The Bullwhip Effect and Order Smoothing in a Laboratory Beergame

David E. Cantor
Coggin College of Business, University of North Florida, Jacksonville, FL 32224
david.cantor@unf.edu

Elena Katok
Smeal College of Business, Penn State University, University Park, PA 16802
ekatok@psu.edu

Production and order smoothing is often a rational cost-minimizing behavior and it has been documented in practice. Changing orders may be costly, but these costs can be difficult to measure using secondary data. Using the controlled setting of the laboratory, we systematically investigate supply chain features that lead to the bullwhip effect and to order smoothing. We find that shorter lead-times not only reduce the bullwhip effect, but also induce order smoothing by retailers. When customer demand has a predictable seasonal component, our participants understand how to smooth orders, and order smoothing is even more pronounced. This behavior is even more prevalent when changing order levels is costly.

1. Introduction and Motivation

Managing a supply chain is a dynamic decision task shown to be prone to systematic errors, collectively referred to as the *bullwhip effect*. The bullwhip effect, which is the tendency for orders to become more variable as one moves away from the final customer and closer to the source of production, tends to lead to dysfunctional outcomes. For example, resources spent drilling for oil and gas fluctuate three times more than actual petroleum production, demand for machine tools is at least twice as variable as automobile sales (the auto industry being the main consumer of machine tools), and production of semiconductors is much more variable than industrial production as a whole (reported in Sterman 2000, pp. 666-667).

Certain structural supply chain factors, (such as those discussed by Lee et al. 1997) amplify the bullwhip effect, while other structural factors might actually mitigate it. A number of behavioral factors have also been shown to contribute to the bullwhip effect in laboratory settings (see Sterman 1989, Croson and Donohue 2003, 2006, Croson. et al 2007, Wu and Katok 2006). In general, behavioral issues aside, firms should smooth their orders relative to their sales when the cost of varying orders is higher than the cost of holding inventory. Cachon et al. (2007) make this basic argument and use secondary data to measure some of the factors that contribute to the bullwhip effect and order smoothing. They find, among other things, that order smoothing is especially prevalent in industries that exhibit seasonality and that retailers and manufacturers tend to smooth orders, while wholesalers, on the contrary, tend to amplify them.

Our study helps to bridge the gap between laboratory research that finds indisputable evidence of the bullwhip effect, and field evidence that indicates that the bullwhip effect phenomenon is less pervasive. We use the controlled setting of the laboratory to systematically manipulate three supply chain features shown to be important in the field—the length of the lead times, demand seasonality, and cost structure—in order to isolate the effect they have on order smoothing behavior.

In studying supply chain management problems, we view laboratory and field methods as compliments (and not substitutes). The two methods can be most effective when used in tandem (see for example Croson and Donohue (2002)), and this is how we use them in our study. Our work is partly motivated by Cachon et al. (2007) who identify several structural factors that cause firms to smooth production. While some of those factors may be difficult to measure and isolate in the field, especially for individual firms, we designed a set of laboratory experiments that manipulate these factors with a great deal of precision and control. At the same time, the fact that the supply chain attributes we manipulate are ones that have been shown to matter in the field increases the external validity of our laboratory experiments.

We use a simplified version of the *Beer Distribution Game* (subsequently simply the beergame) in our experiments. The relative simplicity of the beergame has upsides and downsides. Undoubtedly, real supply chain structures are different than the beergame in many ways (they are not simple serial chains, decision-making is made by groups of managers, these decision-makers have access to ERP systems and other decision support tools, the decision-makers' incentives are unknown, etc.) But there is an important similarity. Since managing both, a real supply chain and a simulated beergame supply chain are dynamic decision tasks, decision-makers in both settings must, to one degree or another, rely on intuitive judgments.

In the field, in spite of sophisticated decision support tools, the notion of optimal ordering policy is often not well defined due to the complexity of the environment, so decision-makers rely on experience and intuition. In the laboratory beergame, although optimal policies can be derived theoretically (although usually assumptions required are quite strong[1]) they are generally not transparent. For example, in the beergame with a known and stable customer demand distribution, a policy that would minimize the total supply chain cost (at least for a sufficiently long time horizon) is for the retailer to carry sufficient inventory to balance the marginal cost of overages and underages, and to place orders equal to average demand. The rest of the players should carry no safety stock and also place orders equal to average demand. This policy would lead to order smoothing by the retailer, and generally nothing resembling this behavior is observed in laboratory beergames. An exception is Wu and Katok (2006) who report that in the treatment with pre-game communication by participants experienced with operating a centralized system, a few (by no means all) of the teams follow a policy close to this optimal. Lacking a transparent optimal policy, laboratory beergame participant rely on intuitive judgment, just as their real-world counterparts.

Since managers of supply chains in practice and managers of laboratory beergame supply chains are faced with a dynamic task that requires them to rely on their intuition and judgment, a laboratory experiment that uses a simulated beergame supply chain can inform us about supply chain management in the real world. In this study we look at the effect of lead times, demand seasonality, and the cost of changing orders, the latter two being factors that Cachon et al. (2007) identified as production smoothing drivers in industry. We find strong evidence that retailers smooth orders in response to seasonal demand and even more so when costs to changing orders are present. Thus we find that the drivers of production smoothing in the field are also the drivers of production smoothing in the laboratory. This finding contributes to the external validity of laboratory studies of the bullwhip effect (in general, not just our own) as well as to the internal validity of field studies of the bullwhip effect—laboratory and field research work in tandem to further knowledge about a difficult and complex problem.

[1] Assumptions that are particularly strong have to do with a player's belief about the actions of the other players. Croson et al (2007) for example find the bullwhip effect in settings with constant and known customer demand of 4 units per week, that is common knowledge, and the behavior persists even in a treatment in which participants are told that the cost-minimizing policy is to always order the same amount as the order received. The explanation is that even though players understand the optimal solution themselves, they do not believe that all other team members understand it also. Consequently, many players seek additional safety stock, thus setting off a cycle of backlogs.

In the following section we summarize some of the relevant literature. Section 3 describes the experimental design and implementation. We present our experimental hypotheses and results in Section 4 and summarize and discuss our findings, as well as offer directions for future research in Section 5.

2. Literature Review

There is a long history of empirical research into production smoothing and the bullwhip effect in the economics and operations management literature (see for example Cachon et al. 2007, Blanchard 1983, Blinder 1986, and Miron and Zeldes 1988 as well as Lee et al. 1997 and Chatfield et al. 2004). Using secondary data, Cachon et al. (2007) measure and discuss various factors that contribute to the bullwhip effect and production smoothing. They argue that a major reason firms should smooth production has to do with a firm's cost structure. Firms face a cost to drastically change production levels, and when those costs exceed inventory holding cost, inventory should be used to smooth production. Cachon et al. (2007) find that production smoothing is especially prevalent in industries which exhibit demand volatility (e.g., seasonality). Building on Cachon et al. (2007)'s framework, we discuss three factors that may contribute to the bullwhip effect and production smoothing.

