Contracting in Supply Chains: A Laboratory Investigation

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

Diana Wu
School of Business, University of Kansas, Lawrence, KS 66045
dianawu@ku.edu.

We investigate the performance of three commonly-studied supply chain contracting mechanisms: the wholesale price, the buyback, and the revenue-sharing contracts. The simple setting we consider is one with a two-member supply chain in which the retailer faces the newsvendor problem, the supplier has no capacity constraints, and the delivery occurs instantaneously. This is the setting for which clear theoretical benchmarks are known: there exists a family of coordinating buyback and revenue-sharing contracts that differ only in how the profit is distributed between the retailer and the supplier. We compare the three mechanisms in a laboratory setting that controls for strategic interactions between the retailer and the supplier, and find that although the buyback and the revenue-sharing contracts improve supply chain efficiency relative to the wholesale price contract, the improvement is smaller than the theory predicts. We also find that although the buyback and the revenue-sharing contracts are mathematically equivalent, they do not generally produce equivalent outcomes for retailers but they do for suppliers.

Keywords: Supply Chain Contracts, Experimental Economics

History: This paper was originally submitted in June 2006. Current draft date: May 2, 2007.

1. Introduction and Motivation

There has been a great deal of interest in analyzing contracting mechanisms that can be used to coordinate supply chains. The question of how to align the economic incentives of supply chain partners is important because whenever a supplier charges a wholesale price to a retailer in excess of his own production cost, *double marginalization* (Spengler 1950) causes the retailer to order less than the first-best optimal amount. These smaller retailer orders imply that both supply chain members forego potential profits. Much of the past research focused on the analytical design of contracting mechanisms that can help avoid this inefficiency by providing retailers incentives to order more (see Cachon 2003 for a review). These theoretical studies provide understanding and characterization of various supply chain contracts. However, they are not sufficient to explain why one contractual form should be adopted over another in practice. Probably because of the difficulties in obtaining field data, empirical investigations on how different contracts affect supply chain performance are relatively rare and inconclusive. To bridge the gap in the literature, we provide a laboratory test of three widely-used contracts in controlled settings designed to conform to assumptions of the contracting models being tested. Prior laboratory research primarily focused on the newsvendor retailer's behavior when faced with a simple wholesale price contract. Behaviors are found to be systematically different from the risk-neutral profit-maximizing benchmark, as well as inconsistent with a number of behavioral decision-theoretic models (Schweitzer and Cachon 2000). The behavior of either retailers or suppliers when faced with coordinating contracts—the focus of our paper—has not previously been studied in the laboratory.

We extend prior research on supply chain coordination and contracting in three ways. First, we investigate the retailers' behavior when faced with coordinating contracts and find those contracts to be significantly less effective than the theory suggests. Second, we study the suppliers' behavior in structuring coordinating contracts, and find that suppliers are not able to offer contracts that fully coordinate the supply chain, even when retailers are programmed to order optimally given contract parameters. Third, we compare two equivalent coordinating contracts, from both, the retailers' and the suppliers' perspectives, and find that they do not induce equivalent behavior by retailers. Our explanation for the difference in performance based on prospect theory and framing (Kahneman and Tversky 1979) organizes the data reasonably well.

This paper is structured as follows. In the next section we discuss some of the related literature, analytical, empirical, and experimental, to provide the necessary background for our study. In section 3 we describe the details of the experimental design and the laboratory protocol we used. We then present the details and the results of the retailer game (section 4) and the supplier game (section 5). In section 6 we develop a prospect-theory-based explanation for the differences observed in the retailers' behavior under two equivalent coordinating contracts, and present additional data that tests this explanation. In section 7 we summarize our finding, point to limitations and directions for future research, and discuss managerial implications of our work.

2. Analytical Background and Related Literature

In the simplest supply chain contracting setting analyzed theoretically, and the one we investigate here, the retailer faces the classic newsvendor problem, and orders from a supplier. The supplier has no capacity constraint and delivers instantaneously. In the baseline model (Spengler 1950), the *wholesale price contract*, the retailer faces an exogenous stochastic demand with distribution $F()$ and an exogenous market price p, and suffers losses whenever his actual order quantity q differs from the realized demand D. The retailer maximizes his expected profit by balancing off the cost of ordering too much or too little, and to do that he sets his order q to satisfy

$$F\left(q^*\right) = \frac{p-w}{p},$$

which is known as the *critical fractile*.

In contrast, the supplier incurs no risk because when the production cost is c and the wholesale price is $w > c$, he simply makes a profit of $(w-c)\times q$ on the retailer's entire order. The wholesale price that maximizes the supplier's profit in the wholesale price contract depends on the demand distribution. If $F()$ is uniform from A to B then the optimal wholesale price w^* is given by

$$w^* = \min\left\{p, \frac{c}{2} + \frac{p}{2}\frac{B}{B-A}\right\}. \qquad (1)$$

To *coordinate* the supply chain, a contract must give the retailer incentives to order the same amount that would be optimal in a centralized setting. Cachon (2003) provides a review of the analytical work that investigates various contractual arrangements that facilitate this coordination. One useful class of arrangements shares the demand risk between the retailer and the supplier by making the supplier's profit depend on realized sales.

Cachon and Lariviere (2005) look at two such *risk-sharing* contracts, the *buyback* and the *revenue sharing*, and show that the two are mathematically equivalent. In the buyback contract the supplier pays the retailer a rebate of b on all unsold units, thus assuming some of the risk associated with over-ordering. In the revenue sharing contract, the supplier induces higher retailer order through a lower wholesale price, but in return he receives a portion of the gross revenue *r*. If we let $0 < \lambda < 1$ be the retailer's share of the total profit, a continuum of coordinating risk-sharing contracts can be constructed, one for each λ. For a buyback contract to coordinate the supply chain, pairs of parameters $\{w_{BB}, b\}$ need to satisfy

$$\begin{aligned} p - b &= \lambda p \\ w_{BB} - b &= \lambda c \end{aligned} \tag{2}$$

Cachon and Lariviere (2005) also show that the revenue sharing contract $\{w_{RS}, r\}$ is equivalent to the buyback contract $\{w_{BB}, b\}$ when

$$\begin{aligned} r &= b \\ w_{BB} &= w_{RS} + b \end{aligned} \tag{3}$$

We chose to investigate risk-sharing coordinating contracts in this paper because this class of contracts is simple and the buyback contracts are quite commonly observed in industries such as publishing, computer software and hardware, and pharmaceuticals (Padmanabhan and Png 1995). We thus believe they represent a logical point for the initial study of coordinating mechanisms. To provide the reader with a sense of the variety of other contracting arrangements we briefly mention several most recent analytical studies (we direct the reader to Cachon 2003 for a comprehensive review of the literature). Choi et al. (2004) develop a supply contract menu that combines supplier's service level and expected backorders. Kamrad and Siddique (2004) show the flexibility of supply contracts can be Pareto improving. Corbett et al. (2004) examine the value to a supplier of being able to offer contracts that are more general than the wholesale price, including two-part linear and nonlinear schemes. Attention has also been paid to procurement contract (Wu and Kleindorfer 2005), option contract (Burnetas and Ritchken 2005, Kleindorfer and Wu 2003), warranty contract (Balachandran and Radhakrishnan 2005), and target-rebate contract on false failure returns (Ferguson et al. 2005).

Field studies on supply chain contracts are limited. Mortimer (2004) analyzes the effect of using revenue sharing contract on supply chain members' profits and consumer welfares in video-rental industries. Gopal et al. (2003) study contract choice for offshore software development projects. Azoulay and Shane (2001) use evidence from business format franchising industries to argue that different information about contracting of entrepreneurs determines the efficiency of the contract implemented.

Laboratory studies related to contracting that investigate the retailer's behavior focus almost exclusively on wholesale price contracts. Schweitzer and Cachon (2000) find that retailers place orders that tend to be between the optimal orders and the average demand and note that this "pull-to-center" effect cannot be explained by risk preferences, loss aversion, or prospect theory. They suggest that this behavior is consistent with minimizing ex post inventory error and the anchoring and insufficient adjustment heuristic. Bolton and Katok (2007) also find the "pull-to-center" effect and additionally show that performance does improve over time with extensive experience, although slowly, and restricting decision-makers to placing standing orders[1] speeds up learning substantially. Lurie and Swaminathan (2005) report a similar finding, that too frequent feedback can degrade performance and slow down learning. Benzion et al. (2005) vary the demand distribution and find that orders are affected by both the average demand and the last period's demand, but this bias is weakened slowly over time—participants learn. Bostian, Holt and Smith (2007) find that an adaptive learning model explains the "pull-to-center" effect.

Although the retailer's ordering behavior when faced with the wholesale price contract is reasonably well studied, there is little work on behaviors in other types of contracting arrangements. The only study we are aware of is Katok, Thomas and Davis (2007), who investigate service level agreements and find that in general longer review periods are more effective at inducing higher orders. The present study is the first to investigate the performance of risk sharing coordinating contracts in the laboratory.

Several recent papers study suppliers' behavior, all in settings with human subjects in both, the supplier and the retailer roles. In the settings with two human players, participants tend to gravitate towards contracts that split profits equally. Keser and Paleologo (2004) find that retailers are likely to

[1] In this setting, a *standing order* refers to a restriction that forces a retailer to place one order that is used for several consecutive periods (in the case of Bolton and Katok (2007) this was 10 periods).

reject contracts with high wholesale prices and report a tendency for players to choose wholesale price contracts that split profits approximately equally when the entire order is sold. Ho and Zhang (2004) look at a bilateral monopoly setting, and compare wholesale price contracts with two-part linear and two-part non-linear tariffs. They find that, contrary to theoretical predictions, two-part linear and two-part non-linear tariffs fail to coordinate the supply chain or yield equivalent allocations of profit, and, moreover, they do not produce results that are significantly more efficient than the wholesale price contract. The designs in which human retailers and suppliers interact, although realistic, do not control for social preferences (Bolton and Ockenfels 2000) and this makes the results in the two above-mentioned studies difficult to interpret. For example, lower-than-optimal wholesale price induces higher retailer orders, and therefore, in a high profit setting, it can be attributed to the anchoring and adjustment bias (Schweitzer and Cachon 2000). However, the same low wholesale price also causes more equal distribution of the supply chain profits between the two parties and thus can be due to decision makers' inequality aversion preferences. By looking at retailers and suppliers independently, as we do in the present study, we can separate decision biases from social preferences.

