No. 39

UNIVERSITY OF COLOGNE

WORKING PAPER SERIES IN ECONOMICS

MANAGERS AND STUDENTS AS NEWSVENDORS - HOW OUT-OF-TASK EXPERIENCE MATTERS

GARY E BOLTON
AXEL OCKENFELS
ULRICH THONEMANN

Department of Economics
University of Cologne
Albertus-Magnus-Platz
D-50923 Köln
Germany

http://www.wiso.uni-koeln.de

Managers and Students as Newsvendors

How Out-of-Task Experience Matters

Gary E Bolton*
Penn State University, USA

Axel Ockenfels *
University of Cologne, Germany,

Ulrich Thonemann*
University of Cologne, Germany

August 25, 2008

We compare how freshmen business students, graduate business students and experienced procurement managers perform on a simple inventory ordering task. We find that, qualitatively, managers exhibit ordering behavior similar to students, including biased ordering towards average demand. Experience, however, affects subjects' utilization of information. The managers' work experience seems most valuable when there is only historical demand data to guide decision making, while students better utilize analytical information and task training. As a result, when information necessary to solve the problem to optimality is added to historical information, students catch up to the managers, and students with classroom experience in operations management outperform managers.

* Gary Bolton, Smeal College of Business, Penn State University, University Park, PA 16802 USA, gbolton@psu.edu (www.personal.psu.edu/faculty/g/e/geb3/). Axel Ockenfels, University of Cologne, Department of Economics, ockenfels@uni-koeln.de (ockenfels.uni-koeln.de/ao). Ulrich Thonemann, University of Cologne, Department of Business Administration, Albertus Magnus Platz, D-50923 Cologne, Germany, ulrich.thonemann@uni-koeln.de (www.wiso.uni-koeln.de/scmms/team/teamThonemann.html). We thank Elena Katok for developing and sharing the inventory software with us, Felix Lamouroux for very helpful research assistance, and the participants in numerous seminars in the U.S. and Germany for very helpful comments. And also to Mirko Kremer for a close read of the paper. Bolton gratefully acknowledges the support of the National Science Foundation. Ockenfels gratefully acknowledges the support of the Deutsche Forschungsgemeinschaft.

1 Introduction

How do work and classroom experiences affect managerial decision making? Here we examine this question in the context of the newsvendor problem, a foundational paradigm in management science. In the problem's most basic form, the newsvendor faces stochastic demand for a perishable product. Prior to seeing the actual demand draw, she must decide how much of the product to stock in inventory. Her realized profit is a function of the unit cost, the unit revenue, the order quantity and the realized demand. A key insight from the theoretical analysis of the newsvendor problem is that the order quantity that maximizes expected profit depends on a comparison of the marginal costs associated with being over and under stocked (Arrow et al. 1951).

Laboratory studies of the newsvendor problem provide descriptive insight into the intuitive judgments and information handling behind stocking behavior. These studies, mostly performed on college students, find that orders tend to be biased away from the expected profit maximizing order and towards the average demand, an effect known as "anchoring bias." This bias is at odds not only with expected profit maximization but also with alternative risk profiles (Schweitzer and Cachon 2000, Schultz et al. 2007). The bias persists for a variety of demand distributions (Benzion et al. 2008). It is influenced by task complexity and framing (Kremer, Minner and Wassenhove 2007).

Anchoring bias has proven robust to a variety of in-task experiences. It persists even with substantial task repetition, and descriptive statistical information on the performance of chosen or foregone orders has little positive influence (Bolton and Katok 2008). More frequent feedback can actually degrade performance (Lurie and Swaminathan 2007). While here is a good deal of heterogeneity in the individual ordering patterns behind anchoring bias (Moritz 2008), a common feature involves adaptive learning-by-doing behavior that insufficiently adjusts orders to the optimum, even when the experiment provides demand distribution and profit information amenable to deductive insight. The data shows fit with bounded rationality models of adaptive learning (Bostian, Holt and Smith 2008), decision noise and optimization error (Su 2008), and overconfidence bias in which subjects underestimate the variance in demand (Croson, Croson and Ren 2008).[1]