2.1 Demand Volatility

Cachon et al. (2007) identify demand volatility as a contributing factor to production smoothing. Indeed, demand volatility has been explored in beergame experiments. In a three-echelon supply chain, Steckel et al. (2004), test the effect of the customer demand distribution and the availability of the point-of-sale (POS) information. They find that the bullwhip effect is most pronounced with the step-up demand, but it also persists with the S-shape demand function with and without errors. Interestingly, the POS information only improves performance in the step-up demand condition; it actually increases costs in the other conditions. Croson and Donohue (2003) test the effect of POS information in a 4-echelon game with a known customer demand distribution (uniform 0 to 8) and find that this information does reduce order oscillation, especially at the distributor and factory levels. Lastly, Croson et al. (2007) eliminate customer demand uncertainty entirely by making the demand constant at 4 units per period, and making this information common knowledge to all players in the beergame. Interestingly, this manipulation does not eliminate the bullwhip effect, but on the contrary increases it, especially in the treatment with zero initial on-hand inventory. This result is, on the face of it, surprising, because the system starts out in equilibrium—constant and known demand means that no safety stock is needed, so 0 on-hand inventory is the optimal amount of inventory. Croson et al. (2007) attribute the anomalous result to a lack of trust by some of the players in their team-members' abilities to follow the optimal policy.

2.2 Cost Structure

Cachon et al. (2007) also find that cost shocks contribute to production smoothing. Surprisingly, one disconnect between field and laboratory research is in that the supply chain's cost structure has never been varied in the laboratory. To be sure, it is difficult to study cost structure in the field also. No doubt firms have capacity constraints and these constraints make changing production levels costly. While the exact structure of these costs in the field may be difficult to ascertain, their presence should make production smoothing more attractive. Changing order quantities was free in all previous laboratory studies, and since these costs are an important reason to smooth production, not having these costs makes the bullwhip effect more likely.

2.3 Lead Times

Lead time length is a contributing factor to the bullwhip effect and has been explored in previous laboratory beergame research. Longer lead times increase the amount of inventory in the supply line relative to the amount of inventory on-hand. Since players persistently underweight the supply line (Sterman 1989), a longer supply line should increase the bullwhip effect. Steckel et al. (2004) confirmed this conjecture. They found that inventory and backorder costs decrease when the lead-times are shorter. The Steckel et al. (2004) study does not report information on order amounts, so it is not clear whether lower costs in their study can be attributed to a reduction in the bullwhip effect or to something else.

3. Design of the Experiment

3.1 Methods

We follow the basic protocol of the "beer distribution game" used in previous experimental studies (see Sterman 1989, Croson and Donohue 2002), but we simplified the setting. Figure 1 shows the diagram of the serial 2-echelon supply chain in our experiment, consisting of a *Retailer* and a *Factory* with exogenous customer demand. Assigned one of these roles, each participant manages her own inventory by placing orders to the upstream supplier for replenishment so as to satisfy demands downstream over multiple periods.

Each period begins with the arrival of shipments, which increases one's on-hand inventory. Next orders placed by the downstream customer are received, which are either filled when inventory is available or become backlogged. Each participant then makes an ordering decision and carries any remaining inventory/backlog over to the next period. The decision task is complicated by the existence of lead-times/delays in the supply chain. All participants saw, on their computer screen, their current inventory, previous order amount placed, customer demand, current amount on backorder, incoming shipments and orders, and total inventory and backorder cost. History on one's own inventory level, orders received and placed is also displayed to participants.

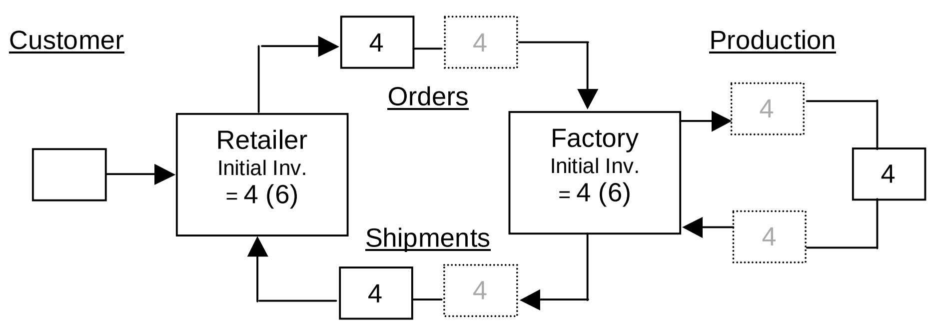


Figure 1. The structure of the 2-echelon version of the "beer distribution game" in our study. The solid boxes represent lead times in the short lead-time treatments. The dotted boxes were added in the long lead-time treatment. Each delay position was always initialized with four units. Initial inventory was 4 units in the short lead-time treatments and 6 units in the long lead-times treatments. The customer demand was U(0,8) and an integer in the random demand treatments and alternating at 0 and 8 in the seasonal demand treatments.

As in most previous work, we initialized orders and shipment in process to be 4 units. Starting inventory at each echelon of the supply chain was 4 units in short lead-time treatments and 6 units in long lead-times treatments. Each team was given an initial endowment, and all participants were told they would incur inventory cost of 0.5 token per unit per week and backorder cost of 1 token per unit per week. Additionally, in treatments with costly order changes, the order change cost was 0.5 tokens per unit of change in order quantity (for example, and increase from 4 to 6 units or a decrease from 4 to 2 units cost 1 token). Final team earnings in tokens were the difference between the initial team endowment and the cumulative holding, backorder, and where applicable order change, costs of all team members. At the end of the session, each team's total earnings were converted to US dollars at a pre-determined exchange rate and divided equally between the two team members. Each participant was provided an additional \$5.00 show-up fee. All experimental sessions lasted 50 periods.

Our design manipulates three factors that we summarize in Table 1. Four treatments include *Short* lead times (2 periods between the retailer and the factory and 1 period production delay for the factory). To check the effect of lead times, a fifth treatment includes *Long* lead times (4 periods between the retailer and the factory and 3 period production delay for the factory). Within the *Short* lead-time condition we have two demand distributions: in the *Random* demand condition customer demand follows the Uniform distribution from 0 to 8 (rounded to the nearest integer) and in the *Seasonal* demand condition demand is 0 in odd-numbered periods and 8 in even-numbered periods. Also within the *Short* lead time condition, we conducted two additional treatments with *Costly* order changes, one with each demand distribution. In total, 126 participants were included in our experiment, each randomly assigned to one of the five treatments summarized in Table 1, as well as to one of the two roles (Retailer or Factory).