3. Experimental Design and Protocol

We organize this study into two games. In the *Retailer Game* the human player is in the role of the retailer, facing a computerized supplier. In the *Supplier Game* the human player is in the role of the supplier, offering contracts to a computerized retailer programmed to place risk-neutral profit-maximizing orders. In both games the retailer faces a single-period inventory-ordering problem with stochastic customer demand (the newsvendor problem). The three contracting arrangements we study are: the wholesale price contract (W), the buyback contract (BB) and the revenue sharing contract (RS). In each session participants make 100 decisions under *two* different contracts, the wholesale price contract and one of the coordinating contracts. This *within-subjects* design allows for the direct comparison of decision for each subject under two different contracts, and in theory it automatically controls for individual differences across subjects. Because subjects have to complete two different tasks, we vary the order of the tasks for approximately half the subjects, and this allows us to control for the order effects in our statistical analysis (Camerer 2003, p. 40-42).

In all treatments we set the supplier's production cost to be $c = 3$ and the retail price to be $p = 12$ to create a setting with potential high supply chain profits[2]. To investigate the effect of loss aversion (Kahneman and Tversky 1979) on behavior we use two different uniform demand distributions. In the DLOW condition $D \sim U(0,100)$, and in the DHIGH condition $D \sim U(50,150)$. Demand realizations are always rounded to the nearest integer. Under these two demand settings, retailers can always lose money under the wholesale price contract, but suppliers cannot. However, both retailers and suppliers can lose money under optimal coordinating contracts in DLOW, but not in DHIGH.

In summary, the baseline study that we present in sections 4 and 5 includes a $2 \times 2 \times 2$ full factorial design with a total of 8 treatments that manipulate the decision-maker's role (Retailer Game and Supplier Game), the two contract types used for a within-subjects comparison (W and BB in two different orders, W and RS in two different orders), and the customer demand distribution (DLOW, DHIGH). We summarize the experimental design and sample sizes in Table 1.

	Demand	Game 1 = W Game 2 = BB	Game 1 = BB Game 2 = W	Game 1 = W Game 2 = RS	Game 1 = RS Game 2 = W
Retailer Game	DLOW	9	6	7	8
	DHIGH	9	8	8	8
Supplier Game	DLOW	9	8	10	8
	DHIGH	11	8	10	10

Table 1. Experimental design and sample sizes.

In addition to the eight baseline treatments summarized in Table 1, we conducted four additional treatments (39 subjects) in the DHIGH demand condition that focus specifically on comparing the buyback and the revenue sharing contracts. We describe these treatments in Section 6.

[2] We focus on the high profit condition (critical fractile > ½) in this paper because it is the setting with higher potential gains from coordination.

[4] The human subject approval for this study required us to store the decision data using subject IDs and we do not have a way of connecting these IDs to individuals. Our recruitment system gives us a way to track individual participants and their earnings by session, and we used this information for the analysis.

All experimental sessions followed the same protocol. Participants arrive at the computer lab at a pre-specified time and read experimental instructions that describe the rules of the game, the use of the software, and the payment procedures. After all participants had a chance to read the instructions, the monitor read instructions to them aloud to ensure common knowledge, used PowerPoint slides to illustrate examples and formulas, and answered questions. Participants then complete 100 rounds under the first of the two contracts. After all participants finished this first game we handed out to them additional instructions describing the second contract, gave them a chance to read these new instructions, read it to them aloud and answered questions, before they began the second game. After completing the second game, participants were paid their actual earnings accumulated from both games in private and in cash. Participants were not allowed to communicate during the experiment.

All sessions were conducted at the Laboratory for Economic Management and Auctions (LEMA) at Penn State Smeal College of Business during the summer of 2005. Each session lasted for approximately 75 minutes and average earnings, including a \$5 participation fee, were \$19. Participants were Penn State students, recruited through a web-based recruitment system, with cash being the only incentive offered. A total of 175 subjects joined in our study. None of them participated in more than one session. The majority of our participants were undergraduates (77%) and the rest were graduate students. About 29% of the students major in Business, approximately the same percentage major in engineering, and the rest are in humanities, social sciences and natural science. Females and males constitute 40% and 60% of the subject pool, respectively. We compared the average earnings[4] by student level, major and gender for each session using a t-test, and found no response biases by those demographic characteristics.

4. The Retailer Game

4.1 Methods

In this game the human decision-maker is in the role of the retailer, and we set the wholesale price w for both demand conditions using (1):

$$w_W^{DLOW} = \min\left\{12, \frac{3}{2} + \frac{12}{2}\frac{100}{100-0}\right\} = 7.5$$

$$w_W^{DHIGH} = \min\left\{12, \frac{3}{2} + \frac{12}{2}\frac{150}{150-50}\right\} = 10.5$$

For the buyback contract, we use (2) and set $\lambda = 1/3$ so that both parties can benefit from coordination, to get $w_{BB} = 9$ and $b = 8$. We then construct the equivalent revenue sharing contract using (3) $r = 9$ and $w_{RS} = 1$. Given those parameters, the order quantity that maximizes the retailer's expected profit under the wholesale price contract is 37.5 in the DLOW condition and 62.5 in the DHIGH condition. Under the two coordinating contracts, the optimal order quantity is 75 in the DLOW condition and 125 in the DHIGH condition.

4.2 Research Hypothesis

Our first two hypotheses follow directly from the theory. Under the wholesale price contract the expected profit-maximizing retailer will order less than the first-best order quantity due to double marginalization, so we have:

Hypothesis 1A (*Double Marginalization*): The retailers' average orders for wholesale price contracts will be below the first-best order quantities.

Specifically, theory implies $q = 37.5$ in the DLOW condition, which is below the first-best quantity of 75, and 62.5 in the DHIGH condition, which is below the first-best quantity of 125.

We expect both the buyback and the revenue sharing contracts to induce higher orders from the retailer than the wholesale price contract does, since in theory both can coordinate the supply chain.

Hypothesis 2A (Coordination): The retailers' average orders for the coordinating buyback and the revenue sharing contracts will be higher than for the wholesale price contract. Theoretical predictions are 75 for DLOW and 125 for DHIGH.

The third hypothesis speaks to the fact that the buyback and the revenue sharing contracts are mathematically equivalent. This equivalence implies that we should not observe any differences in the performance of the two mechanisms.

Hypothesis 3A (Equivalence): The retailers' average orders for buyback contracts will be equal to those in revenue sharing contracts.

Our fourth hypothesis is aimed at linking theoretical predictions with known behavioral biases. Previous literature on ordering behavior in the newsvendor problem that we discussed in section 2 deals with the wholesale price contract only. A finding that Schweitzer and Cachon (2000) were the first to report and that several studies subsequently replicated is that average orders are located between the optimal orders and the average demand. Bostian et al. (2007) refer to this as the "pull-to-center effect." Schweitzer and Cachon (2000) note that the data, while inconsistent with many established behavioral models (risk preferences, prospect theory, loss aversion) is consistent with (i) "anchoring and insufficient adjustment" heuristic, and (ii) a preference for minimizing ex post inventory error. While those two explanations can both account for the pull-to-center effect, they imply different adjustment patterns. The anchoring and adjustment heuristic implies that orders start close to the average demand, and adjust over time in the direction of optimal orders and *away from* average demand. Minimizing ex post inventory error implies that orders are positively correlated with past demand. While not mutually exclusive, the two explanations have different implications about how orders adjust over time, which leads us to our fourth hypothesis.

Hypothesis 4A (Causes for the pull-to-center effect):

1. *Anchoring and insufficient adjustment*: The retailers' average orders for all contracts will start between the expected profit-maximizing benchmark and the average demand and adjust down under the wholesale price contract and up under the two coordinating contracts.
2. *Minimizing ex post inventory error*: The retailers' orders will be positively correlated with past demand.

4.3 Results

To provide a snapshot of the data, we start by reporting retailers' average order quantities and standard deviations over 100 periods across the three contracts and the two demand conditions. In Table 2 we report theoretical benchmarks for the three contracts in both demand conditions and compare them to average retailer orders, for games 1 and 2 separately.

	DLOW			DHIGH		
Contract	Optimal	Game 1	Game 2	Optimal	Game 1	Game 2
Wholesale	37.5	42.24 (5.15)	42.02 (6.75)	62.5	82.51 (9.25)	80.07 (11.15)
Buyback	75	56.16 (5.91)	52.14 (11.27)	125	117.31 (12.26)	109.51 (14.35)
Revenue Sharing	75	64.01 (8.06)	65.80 (14.05)	125	102.64 (18.29)	101.86 (20.69)

Table 2. Average retailer orders and standard deviations (in parenthesis) for the three contracts and two games.

Table 3 summarizes the hypotheses testing results for HA1-HA3. We start the data analysis by comparing average retailer orders for the two games separately. We use the sign test for one-sample comparisons and the Wilcoxon test for two-sample comparisons. We can generally reject the null hypothesis that the average retailer orders for the wholesale price contracts are not different from their theoretical predictions, and average orders for the wholesale price contract are all significantly below the first-best orders. This first set of comparisons is consistent with H1A (double marginalization).

	DLOW		*DHIGH*	
	Game 1	Game 2	Game 1	Game 2
1. $\mathrm{H_o}: q_W = q_W^*$; $\mathrm{H_a}: q_W \neq q_W^*$	**0.0220**	**0.0490**	**0.0001**	**0.0001**
2. $\mathrm{H_o}: q_W = q_{FB}^*$; $\mathrm{H_a}: q_W \neq q_{FB}^*$	**0.0001**	**0.0001**	**0.0001**	**0.0001**
3. $\mathrm{H_o}: q_{BB} = q_{BB}^*$; $\mathrm{H_a}: q_{BB} \neq q_{BB}^*$	**0.0030**	**0.0310**	0.7270	0.1800
4. $\mathrm{H_o}: q_{RS} = q_{RS}^*$; $\mathrm{H_a}: q_{RS} \neq q_{RS}^*$	0.0700	0.4530	0.2890	0.0700
5. $\mathrm{H_o}: q_W = q_{BB}$; $\mathrm{H_a}: q_W \neq q_{BB}$	**0.0003**	**0.0460**	**0.0001**	**0.0001**
6. $\mathrm{H_o}: q_W = q_{RS}$; $\mathrm{H_a}: q_W \neq q_{RS}$	**0.0002**	**0.0015**	**0.0214**	**0.0077**
7. $\mathrm{H_o}: q_{BB} = q_{RS}$; $\mathrm{H_a}: q_{BB} \neq q_{RS}$	0.1996	0.1014	**0.0500**	0.4230

Table 3. Hypotheses testing in the Retailer Game.
Comparisons 1-4 use the sign test and 5-7 the Wilcoxon test.

We can reject the null hypothesis that the average retailer orders for the buyback contract match their theoretical benchmarks in DLOW but not in DHIGH (sign-test p-value=0.143 if data from the two games are pooled). We cannot generally reject the null hypothesis that the average retailer orders for the revenue sharing contract are the same as their theoretical benchmarks. Although in some cases significantly below theoretical benchmarks, average retailer orders for coordinating contracts are always significantly above average retailer orders for wholesale price contracts. The above analysis indicates that coordinating contracts do induce higher orders, consistent with H2A. However, in some cases orders do not increase as much as the theory predicts.

