that could make order batching rational, no price variations that could make forward buying rational, and no horizontal competition among customers for limited supply that could make phantom ordering or inventory hoarding rational.

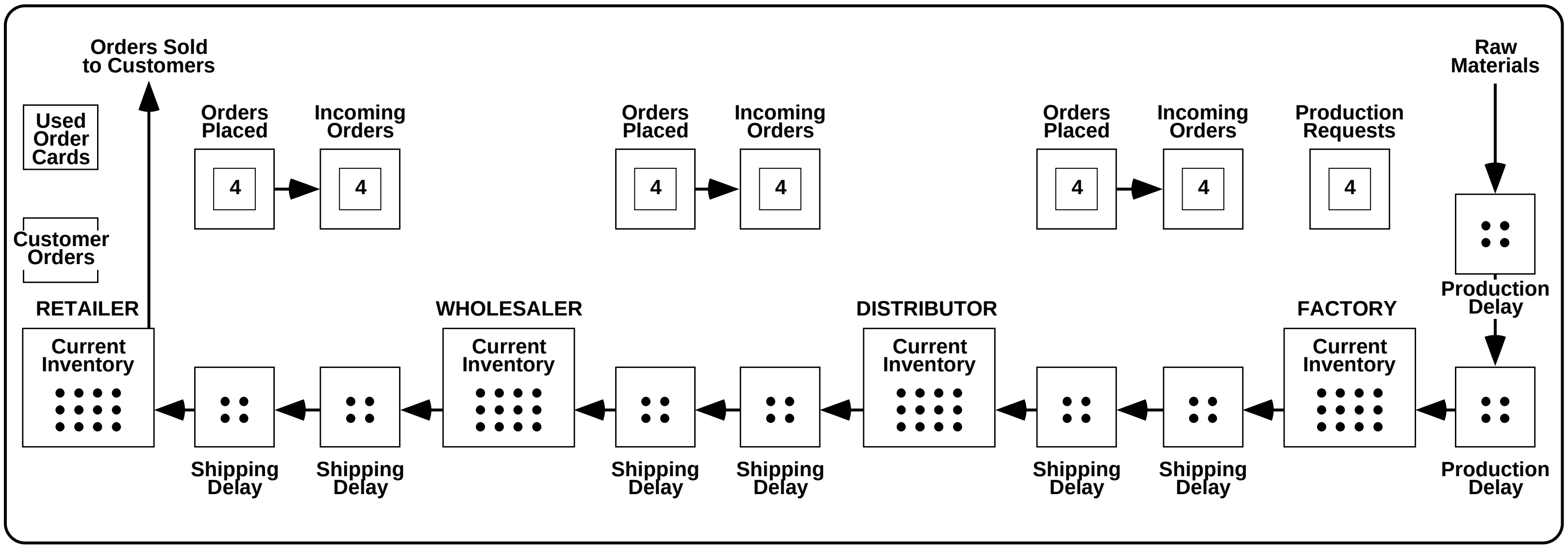

Figure 10 The Beer Distribution Game

The game is a role-play simulation. Each player manages one of the links in the distribution chain from Retailer to Factory. In the game, chips of various denominations represent cases of beer and move through the supply chain from Raw Materials to Customers. Customer Orders are written on a deck of cards. Each week players place orders with the supplier on their left and the factory sets the production schedule. The orders, written on slips of paper, move upstream (left to right). The initial configuration is shown.

Despite the simplicity of the game, however, people do extremely poorly. Among first-time players average costs are typically an astonishing 10 times greater than optimal (Sterman 1989a). Figure 11 shows typical results. In all cases customer orders are essentially constant (except for the small step increase near the start). In all cases, the response of the supply chain is unstable. The oscillation, amplification, and phase lag observed in real supply chains are clearly visible in the experimental results. The period of the cycle is 20-25 weeks. The average amplification ratio of factory production relative to customer orders is a factor of four, and factory production peaks some 15 weeks after the change in customer orders.

Most interesting, the patterns of behavior generated in the game are remarkably similar (there are, of course, individual differences in magnitude and timing). Starting with the retailer, inventories decline throughout the supply chain, and most players develop a backlog of unfilled orders (net inventory becomes negative). In response, a wave of orders move through the chain, growing larger at each stage. Eventually, factory production surges, and inventories throughout the supply chain start to rise. But inventory does not stabilize at the cost-minimizing level near zero. Instead, inventory significantly overshoots. Players respond by slashing orders, often cutting them to zero for extended periods. Inventory eventually peaks and slowly declines. These behavioral regularities are all the more remarkable because there is no oscillation in customer demand. The oscillation arises as an endogenous consequence of the way the players manage their inventories. Though players are free to place orders any way they wish, the vast majority behave in a remarkably uniform fashion.