Table 1. Description of the factors and sample sizes in the treatments in our experiment.

Lead Times	Demand	Order change cost	Sample Size (N)
Long	Random	None	10
Short	Random	None	14
Short	Seasonal	None	11
Short	Random	Costly	15
Short	Seasonal	Costly	13

Lead Times:	Long = 4 (retailer) 3 (factory)
	Short = 2 (retailer) 1 (factory)
Demand:	Random = uniform integer from 0 to 8;
	Seasonal = alternating 0 or 8
Order Change:	Costly = 0.5 tokens per unit of change;
	None = 0
Sample Size:	Number of pairs in the treatment

We conducted all sessions at the Laboratory for Economic Management and Auctions (LEMA) at Penn State, Smeal College of Business, in Fall 2006. Participants, mostly undergraduates, were recruited using the on-line recruitment system, with cash the only incentive offered. Approximately 50% of our subject pool consists of business majors, 30% are engineering majors, and the rest are in natural and social sciences and humanities. Approximately 60% are male and

40% are female. We did not find any statistically significant difference in earnings based on these demographics.

The experiment proceeds as follows. Participants arrive at LEMA at a pre-specified time and are seated at computer terminals. They take approximately 10 minutes to study the written instructions (see Appendix). We then read the instructions to them aloud (to ensure common knowledge of the rules of the game) and answer any questions. Participants complete a brief quiz on the rules of the game, and we go over the answers. After all the quizzes are completed and questions (if any) are answered, participants complete 50 periods of the game. At the conclusion of the sessions, the participants fill out a questionnaire and receive their final payments in private. Average earnings of our experiments were $20, and all sessions lasted approximately 60 minutes. All sessions were conducted using the Z-Tree experimental economics software program (Fischbacher 2007).

4. Research Hypotheses and Results

In this section we present our three research hypotheses and the corresponding results. The unit of our analysis is the standard deviation of orders: for each player we compute the standard deviation of orders for the 50 periods in the session. If amplification (bullwhip) exists, then the standard deviation of orders of the i^{th} echelon in the supply chain, σ_i, exceeds that of its immediate customer σ_{i-1}, (for all $i \in$ {R, F}). If order or production smoothing exists, then the standard deviation of orders of the i^{th} echelon in the supply chain, σ_i, is less than its immediate customer σ_{i-1}, (for all $i \in$ {R, F}). We use the matched-pair Wilcoxon test to measure the amount of amplification within a treatment, and we use the two-sample Wilcoxon test to make comparisons between treatments.

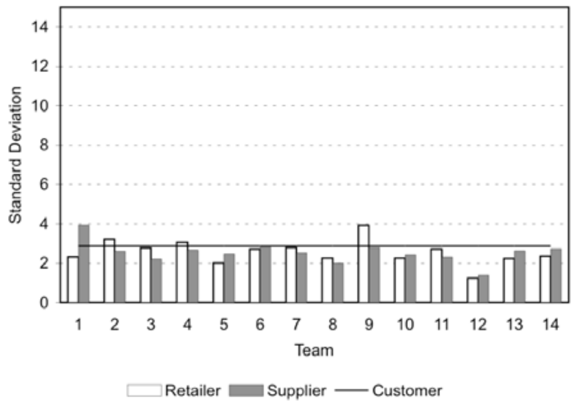
4.1 Lead times

We start by looking at the effect of lead times on performance to determine under what conditions we will observe order smoothing in a 2-player game. In this first experiment we compare two treatments: the *Long* and *Short* lead-time treatments with *Random* customer demand and no costly order change (see Table 1). The initial inventory was 6 units in the *Long* lead-time treatment and 4 units in the *Short* lead-time treatment (the initial on-hand inventory levels had to be slightly different in the two treatments to insure approximately the same initial probability of a stock-out).

Shorter lead-time implies (all else constant) less inventory in the supply line. Managing the supply line is a difficult task for human subjects in laboratory settings (Sterman 1989). As a result human subjects consistently underweight the supply line and this has been identified as a behavioral cause of the bullwhip effect (Sterman 1989). Accordingly, shorter lead-times should lead to production smoothing. Indeed, Steckel et al. (2004) report that shorter lead times lead to lower costs in a three echelon supply chain, but do not report the effect they have on standard deviation of orders (bullwhip). In our setting, we will directly look at the effects lead-times have on production smoothing.

H1: Shorter lead times will induce more oorder smoothing relative to a setting with longer lead times.

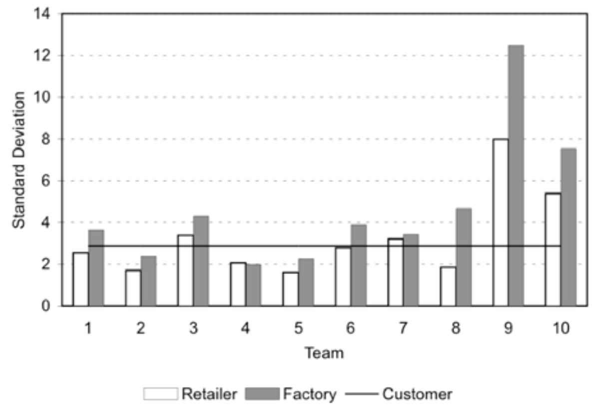
Figure 2 summarizes standard deviations of orders in the short and long lead time treatments. Table 2 summarizes the corresponding hypothesis testing. All p-values indicating statistically-significant differences are in bold. We find that retailers do not amplify customer demand variability in either the short or long lead time treatment, and neither does the factory when lead times are short (Figure 2). When lead-times are long nine out of the 10 factories place orders that are more variable than the retailers' orders (Figure 2). As presented in Table 2, in the long lead time treatment, the median standard deviation of orders for factories is larger than the median standard deviation of orders for retailers. In the short lead-time treatment, 11 out of the 14 retailers place orders that are *less* variable than customer demand (Figure 2), and the median standard deviation of orders for retailers is below that of customer (p-value = 0.039). So we find some evidence of the bullwhip effect in the long lead-time treatment between the retailer and the factory, and we also find support for H1 by finding evidence of order smoothing by the retailer in the short lead-time treatment.



$\mathrm{H_o}: s_C^{Short} \le s_R^{Short}; \mathrm{H_a}: s_C^{Short} > s_R^{Short}$ **0.039**

$\mathrm{H_o}: s_R^{Short} \ge s_F^{Short}; \mathrm{H_a}: s_R^{Short} < s_F^{Short}$ 0.313

(a) Short lead-times.