Due to small sample sizes and the low power of the non-parametric test, the Wilcoxon test does not pick any statistical differences between the buyback and the revenue sharing contracts in DLOW[5] but if we pool the DLOW data from the two games, the differences become significant (p-value = 0.007). In the DHIGH condition the differences are statistically significant in game 1 but not in game 2. In this latter case, a more powerful t-test does not detect any differences in game 2 either (t-test p-value = 0.385). We conclude that overall the data does not support H3A although the differences seem to decrease (and in the case of DHIGH disappear) in the second game.

Since H4A deals with dynamic behavior over time we will test it using a regression model that we fit for each demand condition, contract type and game separately.

$$Q_{it} = Intercept + \beta_t \times t + \beta_D \times D_{i,t-1} \tag{4}$$

In this model the dependent variable Q_{it} is participant i's order in period t, and $D_{i,t-1}$ is the demand participant i observed in period t-1. We present regression estimates for model (4) in Table 4.

[5] A more powerful parametric t-test does pick up a difference in game 1 (p=0.0356) and a marginal difference in game 2 (p=0.0826).

	DLOW		*DHIGH*	
Variable	Game 1	Game 2	Game 1	Game 2
Wholesale Price Contract				
Intercept	37.2444*** (1.1996)	38.8642*** (1.0490)	75.2163*** (1.8407)	74.4162*** (1.8973)
t	-0.0280* (0.0164)	-0.0356** (0.0143)	-0.0713*** (0.0167)	-0.0601*** (0.0172)
D_{t-1}	0.1285*** (0.0162)	0.1009*** (0.0142)	0.1087*** (0.0165)	0.0864*** (0.0170)
Buyback Contract				
Intercept	52.5517*** (1.6667)	44.5837*** (2.0412)	118.2963*** (3.0966)	99.4822*** (2.9195)
t	-0.0561*** (0.0199)	0.0038 (0.0244)	-0.0950*** (0.0246)	0.0856*** (0.0232)
D_{t-1}	0.1287*** (0.0205)	0.1484*** (0.0251)	0.0378 (0.0253)	0.0581** (0.0238)
Revenue Sharing Contract				
Intercept	65.7807*** (1.7678)	60.8390*** (1.8898)	91.8665*** (3.0966)	96.9685*** (3.0966)
t	-0.0950*** (0.0211)	0.0006 (0.0226)	0.0968*** (0.0246)	0.0635*** (0.0246)
D_{t-1}	0.0614*** (0.0217)	0.0998*** (0.0232)	0.0589** (0.0253)	0.0177 (0.0253)

*p-value < 0.10; **p-value < 0.05; ***p-value < 0.01

Table 4. Coefficient estimates and standard errors in parenthesis for model (4).

The anchoring and adjustment heuristic implies that average orders move, over time, in the direction of the optimal solution. We generally do not observe this: the β_t coefficients for the buyback and the revenue sharing contracts should be positive according to the hypothesis, but they are either negative or not significant everywhere except the DHIGH game 2 treatments. The β_t coefficients for the wholesale price contracts are generally negative, which is consistent with anchoring and adjustment. So overall, the evidence for H4A1 (anchoring and adjustment) is mixed. The behavior in the wholesale price contract is generally consistent with that heuristic, as previous studies report. However, orders under coordinating contracts do not generally adjust towards the optimum, but on the contrary, often move away from the optimum and towards the average demand. We find more support for hypothesis, H4A2, because all β_D coefficients are positive, and most of them are statistically significant.

We also use (4) to gain further insight into the dynamic behavior by using the F-test for linear restrictions on the coefficients (Chipman and Rao 1964). The intercepts are not significantly different from 37.5 in DLOW (F-test p-values are 0.8313 in game 1 and 0.1935 in game 2), so average orders start out nearly optimal and correlation with last period's demand causes them to increase over time. Simultaneously, these average wholesale-price orders also decrease over time in response to the economic incentives in the structure of the newsvendor problem. The two effects are in opposite directions, but on balance orders increase over time and average orders for the entire 100 period game are above 37.5. The dynamics are similar in DHIGH, but initial orders are already significantly above 62.5 (F-test p-values < 0.001 in both games).

Intercepts for the buyback and the revenue sharing contracts are below their optimum levels of 75 in DLOW and 125 in DHIGH (F-test p-values < 0.001 in all treatments except in DHIGH game 1 buyback, where p= 0.0304). In fact, the reader can observe from Table 4 that several of these coefficients are even below average demand (not significantly so). In the DHIGH condition in game 1 average buyback orders start out high (but lower than optimum) but then decrease over time. In the other three DHIGH coordinating contract treatments average retailer orders start out at or slightly below average demand and increase over time. In the DLOW condition average orders start above average demand (but below optimum) in game 1, but decrease over time. In game 2 the buyback contract starts our especially low (below average demand), while the revenue sharing contract is above the average demand but below optimum. There is no significant time trend for either of these contracts in game 2.

The bottom line is that while anchoring and adjustment appears to be a reasonably accurate description of behavior under the wholesale price contract (due to negative β_t coefficients), we find no consistent evidence of this behavior under coordinating contracts. In some treatments average orders move towards optimum, and in other treatments they move away from it. But in all treatments there is strong positive correlation between orders and last period demand, indicating that the desire to minimize ex post inventory error (or the regret from having a mismatch between the order and the demand, as

Kremer et al. 2007 point out) plays a role. So on balance, we find support for H4A2, but only for H4A1 in the case of the wholesale price contract.

5. Study 2: The Supplier Game

5.1 Methods

In the second study we look at decisions on contract parameters from the supplier's perspective to better understand the extent to which suppliers are able and willing to offer contracts that coordinate. The design includes an automated retailer programmed to act in accordance with the theory, and this feature helps control for any potential strategic interactions between suppliers and retailers. The retailer is programmed to place expected-profit maximizing orders given a contract offered by the human supplier. Since *F*() is *U*(*A,B*), the automated retailer's order is given by

$$q^* = A + (B-A)\left(\frac{p-w-r}{p-b-r}\right) \tag{5}$$

where $A = 0$ and $B = 100$ in the DLOW condition and $A = 50$ and $B = 150$ in the DHIGH condition. Note that for a uniformly-distributed demand, expected sales given an order of q^* are given by

$$E\left[S\right] = \left(\frac{A+q^*}{2}\right)\left(\frac{q^*-A}{B-A}\right) + q^*\left(\frac{B-q^*}{B-A}\right),$$

and correspondingly, the expected total profit amounts for the retailer, supplier and the total supply chain is given by

$$\begin{aligned}
\pi^R &\equiv E\left[\pi^{Retailer}\right] = (p-r)E\left[S\right] - wq^* + b\left(q^* - E\left[S\right]\right)\\
\pi^S &\equiv E\left[\pi^{Supplier}\right] = (w-c)q^* + rE\left[S\right] - b\left(q^* - E\left[S\right]\right)\\
\pi^T &\equiv E\left[\pi^{Total}\right] = pE\left[S\right] - cq^*.
\end{aligned} \tag{6}$$

We set $b = r = 0$ for the wholesale price contract, and participants selected *w* only. In the buyback contract we set $r = 0$, and participants selected *w* and *b* simultaneously. And in the revenue sharing contract we set $b = 0$, and participants selected *w* and *r* simultaneously.

Since the retailer is automated, the system provides feedback to suppliers about the order quantity that will follow a proposed contract (this is q^* as defined by (5)). Each participant can try different contract parameters and observe the expected (but not the actual) outcome as many times as he chooses before making the final decision for the round. We repeated this procedure for all treatments in the Supplier Game (W, BB and RS) to make certain our participants had access to the relevant information that the theory implicitly assumes they have, thus giving the theory its best shot.

5.2 Research Hypothesis

In formulating research hypothesis we are guided by the theory, which provides normative benchmarks about the families of contracts that *should* be implemented by rational risk-neutral players. The two metrics of interest are (i) the average retailer order quantity q^* as defined by (5), and (ii) the average expected supplier profit π^S [6] as defined by (6). These two metrics are not equivalent because the average retailer order quantity is related to the overall supply chain efficiency, but the supplier's expected profit is also related to the portion of the total profit the supplier extracts from the supply chain. Since retailers are automated, in principle, suppliers can extract the entire profit, but as the results in the following sections show, this does not happen. Looking at both, the suppliers' expected average profit and the retailers' average order helps present a fuller picture.

Our first hypothesis deals with the wholesale price contract. In this contract, the supplier's decision does not involve any risk since his profit is independent of the demand realization—the retailer assumes all demand risk. Our first hypothesis reflects the belief that, absent risk, rational decision-makers will be able to find the expected-profit-maximizing contracts.

Hypothesis 1B (Double Marginalization): Suppliers' average wholesale prices for the wholesale price contract will:

[6] We do not use the actual supplier profit since it depends on the random demand realization and thus is less informative than the expected supplier profit implied by the contract.

1. Maximize suppliers' expected profits. Theoretical benchmarks are w = 7.5 in DLOW which yields an expected supplier profit of 168.75, and w = 10.5 in DHIGH which yields an expected supplier profit of 468.75.
2. Induce average retailers' orders below the first best. Theoretical benchmarks are Q = 37.5 in DLOW and Q = 62.5 in DHIGH.

Our second hypothesis deals with the coordinating contracts, and since the two we are studying are mathematically equivalent, we do not distinguish between them:

Hypothesis 2B (Coordination): Suppliers will select coordinating contracts that will (relative to the wholesale price contract):

1. Induce higher average retailer orders. Theoretical benchmarks for the first-best order are 75 in DLOW and 125 in DHIGH.
2. Result in higher average expected supplier profits. The expected average first-best supply chain profit is 337.5 in DLOW and 787.5 in DHIGH, and since the retailers are automated, suppliers have the ability to extract the entire profit.

The third hypothesis speaks to the fact that the buyback and the revenue sharing contracts are mathematically equivalent. This equivalence implies that we should not observe any differences in the two mechanisms in terms of the average retailer orders or average expected supplier profits.