Modeling Managerial Behavior: Misperceptions of Feedback

To understand the origin of oscillations more formally, Sterman (1989a) tested the decision rule for customer orders described above against the order decisions of players in the game. Orders placed are given by the anchoring and adjustment rule used in the stock management structure developed above. Participants are assumed to anchor on their belief about expected incoming orders (the loss rate from their inventory), then adjust their orders based on the adequacy of their on-hand inventory and supply line of on-order inventory:

$$OR = MAX(0,\ D^e + A_S + A_{SL}) \tag{12}$$

where D^e is expected demand, that is the participant's belief about what the next incoming order will be. Sterman (1989a) assumes that expected incoming orders are formed by exponential smoothing of customer orders or demand, D, given in discrete time by

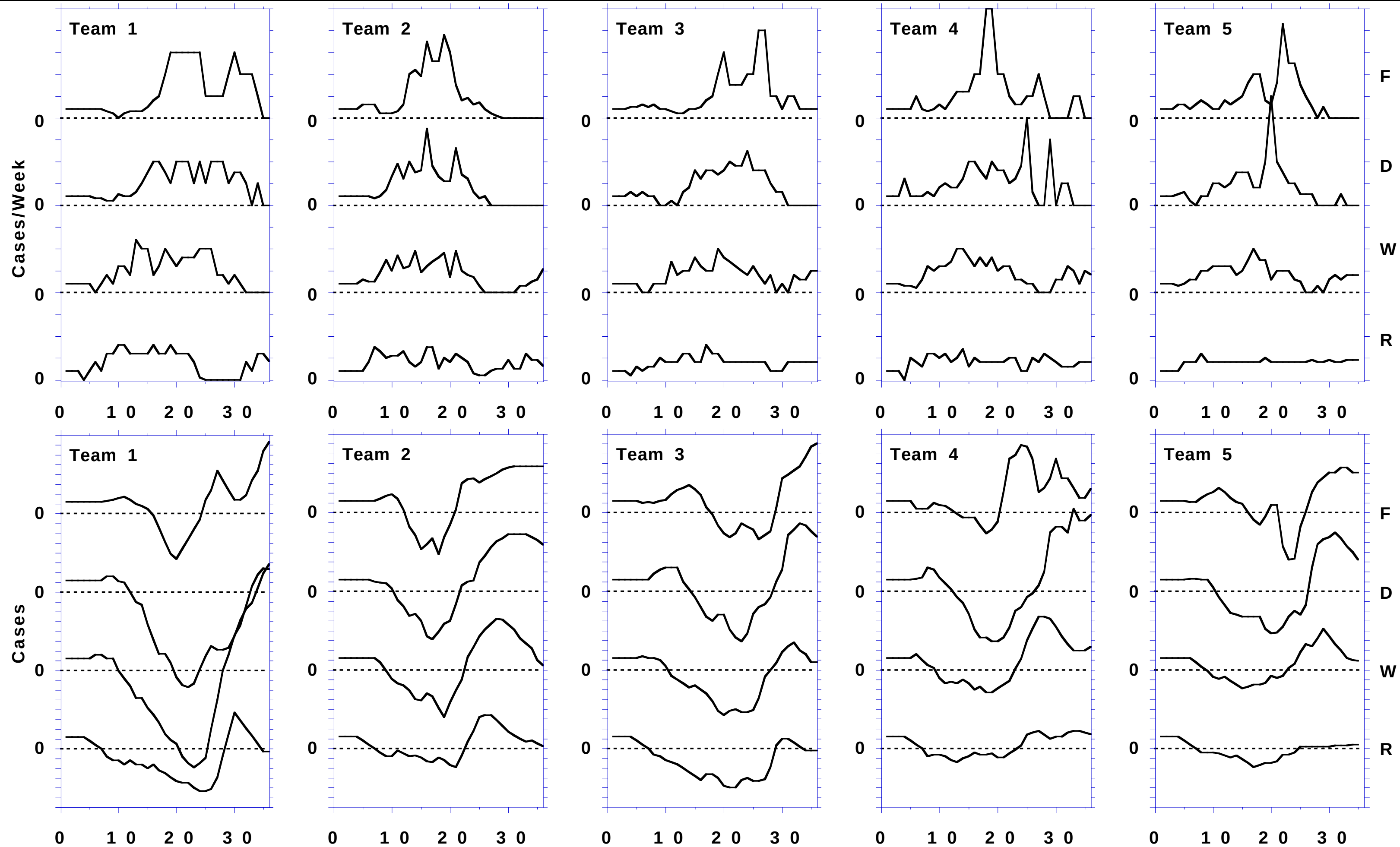

Figure 11 Typical results of the Beer Distribution Game

Top: Orders. *Bottom:* Net inventory (Inventory – Backlog). Graphs show, bottom to top, Retailer, Wholesaler, Distributor, and Factory. Vertical axis tick marks denote 10 units. Note the characteristic oscillation, amplification, and phase lag as the change in customer orders propagates from retailer to factory. *Source:* Sterman (2000).

$$D_t^e = \theta D_t + (1-\theta)D_{t-1}^e \tag{13}$$

where θ is the smoothing parameter.

Clark and Scarf (1960) showed that managers should give as much weight to on-order inventory (the supply line, SL) as on-hand inventory (the stock, S). If so, orders become

$$OR = MAX(0,\ D^e + ((S^* - S) + (SL^* - SL))/\tau_S) \tag{14}$$

However, participants might not fully account for the supply line of orders placed but not yet received. Even if participants can keep track of the supply line, the supply line is likely to be given less weight than on-hand inventory because on-hand inventory is the direct determinant of costs, is highly salient, and is right in front of the players, while the supply line has none of these attributes. If participants underweight the supply line, the ordering rule becomes

$$OR = MAX(0,\ D^e + \big((S^* - S) + \beta(SL^* - SL)\big)/\tau_S) \tag{15}$$