[1] For a comprehensive survey of newsvendor experiments, see Kremer and Minner (2008). Other recent work, using somewhat different methods for analyzing the newsvendor problem, also finds behavior that deviates from theory. In

Our experiment examines the influence of out-of-task experience on newsvendor performance. Both classroom and managerial experience can be expected to influence judgment and information handling, although perhaps in different ways. Classroom treatment of the newsvendor problem exposes students to the broad principles underlying inventory control. Actual procurement experience, which typically comes in addition to classroom experience, provides intensive exposure to practical inventory problems. Procurement managers are also subject to market selection pressure. So we might expect managers and students to approach the newsvendor problem differently.

The experiment samples three experiential groups: freshman business students who have had no course in operations management, graduate business students who have had at least one undergraduate course in operations management, and working managers with practical experience in newsvendor-type procurement. The managers in our sample all have at least one year of experience in newsvendor-type purchasing, with an average of 7.5 years of experience. Their positions at the time of the study ranged from buyer to company vice president. Since the ability to handle information is critical here, the experiment is designed to expose subjects to varying levels of information and task training so that we can examine how the interaction of information and out-of-task experience influences ordering performance. The specific questions we investigate are:

What experience leads to the best use of historical information? Most of the previous studies provided subjects with the information about the underlying demand distribution necessary to solve the problem to optimality. In the first phase of our experiment, we provide subjects with only historical information about demand, the kind of information condition that managers often face in the field.

What experience leads to the best use of analytical information? In our experiment, analytical information is phased in (after exposure to historical data) beginning with information on the demand distribution and then information regarding the expected profit of orders. Both levels of information are sufficient to identify the expected profit maximizing order, but the latter

---

an experiment, Brown and Tang (2006) find evidence for maximization of minimum profit and minimization of maximum loss. Using survey techniques, Corbett and Fransoo (2007) find that entrepreneurs and small businesses make inventory decisions that are consistent in some regards with the newsvendor model but inconsistent with regard to others. Inventory management has been studied in relationship to supply chain behavioral anomalies such as the bullwhip effect; see for example, Croson and Donohue (2006).

requires less deduction. Thus the experiment allows us to compare analytical sophistication among experience groups.

What experience leads to the best use of task training? Most subjects in the previous studies were unfamiliar with the newsvendor problem or at least assumed so. Where students were familiar, there was a significant lag between classroom experience and laboratory study. The training in our experiment is on-the-spot. It provides the tools necessary to determine the expected profit maximizing solution as well as an explanation for the underlying reasoning, distinguishing it from the flat provision of expected profit information.

We find that experience affects subjects' utilization of both historical and analytical information. Managers exhibit somewhat less anchoring bias in Period 1 of the task with only historical demand data for guidance and prior to receiving feedback on ones own performance. With the addition of analytical information or training, however, students catch up, and those with an operations management education, outperform the managers. The findings have implications for what we can expect from experienced managerial judgment as well as the conditions under which education and training are likely to be most successful.

While ours is the first study to examine the role of managerial experience in newsvendor problems, there have been studies of the influence of experience on other kinds of decision tasks. Seigel and Harnett (1964) compared the performance of corporate employees and students on a simple bargaining task and found no appreciable difference. Burns (1985) found that real world auction bidders tend to make more mistakes in the experimental auction than students. Cooper et al. (1999) compared the performance of Chinese students and Chinese managers on a task involving the 'rachet effect', the tendency for administrators to increase the work quota on those who have proven to be most productive in the past. They found that the managers play more strategically – but only when the game is framed in natural environment terms, not when it is framed in abstract terms. We will observe yet a different pattern in our study, although when taken together, a coherent picture of the role of experience begins to emerge (see Section 4).

Section 2 provides a description of the experiment. Section 3 lays out the data analysis and detailed findings. Section 4 discusses implications and concludes.