Hypothesis 3B (Equivalence): Buyback and revenue sharing contracts will be equivalent in terms of:

1. The average retailer order quantities they induce.
2. The average expected supplier profits they generate.

5.3 Results

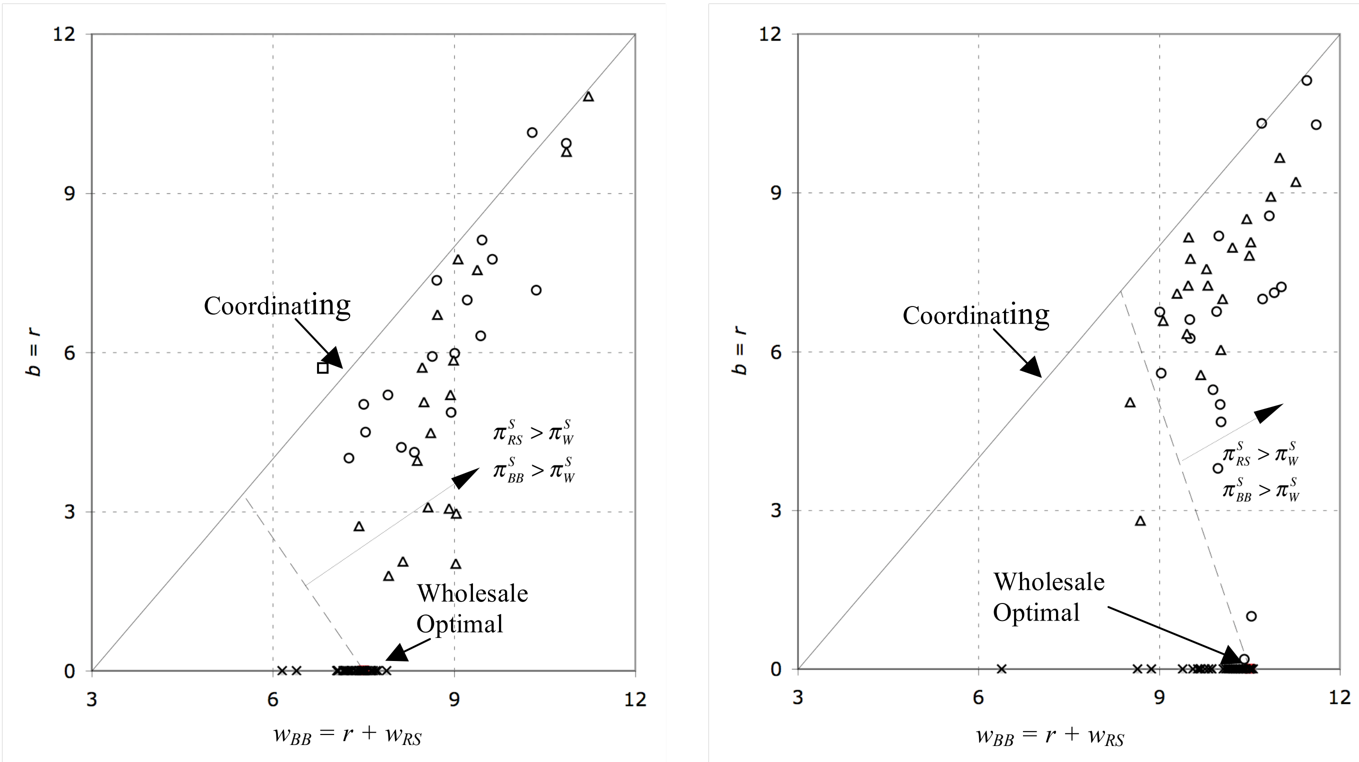
We start the analysis by reporting, in the top panel of Table 5, average contract parameters (w, b, r) across the three contracts and the two games, along with their standard deviations. In the bottom panel of Table 5 we report the averages and standard deviations of the corresponding retailer orders (q) and the suppliers' share of the total expected profit $1-\lambda=\pi^S/\pi^T$. Results of hypotheses tests are in Table 6.

	DLOW				DHIGH			
	Game 1		Game 2		Game 1		Game 2	
Parameter / Contract	w	b/r	w	b/r	w	b/r	w	b/r
Wholesale	7.43 (0.23)		7.26 (0.42)		10.06 (0.43)		9.96 (1.01)	
Buyback	9.24 (1.10)	6.98 (1.72)	8.59 (0.96)	5.76 (1.99)	10.29 (0.64)	6.68 (2.33)	10.25 (0.83)	6.20 (3.24)
Revenue Sharing	3.71 (2.09)	5.26 (2.76)	3.97 (2.04)	4.87 (2.73)	2.90 (1.26)	6.60 (1.70)	2.39 (0.83)	7.87 (1.19)
	q	$1 - \lambda$	q	$1 - \lambda$	q	$1 - \lambda$	q	$1 - \lambda$
Wholesale	38.41 (1.96)	0.66 (0.02)	39.85 (3.43)	0.64 (0.05)	66.44 (3.52)	0.81 (0.04)	67.37 (8.29)	0.80 (0.10)
Buyback	56.77 (11.48)	0.73 (0.11)	58.81 (15.55)	0.68 (0.09)	87.96 (17.84)	0.83 (0.06)	89.68 (22.57)	0.82 (0.08)
Revenue Sharing	48.57 (12.83)	0.74 (0.09)	50.64 (17.13)	0.74 (0.07)	97.55 (9.64)	0.75 (0.07)	97.06 (12.67)	0.82 (0.06)

Table 5. Averages and standard deviations (in parenthesis) for the contract parameters and corresponding order quantities and supplier's profit shares. All figures computed over 100 game periods and averaged across participants.

In Figure 1 we plot the average contracting parameters selected by individual participants —b or r on the y-axis and w_{BB} or $w_{RR}+r$ on the x-axis—for both demand conditions. The solid lines represent combinations of contract parameters that fully coordinate the supply chain, while the dashed lines show the buyback/revenue-sharing contracts that result in the same supplier expected profit as the optimal wholesale price contract (w = 7.5 in DLOW and 10.5 in DHIGH). The average wholesale prices in the DLOW condition are not significantly different from 7.5, but in the DHIGH condition they are slightly lower than 10.5[7]. The vast majority of coordinating contracts are located to the North-East of the optimal wholesale price contract and to the right of the dashed lines, thus they improve both, the total efficiency and the supplier's average expected profit.

[7] These differences in DHIGH remain significant even when we consider the last 50 periods of each game.

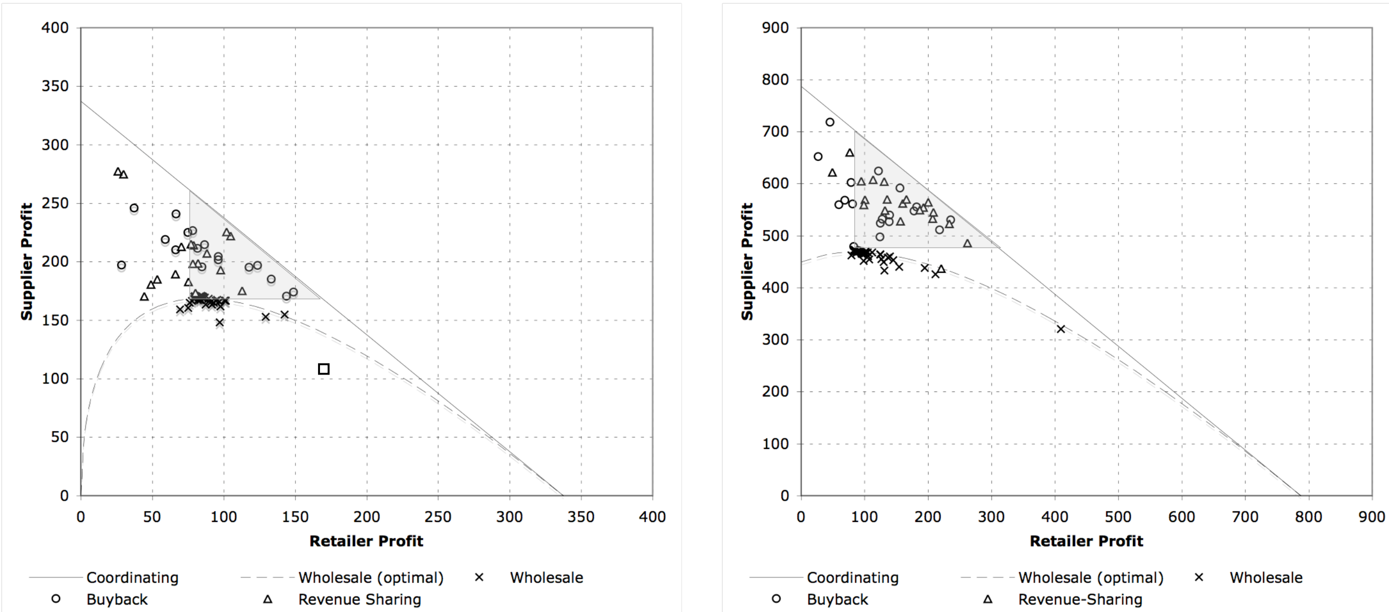


× Wholesale; △ Revenue sharing; ○ Buyback — (a) DLOW

× Wholesale ; △ Revenue sharing; ○ Buyback — (b) DHIGH

Figure 1. Contracting parameters chosen in the Supplier Game compared to theoretical benchmarks.

In Figure 2 we plot contracts in terms of the average expected retailer (on the x-axis) and supplier (on the y-axis) profits they yield for both demand conditions. The straight solid lines in the figure represent profits implied by all possible coordinating contracts, while the concave dashed lines represent profits implied by all possible wholesale price contracts. The supplier's optimal expected profit under the wholesale price contract is 168.75 in DLOW and 468.75 in DHIGH, and the actual average profits are not significantly different from the theoretical benchmark in the DLOW conditions but is lower in the first game of the DHIGH conditions but not in the second. The vast majority of coordinating contracts are located above the wholesale price contracts—they produce higher expected average supplier profits—but almost none are actually on the coordinating contract line. Most of the coordinating contracts in the DHIGH condition (but not in the DLOW condition) are also located to the right of the wholesale price contracts, inside the shaded triangle, indicating that they are Pareto improving relative to the wholesale price contracts. In the DLOW condition a good number of coordinating contracts are located to the left of the wholesale price contracts (outside the triangle), indicating that the retailer profit from the coordinating contract is lower (in absolute terms) than what it would be under the optimal wholesale price contract.



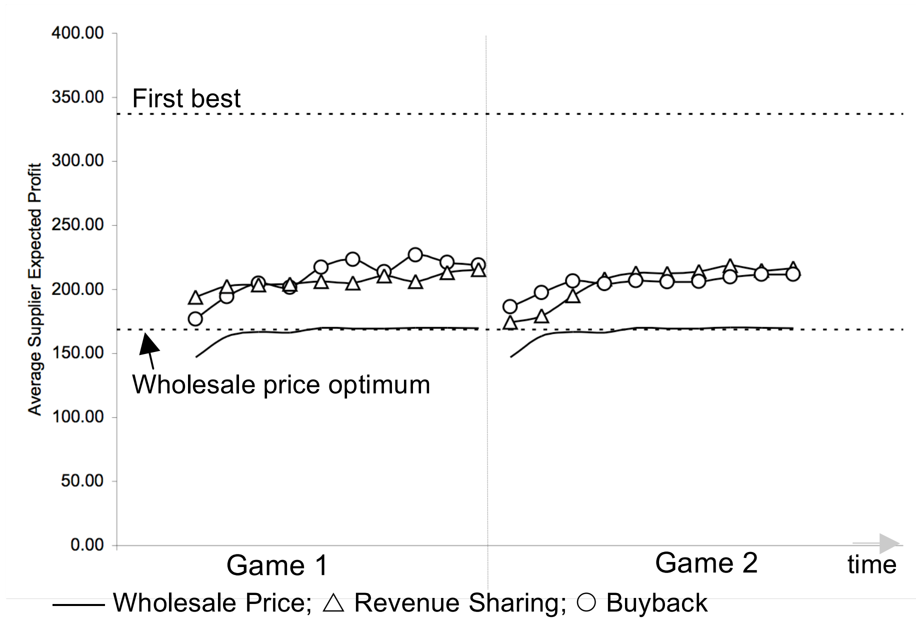
(a) DLOW (b) DHIGH

Figure 2. Profit distribution in the Supplier Game compared to theoretical benchmarks.