where $\beta = \tau_S/\tau_{SL}$ is the fraction of the supply line adjustment taken into account: when people underweight or ignore the supply line, τ_{SL} will be longer than the stock adjustment time; if the supply line adjustment is completely ignored, then $\tau_{SL} \to \infty$ which implies $A_{SL} \to 0$ (Sterman 1989a, 2000). Collecting terms and defining $S' = S^* + \beta SL^*$ yields

$$OR = MAX(0,\ D^e + (S' - (S + \beta SL))/\tau_S) \tag{16}$$

Assuming that the desired stock and desired supply line are constant and including an additive error term yields the system of equations to be estimated:

$$OR_t = MAX(0,\ D_t^e + \alpha\big(S' - (S_t + \beta SL_t)\big) + \varepsilon_t) \tag{17}$$

$$D_t^e = \theta D_t + (1-\theta)D_{t-1}^e \tag{18}$$

where $\alpha = 1/\tau_S$ is the fraction of the perceived inventory discrepancy ordered each week. There are four parameters to be estimated: the forecast updating time, $0 \le \theta \le 1$; the fraction of the perceived inventory discrepancy ordered each week, $0 \le \alpha \le 1$; the fraction of the supply line of unfilled orders taken into account, $0 \le \beta \le 1$; and the desired stock of on-hand and on-order inventory, $0 \le S'$.

Sterman (1989a) estimated eq. (17-18) by nonlinear least squares for a sample of 44 players. Overall, the decision rule worked quite well, explaining 71% of the variance in the order decisions of the participants. The estimated parameters showed that most were using grossly suboptimal cue weights. The average weight on the supply line was only 0.34. Only 25% of the participants considered more than half the supply line and the estimated value of β was not significantly different from zero for fully one-third. To illustrate, Figure 12 compares simulated and actual behavior for the factory in an actual game. The estimated fraction of the inventory discrepancy ordered each week is 0.8—the player reacted aggressively to inventory shortfalls, ordering nearly the entire inventory shortfall each week. At the same time, the estimated fraction of the supply line taken into account is zero—this participant completely ignored the supply line of orders placed but not yet received. As you would expect, aggressively reacting to current inventory shortfalls while completely ignoring the supply line leads to severe instability and high costs. Because it takes three weeks to receive production requested today, the player effectively orders almost three times more than needed to correct any inventory shortfall.