$\mathrm{H_o}: s_C^{Long} \ge s_R^{Long}; \mathrm{H_a}: s_C^{Long} < s_R^{Long}$ 0.539

$\mathrm{H_o}: s_R^{Long} \ge s_F^{Long}; \mathrm{H_a}: s_R^{Long} < s_F^{Long}$ **0.002**

(b) Long lead-times.

Figure 2. The effect of lead times

Table 2. Hypothesis testing. The third row contains median standard deviations of orders in each position. The left-most column states the hypothesis test. The right-most column contains p-values (one-sided) from the Wilcoxon test of each hypothesis.

	Short Lead-times		Long Lead-times		
	Retailer	Factory	Retailer	Factory	Wilcoxon p-value
Hypotheses \ Median:	2.53	2.57	2.65	3.64	
$\mathrm{H_o}: s_R^{Short} = s_R^{Long}; \mathrm{H_a}: s_R^{Short} < s_R^{Long}$	X		X		0.352
$\mathrm{H_o}: s_F^{Short} = s_F^{Long}; \mathrm{H_a}: s_F^{Short} < s_F^{Long}$		X		X	**0.029**

The length of the lead times also has an effect of order variability. As shown in Table 2, the median standard deviation of retailers' orders in the long lead-time treatment is only slightly higher than the standard deviation of the retailers' orders in the short lead-time treatment. However, the factory orders are significantly more variable with long lead times than with short lead times, providing additional support for H1.

4.2 Demand Distributions

In our second experiment we look at the effect of seasonality in customer demand to determine whether seasonal demand increases the likelihood of order smoothing. We use the Cachon et al. (2007) definition of the predictable seasonality ratio—it is the proportion of demand variability that can be attributed to a regular seasonal component:

$$Seasonality\ Ratio = \frac{V(Demand) - V(Seasonally\ Adjusted\ Demand)}{V(Demand)}.$$

In treatments with *Seasonal* demand, customer demand was 0 in all of the odd-numbered periods and 8 in all of the even-numbered periods. Since the variance of the seasonally adjusted demand is 0, the seasonality ratio in these treatments is 1. The seasonality ratio in the treatments with *Random* demand is 0. The variance of customer demand in the treatments with *Seasonal* demand is 16.64 (standard deviation = 4.08), and in treatments with *Random* demand the variance is 8.18 (standard deviation = 2.86).

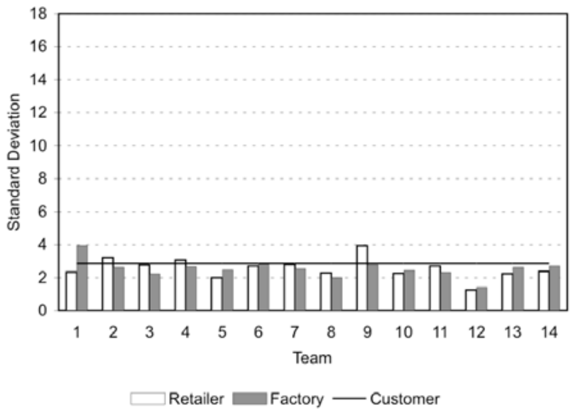
Predictable seasonality ratio in the customer demand has two effects. First, it more than doubles the variance of customer demand, and unless players significantly smooth production, this increased variability should increase the variability of orders in general. Thus we have the following hypothesis:

H2A: A seasonal component in customer demand will increase order variability relative to a setting without such seasonal component.

But if seasonality is the *only* variability in customer demand (as it is in our experiment), customer demand becomes predictable, and the benefits of production smoothing become more transparent in the supply chain. If players smooth production sufficiently, order variability in treatments with *Seasonal* demand may become even lower than in treatments with *Random* demand.

H2B: A seasonal component in customer demand will induce more order smoothing relative to a setting without such seasonal component.

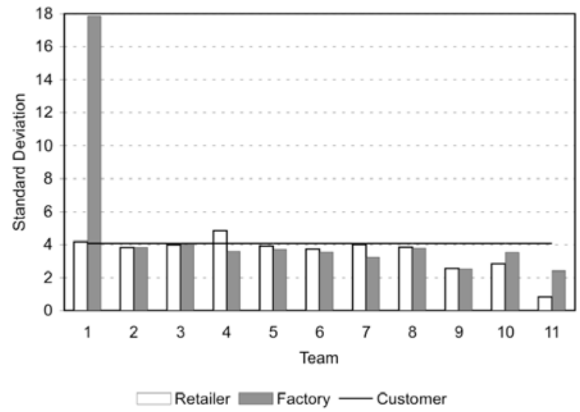
Figure 3 summarizes standard deviations of orders in the four *Short* lead-time treatments (the *Random* demand with no costly order change treatment in Figure 3a is the *Short* lead-time treatment from the previous section) and results of hypotheses test for the bullwhip effect and order smoothing. Table 3 summarizes the hypothesis tests of the comparisons across treatments (as before, p-values indicating statistically-significant differences are in bold).



$\mathrm{H_o}: s_C^{Random} \le s_R^{Random}$; $\mathrm{H_a}: s_C^{Random} > s_R^{Random}$ **0.039**

$\mathrm{H_o}: s_R^{Random} \le s_F^{Random}$; $\mathrm{H_a}: s_R^{Random} > s_F^{Random}$ 0.313

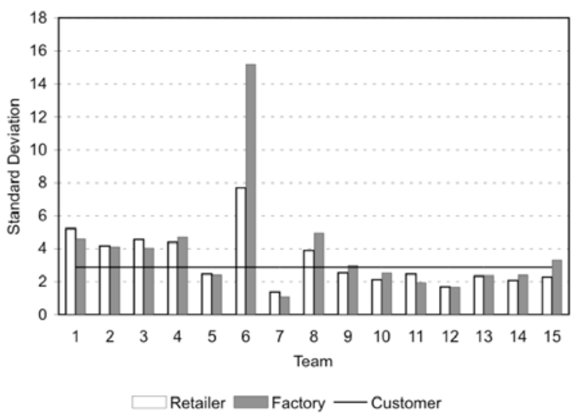
(a) *Random* Demand; no costly order change.