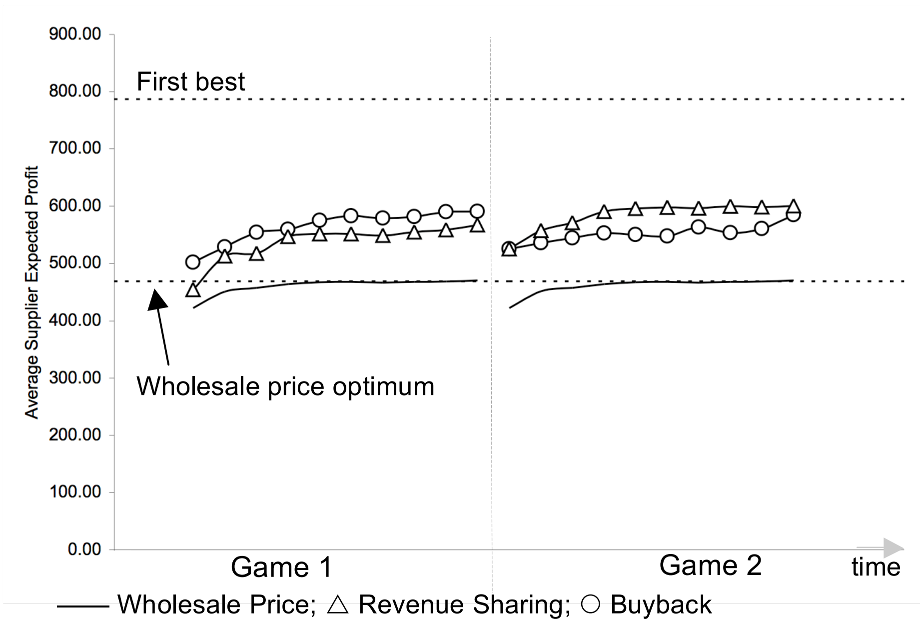
	DLOW		*DHIGH*	
	Game 1	Game 2	Game 1	Game 2
1. $\mathrm{H_o}: w = w^*$; $\mathrm{H_a}: w \neq w^*$	0.6480	0.1670	**0.0015**	**0.0075**
2. $\mathrm{H_o}: q_W = q_W^*$; $H_a: q_W \neq q_W^*$	**0.0414**	**0.0213**	**0.0000**	**0.0000**
4. $\mathrm{H_o}: \pi_W^S = \pi_W^{*S}$; $\mathrm{H_a}: \pi_W^S \neq \pi_W^{*S}$	0.4810	0.3020	**0.0118**	0.9999
4. $\mathrm{H_o}: q_{BB} = q_{BB}^*$; $\mathrm{H_a}: q_{BB} \neq q_{BB}^*$	**0.0078**	**0.0039**	**0.0078**	**0.0009**
5. $\mathrm{H_o}: q_{RS} = q_{RS}^*$; $\mathrm{H_a}: q_{RS} \neq q_{RS}^*$	**0.0078**	**0.0039**	**0.0078**	**0.0009**
6. $\mathrm{H_o}: \pi_{BB}^S = \pi_{BB}^{*S}$; $\mathrm{H_a}: \pi_{BB}^S \neq \pi_{BB}^{*S}$	**0.0078**	**0.0039**	**0.0078**	**0.0009**
7. $\mathrm{H_o}: \pi_{RS}^S = \pi_{RS}^{*S}$; $\mathrm{H_a}: \pi_{RS}^S \neq \pi_{RS}^{*S}$	**0.0078**	**0.0039**	**0.0078**	**0.0009**
8. $\mathrm{H_o}: q_W = q_{BB}$; $\mathrm{H_a}: q_W \neq q_{BB}$	**0.0022**	**0.0001**	**0.0000**	**0.0015**
9. $\mathrm{H_o}: q_W = q_{RS}$; $\mathrm{H_a}: q_W \neq q_{RS}$	0.1053	**0.0165**	**0.0000**	**0.0000**
10. $\mathrm{H_o}: \pi_W^S = \pi_{BB}^S$; $\mathrm{H_a}: \pi_W^S \neq \pi_{BB}^S$	**0.0001**	**0.0000**	**0.0000**	**0.0000**
11. $\mathrm{H_o}: \pi_W^S = \pi_{RS}^S$; $\mathrm{H_a}: \pi_W^S \neq \pi_{RS}^S$	**0.0000**	**0.0000**	**0.0003**	**0.0000**
12. $\mathrm{H_o}: q_{BB} = q_{RS}$; $\mathrm{H_a}: q_{BB} \neq q_{RS}$	0.2350	0.1530	**0.0342**	0.3450
13. $\mathrm{H_o}: \pi_{BB}^S = \pi_{RS}^S$; $\mathrm{H_a}: \pi_{BB}^S \neq \pi_{RS}^S$	0.5929	0.8065	0.7558	0.1489

Table 6. Hypotheses testing in the Supplier Game.
Comparisons 1-7 use the sign test and 8-13 the Wilcoxon test.

In Figure 3 we plot the average expected supplier profit over time for the three contracts and the two games in the DLOW and the DHIGH demand conditions. The average expected supplier profit under all contracts increases in the beginning of the game, and then settles, at optimal levels for the wholesale price contract, and substantially below the first best for coordinating contracts.



(a) DLOW demand condition.



(b) DHIGH demand condition.

Figure 3. Average expected supplier profit over time in the Supplier Game.

The optimal retailer order quantities are $q^* = 37.5$ in DLOW and 62.5 in DHIGH, and average retailer order quantities are slightly higher in both demand conditions and both games. Average retailer orders that the two coordinating contracts induce are not close to the first best. However, coordinating contracts do induce higher average retailer orders than the wholesale price contracts do.

In summary, we find qualitative support for H1B and H2B—wholesale price contracts generally perform close to their theoretical benchmarks and significantly worse than the first best. Coordinating contracts perform better than the wholesale price contracts, consistent with H2B, but contrary to H2B, substantially worse than the first best.

Proceeding to test H3B, we find that the only difference, from the supplier's perspective, between the two coordinating contracts we can detect is in the average retailer order quantity induced in game 1 of DHIGH (the retailer order is lower under the buyback contract than under the revenue sharing contract). So unlike H3A, the data is mostly consistent with H3B.

6. The Buyback and Revenue-Sharing Equivalence

6.1 Methods

To further investigate differences between the two coordinating contracts, we repeated both, the retailer game and the supplier game[8] for the buyback and revenue-sharing treatments, with the following two changes to the experimental procedure (we call these new treatments *DHIGH demand with the DLOW decision frame*):

1. We conducted all games under the DHIGH demand condition (demand uniformly distributed from 50 to 150) but we described this demand to participants as:

   > "There is a guaranteed demand of 50 units. There is an additional demand … which is an integer from 0 to 100, each equally likely."

   We described the decision task as follows:

   > "Your task is to determine how many widgets to order, in addition to the guaranteed 50 units… your order always has to be an integer from 0 to 100."

[8] The main purpose of this additional analysis is to further explore the Retailer game, since the buyback and the revenue sharing contracts are nearly equivalent in the Supplier game. We conducted the corresponding Supplier game treatments primarily for completeness.

Thus, we placed the task into the DLOW frame (since the order is from 0 to 100), and by separating the demand into two parts (the 50 guaranteed units and the up to 100 additional units) we also moved the uncertain demand into the DLOW frame.

2. We conducted the sessions for the two games within subjects, but now each subject played the buyback and the revenue sharing game (no wholesale price game). As before, we varied the order of the games to control for the order effects.

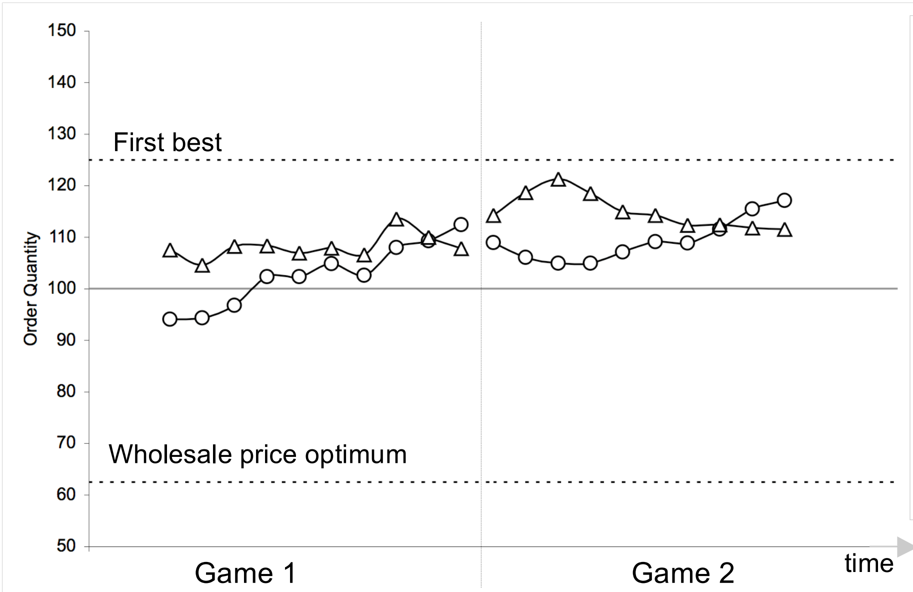
Additional 39 human subjects participated in these new treatments. Of those, twenty participated in the retailer game, with ten playing the buyback/revenue-sharing sequence, and the other ten playing the revenue-sharing/buyback sequence. Nineteen participants played the supplier game, ten in the buyback/revenue-sharing sequence and nine in the revenue-sharing/buyback sequence.

6.2 Retailer Game Results

We summarize results of relevant hypotheses tests in Table 7, and we plot average retailer orders for both contracts in both games over time in Figure 4.