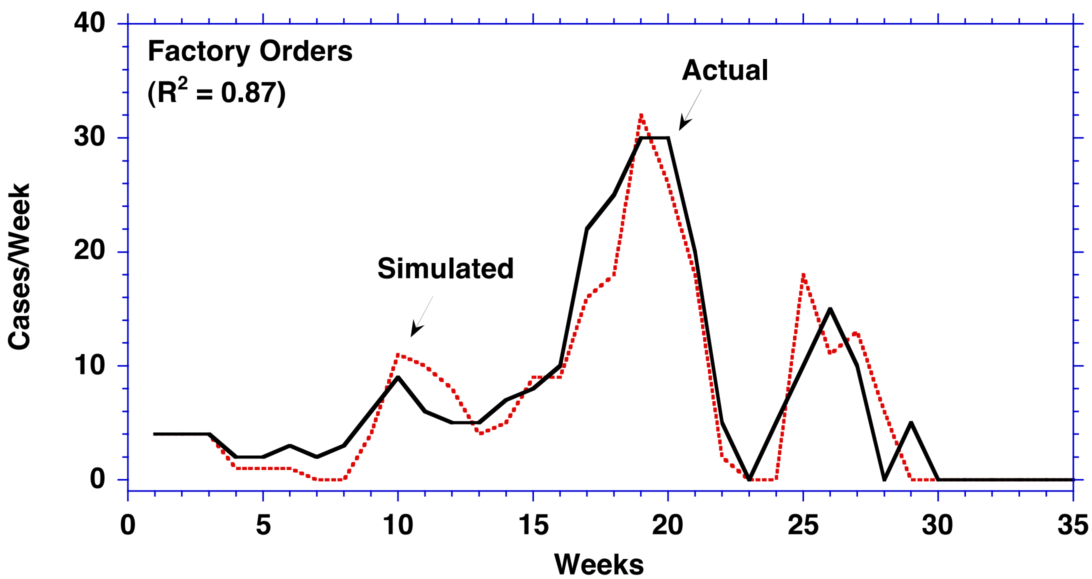

Figure 12 Estimated vs. actual behavior in the beer game

Parameters: θ = 0.55, $S' = 9$, α = 0.80, β = 0. *Source:* Sterman (1989a).

Other experiments with the beer game and similar stock management systems (for example, Sterman 1989b; Diehl &Sterman 1995; Brehmer 1992, Paich & Sterman 1993, Kampmann and Sterman 2013, Croson & Donohue 2006, Croson *et al.* 2013) show that the tendency to ignore time delays and underweight the supply line is robust. In many of these experiments the supply line was prominently displayed to the participants, yet they ignored it anyway. The information we use in decision-making is conditioned by our mental models. If we don't recognize the presence of a time delay or underestimate its length, we are unlikely to account for the supply line even if the information needed to do so is readily available.

Many players find these results disturbing. They argue that they took a wide range of information into account when placing orders and that their subtle and sophisticated reasoning cannot be captured by a model as simple as the anchoring and adjustment decision rule described here. After all, the decision rule for orders only considers three cues (incoming orders, inventory, and the supply line)—how could it possibly capture the way people place orders? Actually, players' behavior is highly systematic and is explained well by the simple stock management heuristic, and furthermore, the cue weights people tend to use are grossly suboptimal and lead to very poor performance, including the oscillation, amplification and phase lag seen in both the game and in real supply chains. People are often surprised how well simple decision rules can mimic their behavior.

In fact, one of the games shown in Figure 11 is a simulation, not the actual play of real people. The simulation uses the decision rule in equations (17-18), with the parameters, for all four players, set to the average estimated values over the full sample in Sterman (1989a). A small amount of random noise was added to the order rate. Can you tell which is the simulation?[7]

E. Discussion

Recognizing and Accounting for Time Delays

The beer game clearly shows it is folly to ignore the time delays in complex systems. Consider the following situation: You are involved in an automobile accident. Thankfully, no one is hurt, but your car is a total loss. Insurance settlement in hand, you visit a dealer and select a new car. You agree on a price, but the model you want is not in stock—delivery will take four weeks. You pay your deposit and leave. The next morning, noticing that your driveway is empty—*Where's my car!*—you go down to the dealer and buy another one. Ridiculous, of course. No one would be so foolish as to ignore the supply line. Yet in many real life situations people do exactly that. Consider the following examples (Table 1 shows how they map into the stock management structure):

- You cook on an electric range. To get dinner going as soon as possible, you set the burner under your pan to "high." After a while you notice the pan is getting quite hot, so

---

[7]Simulated orders were generated by eq. (17-18) with $\theta = 0.36$, $S' = 17$ units, $\alpha = 0.26$/week, and $\beta = 0.34$. The error term was iid normal with mean zero and standard deviation set to the average of the standard errors of the estimated order equation over the full sample.

you turn the heat down. But the supply line of heat in the glowing coil continues to heat the pan even after the current is cut, and your dinner is burned anyway.