$\mathrm{H_o}: s_C^{Seas'nl} \le s_R^{Seas'nl}$; $\mathrm{H_a}: s_C^{Seas'nl} > s_R^{Seas'nl}$ **0.041**

$\mathrm{H_o}: s_R^{Seas'nl} \le s_F^{Seas'nl}$; $\mathrm{H_a}: s_R^{Seas'nl} > s_F^{Seas'nl}$ 0499

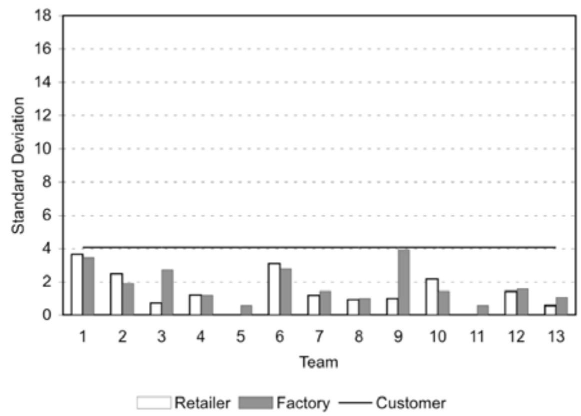
(b) *Seasonal* Demand; no costly order change.



$\mathrm{H_o}: s_C^{Random} \le s_R^{Random}$; $\mathrm{H_a}: s_C^{Random} > s_R^{Random}$ 0.263

$\mathrm{H_o}: s_R^{Random} \le s_F^{Random}$; $\mathrm{H_a}: s_R^{Random} > s_F^{Random}$ 0.804

(c) *Random* Demand; costly order change.



$\mathrm{H_o}: s_C^{Seas'nl} \le s_R^{Seas'nl}$; $\mathrm{H_a}: s_C^{Seas'nl} > s_R^{Seas'nl}$ **0.000**

$\mathrm{H_o}: s_R^{Seas'nl} \le s_F^{Seas'nl}$; $\mathrm{H_a}: s_R^{Seas'nl} > s_F^{Seas'nl}$ 0.847

(d) *Seasonal* Demand; costly order change.

Figure 3. Order variability in *Short* lead-times treatments.

Retailers smooth production in both treatments with *Seasonal* demand (regardless of the presence of the costly order change; Figures 3b and 3d), which provides support for H2B. Recall from the previous section that retailers also smooth orders in the *Random* demand treatment without the costly order change (Figure 3a) (support for H1). There is, however, no evidence of any order smoothing by retailers or the bullwhip effect in the *Random* demand treatment with costly order change (Figure 3c). This latter finding is surprising, but it continues to hold even when team 6, and obvious outlier, is excluded from the analysis. There is no statistically-significant evidence of order smoothing or the bullwhip effect by the factories in any of the *Short* lead-times treatments.

Table 3. Hypotheses testing for the effect of demand distribution. The third row contains median standard deviations of orders in each position. The left-most column states the hypothesis test. The right-most column contains p-values (two-sided) from the Wilcoxon test of each hypothesis.

	Random Demand		Seasonal Demand		
	Retailer	Factory	Retailer	Factory	Wilcoxon p-value
No Costly Order Change					
Median: / Hypotheses	2.53	2.57	3.83	3.59	
$\mathrm{H_o}: s_R^{Random} = s_R^{Seas'nl}; \mathrm{H_a}: s_R^{Random} \neq s_R^{Seas'nl}$	X		X		**0.008**
$\mathrm{H_o}: s_F^{Random} = s_F^{Seas'nl}; \mathrm{H_a}: s_F^{Random} \neq s_F^{Seas'nl}$		X		X	**0.005**
Costly Order Change					
Median: / Hypotheses	3.27	3.88	1.41	1.81	
$\mathrm{H_o}: s_R^{Random} = s_R^{Seas'nl}; \mathrm{H_a}: s_R^{Random} \neq s_R^{Seas'nl}$	X		X		**0.002**
$\mathrm{H_o}: s_F^{Random} = s_F^{Seas'nl}; \mathrm{H_a}: s_F^{Random} \neq s_F^{Seas'nl}$		X		X	**0.007**

Turning to Table 3, in treatments without costly order changes, we find support for H2A: orders are more variable for both, retailers and factories, with *Seasonal* demand than with *Random* demand. The reason for this increase is that *Seasonal* customer demand is substantially more variable than *Random*, so even though retailers smooth orders, they do not smooth them sufficiently to make up for the higher variability of customers' orders.

In treatments with costly order changes the orders are less variable for both, retailers and factories, with *Seasonal* demand than with *Random* demand. The reason for the decrease in variability can be clearly seen in Figure 3d—retailers smooth orders enough to make up for the higher variability of customer orders. This provides additional evidence in support of H2B—seasonal demand induces order smoothing.

4.3 The Cost of Changing Order Amount

Our last set of hypotheses relates to the effect of costly order changes on order and production smoothing.

H3A: When order changes are costly, order and production smoothing should occur relative to a setting in which order changes are free.
H3B: When order changes are costly, order variability should decrease relative to a setting in which order changes are free.

Table 4 summarizes hypotheses tests that compare order variability in treatments with and without costly order changes.

Table 4. Hypothesis testing for the effect of costly order change. The third row contains median standard deviations of orders in each position. The left-most column states the hypothesis test. The right-most column contains p-values (two-sided) from the Wilcoxon test of each hypothesis.

	No Costly Order Change		Costly Order Change		
	Retailer	Factory	Retailer	Factory	Wilcoxon p-value
Random Demand					
Median: / Hypotheses	2.53	2.57	3.27	3.88	
$H_o: s_R^{None} = s_R^{Costly}; H_a: s_R^{None} \neq s_R^{Costly}$	X		X		0.498
$H_o: s_F^{None} = s_F^{Costly}; H_a: s_F^{None} \neq s_F^{Costly}$		X		X	0.285
Seasonal Demand					
Median: / Hypotheses	3.83	3.59	1.41	1.81	
$H_o: s_R^{None} = s_R^{Costly}; H_a: s_R^{None} \neq s_R^{Costly}$	X		X		**0.001**
$H_o: s_F^{None} = s_F^{Costly}; H_a: s_F^{None} \neq s_F^{Costly}$		X		X	**0.001**

When demand is *Random*, the data does not support H3A because there is no evidence of order smoothing in the treatment with *Random* demand and costly order changes (Figure 3c). Recall that there is evidence of order smoothing in the treatment with *Random* demand and no costly order changes, so to the extent that there is evidence, it runs counter to the hypothesis when demand is *Random.*

When demand is *Seasonal*, the data is consistent with H3A because there is strong evidence of order smoothing by the retailers in the treatment with *Seasonal* demand and costly order changes (Figure 3d). In the treatment with *Seasonal* demand and no costly order changes, however, there is also some order smoothing by the retailer (Figure 3b), so the order smoothing in treatments with *Seasonal* demand cannot be attributed exclusively to costly order changes.