	Game 1	Game 2
1. H_o: $q_{BB} = q_W^*$; H_a: $q_{BB} \neq q_W^*$	**0.0019**	**0.0019**
2. H_o: $q_{RS} = q_W^*$; H_a: $q_{RS} \neq q_W^*$	**0.0019**	**0.0019**
3. H_o: $q_{BB} = q_{BB}^*$; H_a: $q_{BB} \neq q_{BB}^*$	**0.0019**	0.1090
4. H_o: $q_{RS} = q_{RS}^*$; H_a: $q_{RS} \neq q_{RS}^*$	**0.0215**	**0.0019**
5. H_o: $q_{BB} = q_{RS}$; H_a: $q_{BB} \neq q_{RS}$	0.0889	0.9690

Table 7. Hypotheses Testing. Comparisons 1-4 use the sign test and 5 the Wilcoxon test.



Summary Information
Average retailer orders over 100 periods, standard deviations (in parenthesis)

Buyback
Game 1: 102.7
(9.8)
Game 2: 109.4
(13.7)

Revenue Sharing
Game 1: 115.0
(15.2)
Game 2: 108.2
(13.4)

△ Revenue Sharing; ○ Buyback

Figure 4. Average retailer order quantities in the two games plotted over time.

Average retailer orders for both of these contracts continue to be above the wholesale price optimum but mostly below the first best. Average retailer orders are marginally lower for buyback than for revenue sharing in game 1 but are not different in game 2. They are also not different if we pool the data in both games ($p = 0.234$). Recall that the average retailer orders were similarly higher under the revenue sharing contract than under the buyback contract in the DLOW condition, indicating that it is the DLOW decision frame (and not the DHIGH demand distribution) that is responsible for the initial differences.

6.3 Retailer Game Discussion

Our results indicate that retailers do not perceive the buyback and the revenue sharing contracts as equivalent, and in this section we offer a potential explanation. In the retailer game, participants start out ordering more under the revenue sharing contract in the DLOW condition, but in the DHIGH condition they order more under the buyback contract. Even though mathematically equivalent, the equivalence of the two contracts is not immediately apparent because the wholesale price is a payment that depends directly on the order amount, and thus can be clearly viewed as a loss, but the effect of the buyback rebate and the revenue share depends on the realization of uncertain demand. There is a substantial amount of evidence that people dislike losses more than they dislike foregone gains (Kahneman and Tversky 1979, Kahneman et al. 1990). In the DLOW condition, the retailer is not guaranteed any revenue (since the demand can be as low as 0) so the potential losses from paying the wholesale price loom large, while potential gains from a rebate (*b*) for unsold units or the foregone losses from having to pay a revenue share (*r*) for the sold units, seem less salient. Consequently, the low up-front wholesale price of the revenue sharing contract is more effective than the high rebate of the buyback contract in inducing high order quantities.

In the DHIGH demand condition the retailer is guaranteed to sell at least 50 units, thus the revenue share (*r*) becomes a salient loss, at least for those 50 units, and now the revenue sharing contract looks like it panelizes high orders. The buyback contract, in contrast, looks like it rewards high orders, since a substantial rebate is paid for unsold units. This (admittedly somewhat tenuous) explanation may well rationalize initial difference between average retailer orders in the first game. These differences, however, shrink over time, even in game 1, and differences by the end of game 1 are quite small. There are no statistically significant differences in retailer orders in the second game in the DHIGH condition.

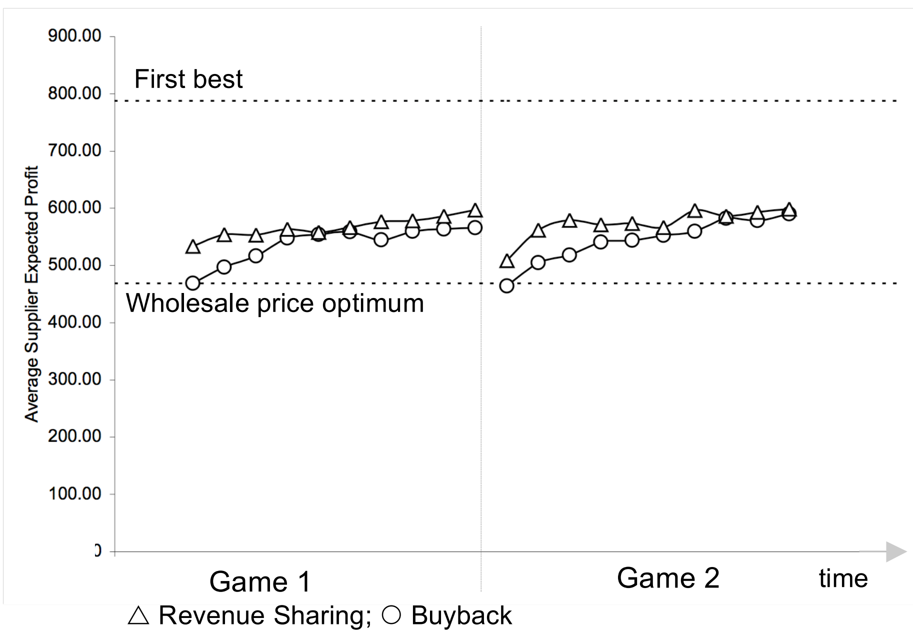
When we describe the DHIGH demand as 50 guaranteed units and a random number of additional units, and frame the decision in terms of the number of additional units to order, we move the task into the DLOW frame, and just as in the DLOW condition, retailers start by ordering more under the revenue sharing contract than under the buyback contract. This is because having to pay the revenue share for the first 50 guaranteed units now does not look like part of the decision. But the differences in order amounts for the two contracts quickly disappear, and no differences can be detected in the second game.

6.4 Supplier Game Results

We plot average expected supplier profit for the two new treatments in both games over time in Figure 5, report averages and standard deviations for expected supplier profits, retailer order quantities, and the suppliers' share of the total expected profit in Table 8 and summarize hypotheses tests in Table 9. As in the supplier game study (Section 5), average supplier profits are significantly below the first best and significantly above the wholesale price optimum expected profits; and similarly, the two coordinating contracts do not differ much in terms of the retailer order quantities they induce or the supplier profits they generate.

	Supplier's Expected Profit		Order Quantity		Supplier's Share	
	Game 1	Game 2	Game 1	Game 2	Game 1	Game 2
Buyback	537.63 (118.94)	543.52 (79.77)	100.36 (21.24)	99.68 (22.00)	0.76 (0.16)	0.78 (0.12)
Revenue Sharing	573.57 (45.42)	566.88 (63.97)	101.27 (13.39)	100.92 (19.39)	0.79 (0.07)	0.79 (0.08)

Table 8. Averages and standard deviations (in parenthesis) of the supplier's expected profit, retailer's order quantity, and the supplier's share of the total expected profit.



Summary Information
Average w, b and r over 100 periods, standard deviations (in parenthesis)

Buyback

	w	b
Game 1:	9.63 (1.55)	6.30 (3.19)
Game 2:	9.88 (1.19)	6.66 (2.34)

Revenue Sharing

	w	r
Game 1:	2.59 (2.33)	7.41 (2.44)
Game 2:	2.34 (0.94)	7.67 (1.36)

Figure 5. Average expected supplier profit in the two games plotted over time.

	Game 1	Game 2
1. $H_o: \pi_{BB}^{S} = \pi_{BB}^{*S}$; $H_a: \pi_{BB}^{S} \neq \pi_{BB}^{*S}$	**0.0019**	**0.0039**
2. $H_o: \pi_{RS}^{S} = \pi_{RS}^{*S}$; $H_a: \pi_{RS}^{S} \neq \pi_{RS}^{*S}$	**0.0039**	**0.0019**
3. $H_o: q_{BB}^{S} = q_{BB}^{*S}$; $H_a: q_{BB}^{S} \leq \pi_{BB}^{*S}$	**0.0215**	**0.0391**
4. $H_o: q_{RS}^{S} = q_{RS}^{*S}$; $H_a: q_{RS}^{S} \geq q_{RS}^{*S}$	**0.0039**	**0.0019**
5. $H_o: \pi_{BB}^{S} = \pi_{RS}^{S}$; $H_a: \pi_{BB}^{S} \neq \pi_{RS}^{S}$	0.3913	0.3913
6. $H_o: q_{BB} = q_{RS}$; $H_a: q_{BB} \neq q_{RS}$	0.8380	0.4380
7. $H_o: \lambda_{BB} = \lambda_{RS}$; $H_a: \lambda_{BB} \neq \lambda_{RS}$	0.6534	0.8383

Table 9. Hypotheses Testing. Comparisons 1-4 use the sign test and 5-7 the Wilcoxon test.

7. Limitations, Conclusions and Managerial Implications

7.1 Limitations and Directions for Future Research

There are two main limitations we wish to discuss, that also point to fruitful directions for future research. The first has to do with the subject pool. We used a subject pool that is most common in experimental economics (see Holt 1995, Kagel and Roth 1995): students, mostly undergraduates, recruited through advertisements offering to earn cash. Our subject pool is representative of the larger student population at Penn State in terms of gender, majors, and to the extent we are able to determine, ethnicity. There have been several studies comparing the performance of students and professionals in laboratory experiments, and very often these studies find no difference (see Plott 1987, Ball and Cech 1996, Katok, Thomas and Davis 2007). Specifically in the context of the newsvendor problem, Bolton, Ockenfels and Thonemann (2007) look at the effect of instructional learning, and find that teaching how to solve the newsvendor problem is highly effective with students, but has almost no effect on managers. Clearly, the effect of the subject pool on outcomes of operations management experiments is a complicated and important question, deserving of future study.

The second limitation is that we study the behavior of retailers and suppliers separately—our retailers and suppliers do not interact. We designed our experiments in this way intentionally, because we wanted to understand the individual decision making, unaffected by social utility considerations, such as inequality aversion or preferences for fairness (Bolton and Ockenfels 2000). In that, our present study represents an intermediate step because it provides a comparison that can be used to separate the effect of social preferences from other individual decision-making biases. But real contractual arrangements are negotiated with human participants on both sides, so a better understanding of how these arrangements compare in reality, should include a study with human retailers and suppliers interacting. Wu (2007) follows this direction and extends our experimental design to investigate the strategic interactions between supply chain partners under different contracts. They observed significant behavioral deviations from the theoretical benchmarks that can be explained by both social preferences for fairness and economic incentives for establishing successful long-term relationships.

7.2 Summary and Managerial Implications

We present a laboratory study in which we compare the performance of the wholesale price contract and two coordinating risk-sharing contracts: buyback and revenue sharing. We compare these mechanisms in two ways: first, from the retailer's perspective, we look at how retailers respond to different mechanisms. Second, from the supplier's perspective, we look at the suppliers' willingness and ability to take advantage of coordinating contracts.