2 Method

The experiment has six treatments. Within each treatment, subjects play the same newsvendor game with phased information, explained in Subsection 2.1. Across treatments, we manipulate the experience subjects have as well as whether they get task training, explained in Subsection 2.2. Details of the experimental protocol are reviewed in Subsection 2.3. Table 1 provides an overview of the activities within a treatment.

Table 1. Overview of activities within treatments

Activity	Description
1. Task training (not included in all treatments)	In treatments with „Task training", subjects watch a one-hour video on the newsvendor problem. In all other treatments, this activity is entirely omitted
2. Information: Briefing and sample demand	Subjects read a two page briefing, including a graph showing the demand of the previous 50 periods (Appendix A.1)
3. Ordering: Phase 1 = Periods 1 – 40	Subjects place 40 orders, receiving feedback after each order (Appendix A.2)
4. Information: Demand distribution	Subjects receive a handout stating that demand is uniformly distributed between 1 and 100 and uncorrelated
5. Ordering: Phase 2 = Periods 41 – 80	Subjects place 40 orders, receiving feedback after each order. The process is the same as in Phase 1
6. Information: Optimal solution	Subjects receive a handout with a graph that shows how the expected profit of a period depends on the order quantity
7. Ordering: Phase 3 = Periods 81 – 100	Subjects place 20 orders, receiving feedback after each order. The process is the same as in Phases 1 and 2
8. Payment	Subjects are paid based on the performance across all 100 Periods

2.1 The newsvendor game

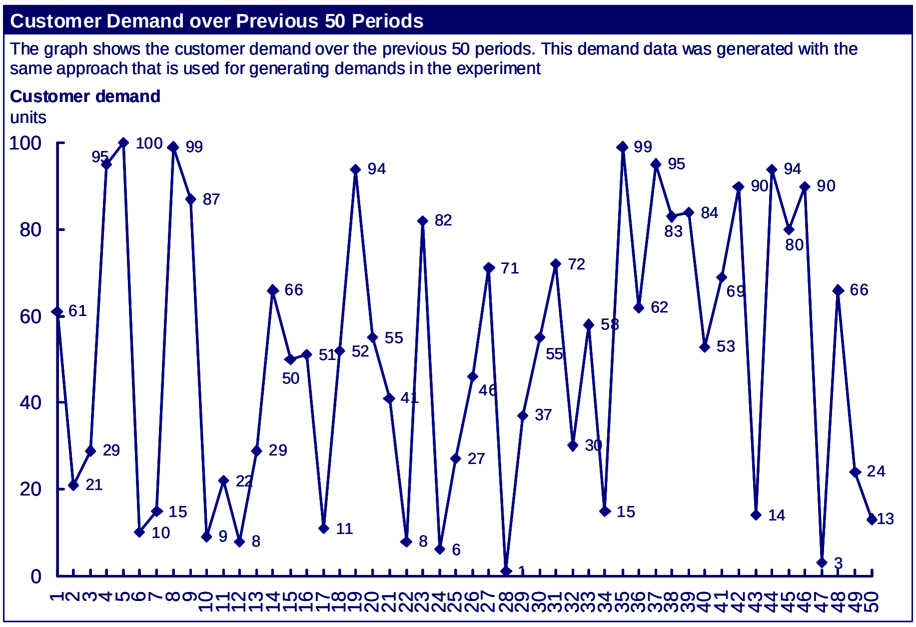
All treatments involve a 100 period newsvendor game with the same demand and economic parameters, and such that expected profits are maximized at an order quantity of 75 (equivalent to 75 percent of the maximum possible period demand). To simplify comparisons, all newsvendors faced the same sequence of independent draws.

Subjects in all treatments received the same basic pre-play briefing on the structure of the game (item 2 of Table 1). In synopsis: The purchase price to be paid for buying a unit of the "generic product" is 3 talers per unit and it can be sold to customers for 12 talers per unit. At the beginning of each of the 100 periods, the subject chooses an order quantity, after which demand for the period is revealed and profits are calculated. The briefing provides examples to illustrate profit calculation. Subjects are told that any unsold inventory remaining at the end of the period is “worthless” and not carried over to the next period. Subjects were also told the taler-to-euro exchange rate that would determine their total game earnings at the end of play. In addition, managers were asked for some biographical data.