When demand is *Random*, the data is not consistent with H3B (Table 4) because we cannot reject the null hypothesis that order variability is the same in the treatment with costly order changes as in the one without. When the demand is *Seasonal*, however, there is strong evidence that orders are less variable with costly order changes, consistent with H3B. Our results indicate that there is an interaction effect between *Seasonal* demand and costly order changes that induce order smoothing. Order smoothing that we observe cannot be attributed to costly order changes alone because we do not observe order smoothing in the treatment with *Random* demand.

5. Discussion and Conclusions

We investigate the bullwhip effect and order smoothing in the context of a simplified version of the beer distribution game. Our beergame version includes two echelons (the *Retailer* and the *Factory*) instead of the standard four-echelon version, but in other ways we kept our implementation close to the previous literature (see Figure 1 for the graphical representation of the game in our study).

The first issue we examine is the effect of lead-times. We observe the bullwhip effect in the treatment with lead-times that are the same as in the standard beergame implementation (the *Long* lead-times treatment). The bullwhip effect disappears in the *Short* lead-time treatment, however, (moreover, retailers smooth orders) and this is consistent with the Sterman (1989) notion that underweighting the supply line is the main behavioral cause of the bullwhip effect. The size of the supply line is proportional to lead times— longer lead-times imply a longer supply line. Since participants do not manage the supply line correctly due to supply line underweighting, it follows that shorter supply-lines should induce less bullwhip effect, as they do in our study.

Having connected our simplified two-echelon beergame to the literature by replicating the bullwhip effect in this simplified setting with *Long* lead-times, we proceed to explore order smoothing in the context of the two-echelon beergame with *Short* lead-times. We intentionally designed this setting with the express purpose of creating an environment in which we are likely to observe order smoothing by eliminating some of the complexity present in the standard beergame. Thus, the goal of our research is to better understand the factors that induce order smoothing, which is different from the goal of most previous laboratory beergame research, which was to better understand factors that induce the bullwhip effect, ways for mitigating it, and propose new behavioral heuristics that explain behavior (Dogan and Sterman 2006). Understanding factors that lead to order smoothing is the contribution of our paper, and ours is the first study to investigate these factors in the controlled and incentive-compatible setting of a laboratory.

The two factors that we investigate are ones that Cachon et al. (2007) found to be related to production smoothing in the field: one is the effect of predictable seasonal demand, and the second is the effect of costly order changes. It turned out that our main finding has to do with the *interaction effect* of these two factors. Regular seasonal demand makes it easy for laboratory participants to understand *how* to smooth production. Consequently, retailers smooth production in both treatments with *Seasonal* demand, but more so when order changes are costly. Why? While order smoothing by the retailers is beneficial to the supply chain as a whole whether or not there is an explicit cost to changing orders, Retailers have to assume a disproportional amount of holding costs in order to smooth orders. Some retailers may resist having to carry so much inventory because they think myopically, and do not recognize that in order to minimize the total supply chain cost, their local cost has to increase. But costly order changes make the benefits of order smoothing that much clearer. Now by smoothing orders retailers do not just benefit the supply chain, but they also decrease their own costs by avoiding the order change charge.

Why do costly order changes not lead to order smoothing by retailers when the demand is *Random*? One possibility is that it is not clear to some retailers *how* to smooth random demand. Participants in these treatments are told that the demand is an integer from 0 to 8, each equally likely, but it may well be that human subjects do not understand, left to their own devises, that the mean of this distribution is 4, and that ordering 4 and holding a sufficient amount of inventory is an ordering policy that is likely to lead to lowered costs. There is evidence that even trained scientists fall victim to the "law of small numbers" -- making conclusions based on inappropriately small samples (Tversky and Kahneman 1971). Participants who attempt to learn the customer demand distribution through experience may be susceptible to one of the "gambler's fallacy" biases[2] (Kahneman and Tversky 1972) that has been shown to cause orders to be correlated with previous demand draws (Bolton and Katok 2007). A visual examination of Figure 3c, and comparing it to the rest of Figure 3, reveals that 6 out of 15 retailers place orders that are significantly more variable than demand, while the other 9 retailers place orders that are signifi-

[2] These are versions of incorrectly believing that independent draws are positively correlated (Positive; ex., 'hot hand' fallacy in basketball) or negatively correlated (Negative; ex., believing a number on the roulette wheel is 'due').

cantly less variable. For those six retailers that do amplify order variability, the average increase in the standard deviation of orders relative to the standard deviation of customer orders is 2.11 (in other words, those six retailers double the standard deviation of orders). In the other three *Short* lead-time treatments, the comparable average increase is 0.49. The above informal analysis is suggestive of the fact that some of the retailers simply do not understand *how* to smooth *Random* demand. They want to do something to avoid costly order changes, but their actions backfire.

Our paper is the first to look at order smoothing in a beergame laboratory experiment. Previous beergame studies were primarily concerned with identifying factors that contribute to the bullwhip effect (Sterman 1989, Steckel et al. 2004, Croson at al. 2007) ways to mitigate the bullwhip effect through additional information (Croson and Donohue 2003, 2006) enhancing coordination through education or additional inventory (Croson et al. 2007), providing players with experience and opportunities to communicate (Wu and Katok 2006) or developing behavioral heuristics that explain ordering behavior (Dogan and Sterman 2006). Our paper also helps to bridge the gap between laboratory experiments and field studies, because we examine some of the factors Cachon et al. (2007) identified as contributing to order smoothing in the field, and find how these factors also contribute to order smoothing in the laboratory. Thus our laboratory study gains external validity because the factors we manipulate in the laboratory are the same factors that Cachon et al. (2007) identified as contributing to production smoothing in the field.

One issue that is sometimes mentioned about beergame experiments in general is that these results are due to the poor understanding of the game by the participants. In other words, if only participants knew the optimal solution, the bullwhip effect would disappear. We would like to point at mounting evidence that the solution is unlikely to be this simple. The setting in Croson et al. (2007) has an optimal solution that is transparent, but participants do not follow it even in the treatment in which this solution is explained to them. Even the most extreme behavior in that study is consistent with a behavioral decision-making heuristic developed by Dogan and Sterman (2006). Participants in Wu and Katok (2006) discover on their own a near-optimal ordering policy through experience, in a setting that is significantly more complicated than that in the Croson et al. (2007) study. In fact, in Wu and Katok (2006), when pre-game communication is allowed, some teams are able to come close to implementing this policy. In our study, participants figure out not only how to avoid the bullwhip effect, but also, in several treatments, how to smooth orders— a task that may be considered more complex, since the bullwhip effect can be eliminated through simply matching orders placed with orders received, but successful order smoothing requires more sophisticated actions. A lack of understanding of how to participate in beergame experiments cannot account for results in all of the above studies because there is evidence that laboratory participants are often quite sophisticated and rational. Laboratory experiments that report evidence of this rationality seemingly breaking down should not be dismissed, but on the contrary, they should be used to point out aspects of human behavior or decision-making processes that have not previously been correctly understood. This, in turn, will lead to better models for researchers and to better decision support tools for practitioners.