We find that consistent with earlier studies, retailers on average place orders that are between the profit-maximizing order and the average demand. When faced with the wholesale price contracts, average retailer orders are higher than the expected-profit maximizing benchmark, and they adjust in the direction of optimal order over time. This initial over-ordering makes wholesale price contracts perform better in the laboratory than in theory. Coordinating contracts induce higher orders than do the wholesale price contracts, but fall short of the first best, so they perform worse in the laboratory than in theory.

Our study was the first to examine coordinating contracts in the laboratory, and we find that for those contracts especially, the way behavior changes over time is more consistent with preferences for minimizing ex post inventory error than with the anchoring and adjustment heuristic. We find that the two mathematically equivalent risk-sharing contracts do not induce identical initial retailer behavior in the laboratory, but differences decrease with experience. We offer a framing explanation for this initial lack of equivalence. How participants perceive the demand distribution is an important factor: the buyback contract emphasizes the benefit of placing high orders, so it is more effective when the demand distribution is framed in terms of having some minimum amount as well as a possibility of an upside. The revenue-sharing contract emphasizes low upfront cost, so it is more effective when the demand distribution makes the decision look like it has less of an upside and a serious downside. Our study in which we used the DHIGH demand distribution with the DLOW decision frame demonstrated that the framing explanation is useful for organizing this data.

The main finding from the supplier game is that while suppliers easily figure out how to use the wholesale price contract to achieve near-optimal performance, using the coordinating contracts (that require suppliers to set two parameters instead of one) is more problematic. Our suppliers are able to learn to use coordinating contracts to increase the retailers' orders relative to the wholesale price contract, but the resulting average retailer orders are still far below the first best. With human retailers, who are likely to place orders that are even lower than our computerized retailers, the effectiveness of coordinating contracts may well be even worse.

Our results indicate that decision support tools for contract design can increase the effectiveness of contractual arrangements for both, suppliers and retailers. From the retailers' perspective, tools are needed to counteract the demand following behavior resulting from trying to minimize the ex post inventory error. It is this behavior that decreases the effectiveness of coordinating contracts by making retailers less responsive to economic incentives (see Kremer et al. 2007). From the suppliers' perspective, decision support tools that help set contract parameters properly may well go a long way to increase not only the total supply chain efficiency, but also the suppliers' profit. When buyback contracts are prevalent in an industry, the natural tendency of contract designers may be to set both, the wholesale price and the rebate too low. Alternatively, a lack of understanding of how contracting mechanisms work, may lead to such obviously sub-optimal contracts as offering the full rebate, as for example, is the standard in the pharmaceutical industry. While such contracts may be rational in the face of retailer under-ordering, they are sure to lead to over-ordering, so effective decision support tools on both sides offer a promise for decreasing waste and increasing profitability.

Acknowledgments

This research was funded in part by the National Science Foundation Doctoral Dissertation Grant (# SES-0412900) and by the Smeal College of Business Doctoral Dissertation Award, Penn State University. We thank Axel Ockenfels and the Deutsche Forschungsgemeinschaft for financial support through the Leibniz-Program. Katok gratefully acknowledges the support from the Smeal College of Business and the Center for Supply Chain Research (CSCR) at Penn State's Smeal College of Business through the Smeal Summer Grants Program.

References

Azoulay, P. and Shane, S. (2001), Entrepreneurs, contracts, and the failure of young firms, *Management Science*, 47(3), 337-358.

Balachandran, K. R., and Radhakrishna, S. (2005), Quality implications of warranties in a supply chain, *Management Science*, 51(8), 1266-1277.

Ball, S. B., and P. Cech (1996), Subject Pool Choice and Treatment Effects in Economic Laboratory Research, *Experimental Economics*, vol. 6, 239- 292.

Benzion, U., Cohen, Y, Peled, R. and Shavit, T. (2005), Decision-making and the newsvendor problem—An experimental study, Working paper.

Bolton, G., Ockenfels, A. (2000) A theory of equity, reciprocity, and competition. *American Economics Review*, 90(1), 166–193.

Bolton, G., Ockenfels, A. and Thonemann U. (2007), The effect of subject pool on newsvendor behavior. Working Paper.

Bolton, G. and Katok, E. (2007), Learning-by-doing in the newsvendor problem: A laboratory investigation, Penn State Working Paper http://lema.smeal.psu.edu/katok/news_post.pdf.

Bostian, A,J. A., Holt, C.A. and Smith, A.M. (2007), The Newsvendor "Pull-to-Center Effect:" Adaptive Learning in a Laboratory Experiment, University of Virginia Working Paper.

Burnetas, A. and Ritchken, P. (2005), Option pricing with downward-sloping demand curves: The case of supply chain options, *Management Science*, 51(4), 566-580.

Cachon, G. (2003), Supply chain coordination with contracts, *Handbooks in Operations Research and Management Science: Supply Chain Management*, edited by Steve Graves and Ton de Kok, North Holland.

Cachon, G. and M. Lariviere (2005) Supply chain coordination with revenue sharing contracts, *Management Science*, 51(1), 30-44.

Camerer, C.F. (2003), *Behavioral Game Theory Experiments in Strategic Interactions*, Princeton University Press, Princeton, NJ.

Chipman, J. S. and Rao M. M. (1964), The treatment of linear restrictions in regression analysis, *Econometrica*, 32(1/2), 198-209.

Choi, K., Dai, J. G. and Song, J. (2004), On measuring supplier performance under vendor-managed-inventory programs in capacitated supply chains, *Manufacturing & Service Operations Management,* 6(1), 53-72.

Corbett, C. J., Zhou, D., and Tang, C. S. (2004), Designing supply contracts: contract type and information asymmetry, *Management Science*, 50(4), 550-559.

Ferguson, M. E., Guide, V. D., and Souza, G. C. (2005), Supply chain coordination for false failure returns, Working paper.

Gopal, A., Sivaramakrishnan, K., Krishnan, M. S., and Mukhopadhyay, T. (2003), Contracts in offshore software development: An empirical analysis, *Management Science*, 49(12), 1671-1683.

Ho, T. and Zhang, J. (2004), Does format of pricing contract matter? Working Paper.

Holt, Charles A. (1995) Industrial Organization: A Survey of Laboratory Results, in *Handbook of Experimental Economics*, edited by J. Kagel and A. Roth, Princeton, N.J.: Princeton University Press, 349-443.

Kagel, John H. and Alvin E. Roth, eds. (1995), *Handbook of Experimental Economics*, Princeton: Princeton University Press.

Kahneman, D., and Tversky, A. (1979), Prospect theory: An analysis of decision under risk, *Econometrica*, 47, 263-291.

Kahneman, D., Knetsch, J.L and Thaler, R.H. (1990), Experimental tests of the endowment effect and the Coase theorem, *Journal of Political Economy*, 98, 1325-1348.

Kamrad, B. and Siddique, A. (2004), Supply contracts, profit sharing, switching, and reaction options, *Management Science*, 50(1), 64-82.

Katok, E., Thomas, D. and Davis, A. (2007), Inventory service level agreements as coordination mechanisms: the effect of review periods. *Manufacturing & Service Operations Management*, conditionally accepted.

Keser, C. and Paleologo, G. (2004), Experimental investigation of retailer-supplier contracts: The wholesale price contract, IBM Research Working Paper.

Kleindorfer, P. R. and Wu, D. J. (2003), Integrating long- and short-term contracting via business-to-business exchanges for capital-intensive industries, *Management Science*, 49(11), 1597-1615.

Kremer, M., Minner, S., and Van Wassenhove, L. (2007), Regret and framing effects in the newsvendor model - the impact of framing and complexity, INSEAD Working Paper.

Lurie, N. H. and Swaminathan, J. M. (2005), Is timely information always better? The effect of feedback frequency on performance and knowledge acquisition, Working paper.

Mortimer, J. H. (2004), Vertical contracts in the video rental industry, Harvard Working paper.

Padmanabhan, V. and Png, I.P.L. (1995), Returns policies: make money by making good, *Sloan Management Review*, 37 (1), 65-72.

Plott, C. (1987), Dimensions of parallelism: some policy applications of experimental methods, in A. Roth, *Experimental Economics: Six Points of View*, Cambridge, University Press.

Schweitzer, M. and Cachon, G. (2000), Decision bias in the newsvendor problem: Experimental evidence, *Management Science*, 46(3), 404-420.

Spengler, J. J. (1950), Vertical integration and antitrust policy, *The Journal of Political Economy*, 58(4), 347-352.

Wu, D. J. and Kleindorfer, P. R. (2005), Competitive options, supply contracting, and electronic markets, *Management Science*, 51(3), 452-466.

Wu, Y. D. (2007), The impact of long-term relationship on supply chain contracts: a laboratory study, University of Kansas Working Paper.

On-Line Appendix: Sample Instructions

Sample instructions below are for the Same-Frame treatments. Instructions for DLOW and DHIGH treatments are analogous, but the demand distributions and order decisions are described directly, and profit formulas do not include the fixed component. The instructions for the Wholesale Price contract are also presented analogously, but without reference to the additional parameters.

Retailer Game Instructions

In today's study, you will participate in *two games* where you will earn money based on your own 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.

The Game Scenario
You are involved in the management of a supply chain that produces and sells widgets over multiple rounds. There are two members in each chain, a *Retailer* and a *Supplier*.

You are the retailer. The supplier is automated.

The Retailer buys widgets from the Supplier and sells widgets to the customer at **12** francs per unit. The supplier produces retailer's orders at **3** francs per unit. In each round, widgets are ordered and produced before you find out the actual customer demand.

Customer Demand
The customer demand per round consists of two parts:
1. There is a guaranteed demand of 50 units
2. There is an additional demand we will call **D**, which is an integer from 1 to 100, each equally likely. That is, there is a 1/100 chance that additional demand will be any one of the integers from 1 to 100

The total demand for the round is the sum of the guaranteed 50 units and the additional demand for that round. The additional demand drawn for any one round is independent of the additional demand for the earlier rounds. So a small or large additional demand in a round has no influence on whether additional demand is small or large in any other rounds.

Your task
Your task is to determine how many widgets to order, in addition to the guaranteed 50 units, so as to maximize your own total profits for the session. Your order always has to be an integer from 1 to 100. The number of additional units you order is called **Q.**

Feedback Information
In each round, after placing your order, you will be reminded of the order you just placed, the additional demand realized, the total customer demand, your own profit, and your automated supplier's profit for that round. The computer will also display the history of the above information for all pervious rounds.

How you will be paid
You will participate in two separate games, each lasting 100 rounds. The games will differ in the terms of the contract supplier offers. Your total earnings from both games will be converted to US dollars at the rate of 2000 francs per dollar, added to your participation fee of $5 and paid to you in cash at the end of the session.