- You are surfing the worldwide web. Your computer did not respond to your last click. You click again, then again. Growing impatient, you click on other buttons—any buttons—to see if you can get a response. After a few seconds, the system executes all the clicks you stacked up in the supply line, and you end up far from the page you were seeking.
- You arrive late and tired to an unfamiliar hotel. You turn on the shower, but the water is freezing. You turn up the hot water. Still cold. You turn the hot up some more. Ahhh. Just right. You step in. A second later you jump out screaming, scalded by the now too-hot water. Cursing, you realize that once again, you've ignored the time delay for the hot water to get to your shower.
- You are driving on a busy highway. The car in front of you slows slightly. You take your foot of the gas, but the distance to the car in front keeps shrinking. Your reaction time and the momentum of your car create a delay between a change in the speed of the car ahead and a change in your speed. To avoid a collision, you have to slam on the brakes. The car behind you is forced to brake even harder. You hear the screech of rubber and pray you won't be rear-ended.
- You are young, and experimenting with alcohol for the first time. Eager to show your friends you can hold your liquor, you quickly drain your glass. You feel fine. You drink another. Still feeling fine. You take another and another. As consciousness fades and you fall to the floor, you realize—too late—that you ignored the supply line of alcohol in your stomach and drank far too much.[8]

How often you have fallen victim to one of these behaviors? Few of us can say we've never burned our dinner or been scalded in the shower, never drunk too much or been forced to brake hard to avoid a collision.

Recognizing and accounting for time delays is not innate. It is behavior we must learn. When we are born, our awareness is limited to our immediate surroundings. Everything we experience is *here* and *now*. All our early experiences reinforce the belief that cause and effect are closely related in time and space: When you cry, you get fed or changed. You keep crying until mother or father appears, even when you hear your parents say, "We're coming" (i.e., despite knowledge that your request for attention is in the supply line). As all parents know, it takes years for children to learn to account for such time delays. When my son was two he might ask for a cup of juice: "Juice please, Daddy." "Coming right up," I'd say, taking a cup from the shelf. Though he could see me getting the cup and filling it up, he'd continue to say, "Juice, Daddy!" many times—ever more insistently—until the cup was actually in his hand.

[8]Tragically, young people die every year from alcohol poisoning induced by aggressive drinking (a short stock adjustment time, τ_S, and failure to account for the supply line of alcohol they've already ingested, $\beta \approx 0$).

Learning to recognize and account for time delays goes hand in hand with learning to be patient, to defer gratification, and to bear short-run sacrifice for long-term reward. These abilities do not develop automatically. They are part of a slow process of maturation. The longer the time delays and the greater the uncertainty over how long it will take to see the results of your corrective actions, the harder it is to account for the supply line.[9]

You might argue that by the time we become adults we have developed the requisite patience and sensitivity to time delays. There may be no cost to saying "juice" a dozen times, but surely when the stakes are high we would quickly learn to consider delays. You don't burn yourself in your own shower at home—you've learned where to set the hot water faucet to get the temperature you like and to wait long enough for the water to warm up. Most people learn to pay attention to the supply line of alcohol in their system and moderate their drinking. The conditions for learning in these systems are excellent. Feedback is swift, and the consequences of error are highly salient (particularly the morning after). There is no doubt in either case that it was the way you made decisions—the way you set the faucet or drank too fast—that caused the problem. These conditions are frequently not met in business, economic, environmental, and other real world systems. In the real world, cause and effect are obscure, creating ambiguity and uncertainty. The dynamics are much slower, and the time required for learning often exceeds the tenure of individual decision makers.