We conclude by mentioning that our study provides significant contributions to the literature and that it establishes a linkage between field and laboratory research but it does have several limitations that can be addressed through future research. One is to look at the effect of supply chain length. Cachon et al. (2007) find, looking at three levels of field data (retailers, wholesalers, manufacturers), that retailers and manufacturers are more likely to smooth orders, while wholesales are more likely to amplify the variability. To check whether this regularity holds in the laboratory requires at least a 3-echelon supply chain. Another direction for future research is to look at the effect of price variability. Cachon et al. (2007) find evidence that price variability contributes to amplification. They use price variability as a proxy for promotional activity and cost shocks because direct data on the cost of production factors and on promotional activity is not readily available. The laboratory provides an opportunity to test a hypotheses about the effect of cost shock and promotional activity directly. Yet a third direction for future research is to look at the effect of demand shocks. Cachon et al. (2007) report that the amount of autocorrelation in demand does not have a significant relationship with the bullwhip effect. They also note that most industries in their sample have negative demand autocorrelation, so field data does not provide an opportunity for a clean test. A demand stream with a positive autocorrelation, however, can be easily implemented in the laboratory.

Acknowledgments

This research was funded by the Center for Supply Chain Research (CSCR), Smeal College of Business, Penn State University. The authors also thank the support from Laboratory of Economics Management and Auctions (LEMA) at PS and Axel Ockenfels and the Deutsche Forschungsgemeinschaft for financial support through the Leibniz-Program. The first author was a visiting assistant professor at Penn State when this research was conducted and appreciates PSU's support of this project.

References

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

Blinder, A. 1986. Can the production smoothing model of inventory behavior be saved? *Quarterly Journal of Economics* **101** 431-53.

Bolton, G. and Katok, E. 2007, Learning-by-Doing in the Newsvendor Problem: A Laboratory Investigation of the Role of Experience and Feedback, Penn State Working Paper.

Cachon, G., T. Randall, G. Schmidt. 2007. In search of the bullwhip effect. *Manufacturing and Service Operations Management*, in press.

Chatfield, D.C., J.G. Kim, T.P. Harrison, and J.C. Hayya. 2004. The bullwhip effect-impact of stochastic lead time, information quality, and information sharing: A simulation study. *Production and Operations Management* **13** 340-354.

Croson, R. and K. Donohue. 2003. Impact of POS data sharing on supply chain management: An experimental study. *Production and Operations Management* **12**1 1-12.

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

Croson, R., and Donohue, K. 2002. Experimental economics and supply-chain management. *Interfaces* **32**5 74-83.

Croson R., K. Donohue, E. Katok and J. Sterman. 2007. Order stability in supply chains: coordination risk and the role of coordination stock, Penn State Working Paper.

Dogan, G. and Sterman, J.D. 2006. "I'm not hoarding, I'm just stocking up before the hoarders get here" Behavioral causes of phantom ordering in supply chains, MIT Working paper.

Fischbacher, U. 2007. z-Tree: Zurich Toolbox for Ready-made Economic Experiments, *Experimental Economics* 10(2), 171-178.

Kahneman, D. and Tversky, A. 1972, Subjective probability: A judgment of representativeness, Cognitive Psychology, 3, 430-454.

Lee, H., P. Padmanabhan and S. Whang. 1997. Information distortion in a supply chain: The bullwhip effect. *Management Science* **43** 546-58.

Miron, J. and S. Zeldes. 1988. Seasonality, cost shocks, and the production smoothing model of inventories. *Econometrica* **56** 877-908.

Steckel, J.H., S. Gupta, and A. Banerji. 2004. Supply chain decision making: Will shorter cycle times and shared point-of-sale information necessarily help? *Management Science* **50**4 458-465.

Sterman, J. 1989. Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science* **35**3 321-340.

Tversky, A. and Kahneman, D. 1971, The belief in the law of small numbers, Psychological Bulletin, 76, 105-110.

Wu, Y. and E. Katok. 2006. Learning, communication, and the bullwhip effect. *Journal of Operations Management* **24**6 839-850.

Appendix

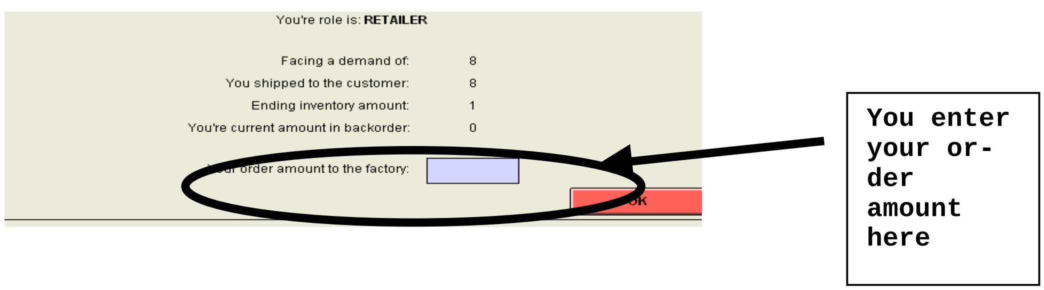
In today's study, you will participate in one game where you will earn money based on your own ordering decisions. If you follow the instructions carefully and make good decisions, you could earn a considerable amount of money. The unit of currency for this session is called a franc.

Description of the Game: You sell "widgets." In this game you order "widgets" over multiple rounds from a supplier. You will find out the demand from your customer before you place an order.

Each of you will participate in teams of two. One of you will be assigned the role of a Retailer; the other will be assigned the role of a Factory.

Your decision is to select an order quantity (or just simply place an order) from your supplier. Orders come from suppliers and are shipped to customers. The Factory's customer is the Retailer. The Retailer's customer is the consumer that is programmed into the computer software.

A sample ordering box is provided below:



Game Incentives:

If you have unsold "widgets" at the end of a period, called **Overages**, this quantity will be carried over to the beginning of the next period, and the holding cost of each unit in inventory is **0.50 franc**.

If you do not have enough "widgets" to meet your customer's demand, these units will enter your **Backlog**, which will also be carried over to the beginning of the next period, and the cost of each unit in backlog is **1.00 franc**.