Buyback Contract Game

Contract terms
The supplier offered you a contract that requires you to pay a wholesale price of **9** francs per unit you order, and agrees to buy back any units unsold at the end of the period at a rebate of **8** francs per unit. Remember that you sell each unit for **12** francs per unit.

Calculating Your Profit
The guaranteed demand is 50 units. For those units you will earn a profit of $50\times(12-9)=150$ francs. You will earn additional profit based on the number of additional units you order and sell.

When your order Q turns out to be the same or lower than the additional customer demand D, your total profit for the round is:

$$\begin{aligned} \text{Your Profit} &= 150 + (12 - 9) \times Q \\ &= 150 + \quad 3 \qquad \times Q \end{aligned}$$

For example, if the additional demand is 80 and you order 60, then your total profit for the round is 150 + 3 × 60 = 330 francs. Note that when the number of widgets ordered is less than demand, you lose opportunities for sales.

When your order Q turns out to be higher than the additional customer demand D, your total profit for the round is:

$$\begin{aligned} \text{Your Profit} &= 150 + (12 - 9) \times D - (9 - 8) \times (Q - D) \\ &= 150 + \quad 3 \qquad \times D - \quad 1 \qquad \times (Q - D) \end{aligned}$$

For example, if the additional demand is 40 and you order 60, then your total profit for the round is 150 + 3 × 40 – 1 × (60 – 40) = 250 francs. Note that when the number of widgets ordered exceeds demand, you must dispose of the unsold units (since extra widgets go stale after a round, and cannot be carried as inventory into future rounds), and thus incur cost for excess widgets. Since the supplier will refund you 8 francs for each unsold unit, it costs you 9 – 8 = 1 per unit.

Calculating Supplier's Profit
The supplier will earn (9 – 3) × 50 = 300 francs per round for the guaranteed demand, (9 – 3 = 6) for each additional unit you order and sell, and 9 - 8 – 3 = -2 franks for each unit you order and not sell.

For example, if you order 60 and the additional demand is 80, the supplier's profit for the round is 300 + 6 × 60 = 660 francs; if you order 60 and the additional demand is 40, then the supplier's profit for the round is 300 + 6 × 40 – 2 × (60 – 40) = 500 francs.

Revenue sharing Contract Game

Contract terms
The supplier offered you a contract that requires you to pay a wholesale price of **1** franc per unit you order, but you agrees to pay an additional **8** francs for each unit you sell. Remember that you sell each unit for **12** francs per unit.

Calculating Your Profit
The guaranteed demand is 50 units. For those units you will earn a profit of $50\times(12-1-8)=150$ francs. You will earn additional profit based on the number of additional units you order and sell.

When your order Q turns out to be the same or lower than the additional customer demand D, your total profit for the round is:

$$\text{Your Profit} = 150 + (12 - 1 - 8) \times Q$$
$$= 150 + 3 \times Q$$

For example, if the additional demand is 80 and you order 60, then your total profit for the round is 150 + 3 × 60 = 330 francs. Note that when the number of widgets ordered is less than demand, you lose opportunities for sales.

When your order Q turns out to be higher than the additional customer demand D, your total profit for the round is:

$$\text{Your Profit} = 150 + (12 - 1 - 8) \times D - 1 \times (Q - D)$$
$$= 150 + 3 \times D - 1 \times (Q - D)$$

For example, if the additional demand is 40 and you order 60, then your total profit for the round is 150 + 3 × 40 – 1 × (60 – 40) = 250 francs. Note that when the number of widgets ordered is greater than demand, you must dispose of the unsold units (since extra widgets go stale after a round, and cannot be carried as inventory into future rounds), and thus incur cost of 1 franc per unit for excess widgets.

Calculating Supplier's Profit
The supplier will earn (1 + 8 – 3) × 50 = 300 francs per round for the guaranteed demand, (1 + 8 – 3 = 6) for each additional unit you order and sell, and 1 – 3 = -2 franks for each unit you order and not sell.

For example, if you order 60 and the additional demand is 80, the supplier's profit for the round is 300 + (8 + 1 – 3) × 60 = 660 francs; if you order 60 and the additional demand is 40, then the supplier's profit for the round is 300 + 6 × 40 – 2 × (60 – 40) = 500 francs.

Supplier Game Instructions

In today's study, you will participate in *two games* where you will earn money based on your own 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.

The Game Scenario
You are involved in the management of a supply chain that produces and sells widgets over multiple rounds. There are two members in each chain, a *Retailer* and a *Supplier*.

You are the supplier. The retailer is automated.

The retailer decides how many widgets to buy from the supplier and sells widgets to the customer at **12** francs per unit. The supplier decides how to charge the retailer by specifying some contract terms and produces retailer's orders at **3** francs per unit. In each round, widgets are ordered and produced before you find out the actual customer demand.

Customer Demand
The customer demand per round consists of two parts:

1. There is a guaranteed demand of 50 units.
2. There is an additional demand we will call **D**, which is an integer from 1 to 100, each equally likely. That is, there is a 1/100 chance that additional demand will be any one of the integers from 1 to 100.

The total demand for the round is the sum of the guaranteed 50 units and the additional demand for that round. The additional demand drawn for any one round is independent of the additional demand for the earlier rounds. So a small or large additional demand in a round has no influence on whether the additional demand is small or large in any other rounds.

Your task
Your task is to determine the contract terms that specify the transfer payment between you and the retailer so as to maximize your own total profits in this session. The automated retailer has been programmed to place orders so as to maximize his own expected profits given the contract terms you offered.

Feedback Information
In each round, after you set the contract terms, the computer will display the optimal order quantity that the automated retailer places and the corresponding expected profit of the retailer. After confirming your decisions, you will receive information on the retailer's order, the additional and total customer demand realized, the retailer's and your own profits for that round. The computer will also display the history of the above information for all pervious rounds.

How you will be paid
You will participate in two separate games, each lasting 100 rounds. The games will differ in the type of the contract that you are asked to use. Your total earnings from both games will be converted to US dollars at the rate of 5000 francs per dollar, added to your participation fee of $5 and paid to you in cash at the end of the session.

Buyback Contract Game

Contract terms
You are asked to set a contract that specifies a wholesale price (**W**) at which you charge for each unit the retailer orders, and a rebate (**B**) at which you agree to buy back any units unsold from the retailer at the end of the period.

Note the wholesale price per unit you offer should be less than 12 francs (the retail price) and greater than 3 francs (your production cost). The rebate per unit you offer cannot be greater then the wholesale price you specify and cannot be less than zero. You are allowed to use at most two decimal places in setting W and B.

Calculating Your Profit
The guaranteed demand is 50 units. For those units you will earn a profit of $(W - 3) \times 50$ francs per round. You will earn additional profit based on the number of units (Q) in addition to the guaranteed demand that your automated retailer orders and sells.

When your retailer's additional order Q turns out to be the same or lower than the additional customer demand D, your total profit for the round is:

Your Profit $= [(W - 3) \times 50] + [(W - 3) \times Q]$

When your retailer's additional order Q turns out to be higher than the additional customer demand D, your total profit for the round is:

Your Profit $= [(W - 3) \times 50] + [(W - 3) \times D] + [(W-B) \times (Q - D)]$

For example, if you set the wholesale price at 9 francs/unit and the rebate at 7 francs/unit, the automated retailer will order 60 widgets in addition to the 50 units of guaranteed demand for the round. If the additional demand is 80, then your profit is $(9 - 3) \times (50 + 60) = 660$ francs for the round. If the additional demand is 40, then your profit is $(9 - 3) \times 110 - 7 \times (60 - 40) = 520$ francs for the round.

Note that for a fixed rebate (B), the higher the wholesale price you offer, the lower the quantity the retailer orders, and vice versa. While for a fixed wholesale price (W), the higher the rebate you offer, the higher the quantity the retailer orders, and vice versa.

Calculating Retailer's Profit
The retailer will earn $(12 - W) \times 50$ francs per round for the guaranteed demand, $(12 - W)$ for each additional unit he orders and sells, and $(B - W)$ francs for each unit he orders and not sell.

For the above example, given W = 9 and B = 7, the automated retailer orders an additional of 60 units for the round. If the additional demand is 80 for the round, the retailer's profit is $(12 - 9) \times 50 + (12 - 9) \times 60 = 330$ francs for the round. If the additional demand is 40, the retailer's profit is $(12 - 9) \times 50 + (12 - 9) \times 40 + (7 - 9) \times (60 - 40) = 230$ francs for the round.

Revenue sharing Contract Game

Contract terms
You are asked to set a contract that specifies a wholesale price (**W**) at which you charge for each unit the retailer orders, and a revenue share (**R**) (or commission) that you charge for each unit the retailer sells.

Note the wholesale price per unit per unit you offer should be less than 12 francs (retail price) and no less than zero. The revenue share per unit you acquire should be less than the retailer's total revenue per unit, which is the retailer price minus the wholesale price you offer, and should be no less than zero. You are allowed to use at most two decimal places in setting W and R.

Calculating Your Profit:
The guaranteed demand is 50 units. For those units you will earn a profit of $(W - 3 + R) \times 50$ francs per round. You will earn additional profit based on the number of units (Q) in addition to the guaranteed demand that your automated retailer orders and sells.

When your retailer's additional order Q turns out to be the same or lower than the additional customer demand D, your total profit for the round is:

Your Profit $= [(W - 3 + R) \times 50] + [(W - 3 + R) \times Q]$

When your retailer's additional order Q turns out to be higher than the additional customer demand D, your total profit for the round is:

Your Profit = [(W – 3 + R) × 50] +[(W – 3 + R) × D] + [(W – 3) × (Q – D)]

For example, if you set the wholesale price at 2 francs/unit and the revenue share at 7 francs/unit, the automated retailer will order 60 widgets in addition to the 50 units of guaranteed demand for the round. If the additional demand is 80, then your profit is (2 – 3 + 7) × (50 + 60) = 660 francs for the round. If the additional demand is 40, then your profit is (2 – 3 + 7) × 50 + (2 – 3) × 60 + 7 × 40 = 520 francs for the round.

Note that, for a fixed revenue share, the higher the wholesale price you offer, the lower the quantity the retailer orders, and vice versa. And for a fixed wholesale price, the higher the revenue share you charge, the lower the quantity the retailer orders, and vice versa.

Calculating Retailer's Profit
The retailer will earn (12 – W – R) × 50 francs per round for the guaranteed demand, (12 – W – R) for each additional unit he orders and sells, and (– W) francs for each unit he orders and not sell.

For the above example, given W = 2 and R = 7, the automated retailer orders an additional 60 units for the round. If the additional demand is 80 for the round, the retailer's profit is (12 – 2 – 7) × 50 + (12 – 7 – 2) × 60 = 330 francs for the round. If the additional demand is 40, retailer's profit is (12 – 2 – 7) × 50 + (12 – 2 – 7) × 40 + (– 2) × (60 – 40) = 230 francs for the round.