The French economist Albert Aftalion recognized in the early 1900s how failure to account for the time delays could cause economic cycles. Using the familiar fireplace as an analogy, his description explicitly focuses on the failure of decision makers to pay attention to the supply line of fuel:

> If one rekindles the fire in the hearth in order to warm up a room, one has to wait a while before one has the desired temperature. As the cold continues, and the thermometer continues to record it, one might be led, if one had not the lessons of experience, to throw more coal on the fire. One would continue to throw coal, even though the quantity already in the grate is such as will give off an intolerable heat, when once it is all alight. To allow oneself to be guided by the present sense of cold and the indications of the thermometer to that effect is fatally to overheat the room.[10]

---

[9]More subtly, our childhood experiences reinforce the idea that there is no cost to ignoring the supply line. Though my son may have said "Juice, Daddy" 10 times before I could fill his order, I brought him only one cup. He didn't take the supply line into account, but I did. In that situation, there is no cost to over-ordering, while patience might not work (dad might get distracted and forget to bring the juice). In many real stock management situations, there is no central authority to account for the time delays and prevent over-ordering.

[10]Quoted in Haberler, G. (1964) *Prosperity and Depression*. London: George Allen and Unwin, pp. 135–136.

While Aftalion argued that “the lessons of experience” would soon teach people not to “continue to throw coal,” he argued that business cycles in the economy arose because individual entrepreneurs focused only on current profitability and failed to account for the lags between the initiation of new investment and its realization, leading to collective overproduction.

Yet even if individuals can’t learn effectively, shouldn’t the discipline imposed by the market quickly weed out people who use suboptimal decision rules? Those who ignore the supply line or use poor decision rules should lose money and go out of business or be fired, while those who use superior decision rules, even by chance, should prosper. The selective pressures of the market should quickly lead to the evolution of optimal decision rules.

The persistent cycles in a wide range of supply chains presented at the start of this chapter suggest Aftalion was right. Learning and evolution in real markets appear to be slow, at best, despite decades of experience and the huge sums at stake, as illustrated by the persistence of business cycles and speculative bubbles such as the bubble in housing construction in the early 2000s that culminated in the financial crisis of 2008 and the so-called “Great Recession.” People tend to discount the experience of prior decades as irrelevant, arguing that the world has changed since the last crisis. Additionally, individual firms usually do not ignore the supply lines of materials on order or capital under construction. The problem is one of aggregation. The individual firm tends to view itself as small relative to the market and treats the environment as exogenous, thereby ignoring all feedbacks from prices to supply and demand. The individual firm may not know or give sufficient weight to the supply lines of all firms in the industry or the total capacity of all plants under construction. Firms tend to continue to invest and expand as long as profits are high today, even after the supply line of new capacity under construction is more than sufficient to cause a glut and destroy profitability. Each investor takes market conditions as exogenous, ignoring the reactions of others. When all investors react similarly to current profit opportunities, the result is overshoot and instability.

F. Summary

Supply chains are fundamental to a wide range of systems and many exhibit persistent instability and oscillation. Every supply chain consists of stocks and the management policies used to manage them. These management policies are designed to keep the stocks at their target levels, compensating for usage or loss and for unanticipated disturbances in the environment. Often there are important delays between the initiation of a control action and the result, creating a supply line of unfilled orders.

This chapter developed a generic model of the stock management structure and showed how it can be customized to various situations. The model explains the sources of oscillation, amplification, and phase lag observed in supply chains. These patterns of behavior are fundamental to the basic physical structure of stock management systems and supply chains. Oscillation arises from the combination of time delays in negative feedbacks and failure of the decision maker to take the time delays into account. Field and experimental studies show that people often ignore the time delays in a wide range of systems.

There is no one single cause for the failure to account for time delays and the supply line. A range of factors, from information availability to individual incentives, all contribute. But behind these apparent causes lies a deeper problem. True, the supply line is often inadequately measured, but if people understood the importance of the supply line they would invest in data collection and measurement systems to provide the needed information. True, compensation incentives often encourage people to ignore the delayed consequences of today's actions, but if investors understood the structure and dynamics of the market they could redesign compensation incentives for their agents to focus on long-term performance. Our mental models affect the design of our institutions, information systems, and incentive schemes. These, in turn, feed back to our mental models. The failure to account for the supply line reflects deeper defects in our understanding of complex systems. Ignoring time delays is one of the fundamental misperceptions of feedback that leads to poor performance in systems with high dynamic complexity. Failure to understand the role of time delays worsens the instability we face and leads to more unpleasant surprises, reinforcing the belief that the world is inherently capricious and unpredictable and strengthening the short-term focus still more.

References

Anderson, E., Fine, C., & Parker, G. (2000). Upstream volatility in the supply chain: The machine tool industry as a case study. *Production and Operations Management* 9(3): 239-261.