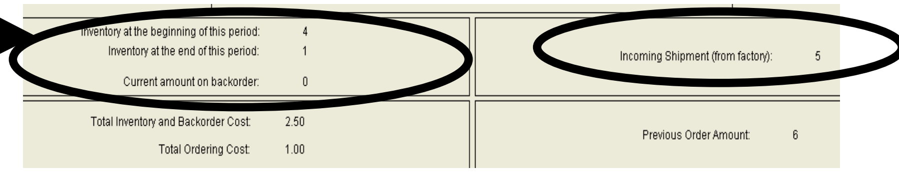
Example: Suppose Starting Inventory is 4, and you received an incoming shipment of 10 units of "widgets." If the customer demand during this period is 10, then:

Ending Inventory = 4 + 10 - 10 = 4, and your overage cost is 4 x 0.50 = 2 francs.

If customer demand in this period turns out to be 18, then

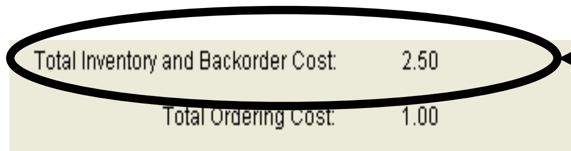
Ending inventory = 4 + 10 – 18 = -4. The backlog of 4 costs 4 x 1 = 4 francs.

The status of your inventory is provided here.



Incoming shipments are added to your beginning inventory.

Costs:
Recall that any inventory that you have on-hand at the end of the period is charged at **0.50 francs** per unit of inventory. Also, any orders that are backlogged is charged **1.00 francs per unit.** This information is available to you as described below.



Status of inventory and backorder cost are provided here.

Ordering Costs:

Suppose you decide to change the amount of widgets that you want to order from your supplier. If you desire to order an amount that is different from your order quantity in the previous period, then you will be charged a fee to do so. The fee that you are charged depends on the extent of the change in order quantity. At a minimum, you will be charged **0.50 francs** multiplied by the change in order quantity level.

Example: Suppose in Period 1 you ordered 10 widgets from your supplier. In Period 2, you desire to order 30 widgets from your supplier. Then, the ordering cost for changing order quantity level is the following:

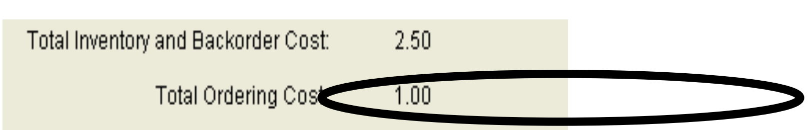
Ordering Cost: 30 – 10 = 20. The ordering change fee is: 20 x 0.50 = **10 francs.**

If you decide to order 10 widgets in Period 2, then there isn't any ordering cost.

Ordering Cost: 10 – 10 = 0. The ordering change fee is 0 x 0.50 = **0 francs.**

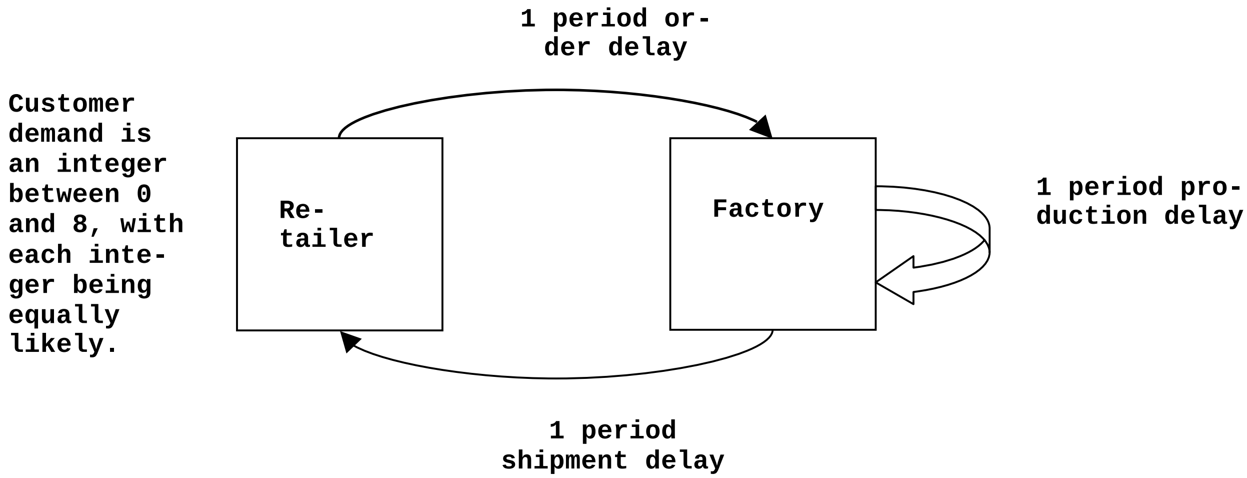
If you decide to order 0 widgets in Period 2, then the ordering cost for changing order quantity levels is the following:

Ordering Costs: 10 – 0 = 10. The ordering change fee is 10 x 0.50 = **5 francs.**



Status of order cost is provided here.

Delays:
Retailer:
There is a **1** period delay from the time that the retailer places an order until it arrives to the factory. There is a **2** period delay from the time that the retailer places an order until that shipment is received by the retailer.
Factory:
There is a **1** period delay from the time that the factory places the order until it arrives to the factory.



Customer Demand:
The retailer's customer demand is an integer from 0 to 8, with each integer from 0 to 8 being equally likely. Demand in one period has no effect on demand in the other period. The customer will demand units of "widgets" from the Retailer. The Retailer will demand units **directly** from the Factory.

How you make money:
Your decision is to select an order quantity for "widgets." You will make **50** decisions in this game. Starting Inventory is set to 4 units at the beginning of the game. For the first two periods, your supplier will provide you with 4 units. You will earn money by making ordering decisions which results in the lowest total supply chain cost for your team. Supply chain cost consists of: total inventory and backorder cost.

How you will be paid:
You will participate in one game, which consists of 50 decisions. Each team is given an *endowment* of *630* tokens at the beginning of the game. All team members' costs will be added together to calculate the total team costs in order to get your final earnings which is based on the below formula: Earnings = (Endowment – Total Supply Chain Costs)/2 * Conversion Rate + Show-up fee
Both team members will earn the same amount. The lower your team's costs are, the more money you will earn in this game. Thus, your objective is to make ordering decisions that minimize the total costs of your team over the entire game. However, it is possible for a team to go bankrupt during the game. If your team's endowment minus your chain costs becomes negative before the 50 weeks end, your earnings will be ONLY the show-up fee.