Armony, M. & Plambeck, E. (2005). The impact of duplicate orders on demand estimation and capacity investment. *Management Science* 51(10) 1505-1518.

Bendoly, E., Croson, R., Goncalves, P., & Schultz, K. (2010). Bodies of Knowledge for Research in Behavioral Operations. *Production and Operations Management* 19(4): 434-452.

Booth Sweeney, L. & Sterman, J. (2000). Bathtub Dynamics: Initial Results of a Systems Thinking Inventory. *System Dynamics Review* 16(4): 249-294.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*. **81** 211-241.

Boudreau, J., Hopp, W., McClain, J., Thomas, L. (2003). On the Interface Between Operations and Human Resources Management. *Manufacturing & Service Operations Management* 5(3) 179-202.

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

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

Cronin, M., Gonzalez, C., & Sterman, J. (2009). Why Don't Well-Educated Adults Understand Accumulation? A Challenge to Researchers, Educators, and Citizens. *Organizational Behavior and Human Decision Processes* 108(1): 116-130.

Croson, R. & Donohue, K. (2006). Behavioral causes of the bullwhip and the observed value of inventory information. *Management Science*. **52**(3) 323-336.

Croson, R., Donohue, K., Katok, E., & Sterman, J. (2013). Order Stability in Supply Chains: The Impact of Coordination Stock. *Production and Operations Management*

Cyert, R. & March, J. (1963/1992). A behavioral theory of the firm. Englewood Cliffs, NJ.: Prentice Hall. Second edition, Cambridge MA: Blackwell.

Diehl, E. & Sterman, J. (1995). Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes*. **62**(2) 198-215.

Forrester, J. W. (1961). Industrial Dynamics. Cambridge: MIT Press.

Gino, F.& Pisano, G. (2008). Toward a theory of Behavioral Operations. *Manufacturing and Service Operations Management* 10(4): 676–691.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, England.

Kampmann, C. & Sterman, J. (Under Review). "Do Markets Mitigate Misperceptions of Feedback?" *System Dynamics Review*.

Lee, H., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management Science* 43(4) 546-558.

Lo, A. & Repin, D. (2002). The psychophysiology of real time financial risk processing. *Journal of Cognitive Neuroscience* 14(3) 323-339.

McCarthy, J., Hocum, C., Albright, R., Rogers, J., Gallaher, E., Steensma, D. et al. (2014). Biomedical System Dynamics to Improve Anemia Control With Darbepoetin Alfa in Long-Term Hemodialysis Patients. *Mayo Clinic Proceedings* 89(1):87-94.

Morecroft, J. (1985). Rationality in the analysis of behavioral simulation models. *Management Science* 31(7) 900-916.

Paich, M. & Sterman, J. (1993). Boom, Bust, and Failures to Learn in Experimental Markets. Management Science, 39(12), 1439-1458.

Rudolph, J. &Repenning, N. (2002). Disaster Dynamics: Understanding the Role of Stress and Interruptions in Organizational Collapse. *Administrative Science Quarterly* 47, 1-30.

Senge, P. (1980). A System Dynamics Approach to Investment-Function Specification and Testing. *Socio-Economic Planning Sciences*, 14(6), 269-280.

Shah, A., Mullainathan,S., & Shafir, E. (2012). Some consequences of having too little. *Science*. 338, 682-685.

Simon, H. (1982). Models of Bounded Rationality. Cambridge: The MIT Press.

Sterman, J. (1989a). Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment. *Management Science.* 35(3), 321-339.

Sterman, J. (1989b). Misperceptions of Feedback in Dynamic Decision Making. *Organizational Behavior and Human Decision Processes.* 43(3), 301-335.

Sterman, J. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*, Irwin/McGraw-Hill.

Sterman, J. (2008). Risk Communication on Climate: Mental Models and Mass Balance. *Science* 322: 532-533.

Sterman, J. (2010). Does formal system dynamics training improve people's understanding of accumulation? *System Dynamics Review* 26(4): 316-334.

Sterman, J., & Booth Sweeney, L. (2007). Understanding Public Complacency About Climate Change: Adults' Mental Models of Climate Change Violate Conservation of Matter. *Climatic Change* 80(3-4): 213-238.

Sterman, J., & Dogan, G. (2014). "I'm not hoarding, I'm just stocking up before the hoarders get here" Behavioral causes of phantom ordering in supply chains. Working paper: MIT Sloan School of Management, Cambridge, MA.

Tversky, K., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.