The experiment phases in **information** about the game parameterization and newsvendor analysis. The briefing pointed out that new information would be added in three phases, before the first period, after the 40th period and after the 80th period, although the nature of this information was not discussed prior to introduction.

Phase 1. Sample demands: Before they place their first order, all subjects are presented with demands of 50 hypothetical periods that took place prior to the game, and are told that the first order they place is for the next period (period 51 of Figure 1). They are not told what distribution generated the draws.

Figure 1. Sample demands given to subjects prior to placing their first order (all treatments)



Phase 2. Demand information: After 40 periods, subjects receive the demand distribution information in Figure 2. Demand is drawn from a uniform distribution between 1 and 100 and independent between periods. With this, there is sufficient information to compute the expected profit maximizing order quantity (Arrow et al. 1951).

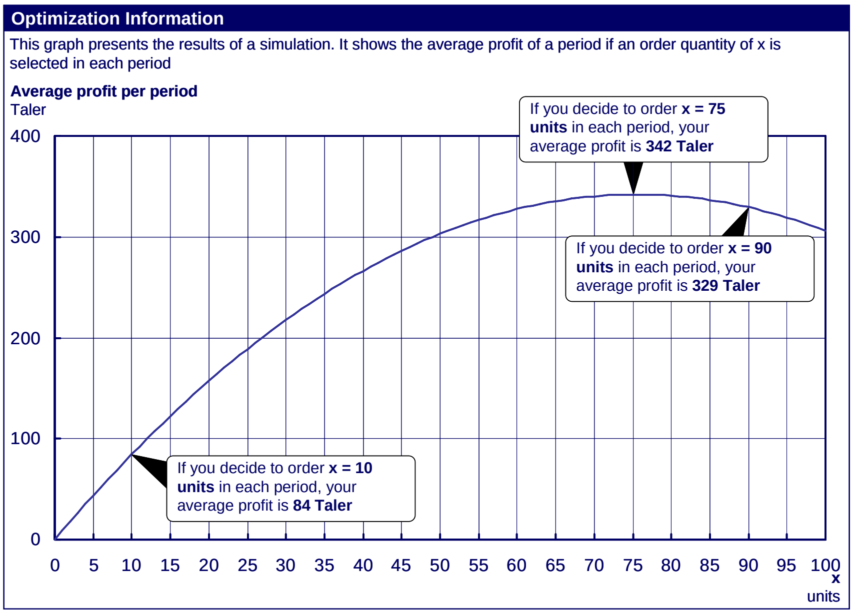
Figure 2. Information provided to subject after period 40

To create a customer demand the computer draws in each period a random number between 1 and 100. All customer demands between 1 and 100 are equally likely.

The demand of the current period is independent of the demand of the previous period. Therefore low or high demands in previous periods do not affect the demands of the following periods.

Phase 3. Expected profit information: After another 40 periods (so prior to period 81), subjects are presented the graph in Figure 3, which shows the expected profit as a function of the order quantity. The graph shows the order quantity (75) that results in the maximum expected profit. While the profit information is redundant in the sense that it could already be derived from the demand information, it is probably not redundant for (some) human subjects, who may have difficulty transforming demand information into the relevant expected profit information (which is in fact confirmed by the data as we show below).

Figure 3. Expected profit information provided to subjects after period 80



2.2 *Treatment manipulations: Experience and training*

The experiment has a fully crossed 3 X 2 design for a total of six treatments. One factor that distinguishes treatments is the **experience** the subjects have. This factor has three levels:

- **Freshmen**: Freshmen business students who have not yet had classroom education in operations management and are unlikely to have had work experience in purchasing.
- **Graduates**: Master degree students who have taken at least one operations management course and have completed their undergraduate studies in business administration. They are unlikely to have had much work experience in purchasing.
- **Managers**: Purchasing managers who have worked for at least one year in a line position making newsvendor-type decisions, such as ordering promotion items. 80% of the subjects were recruited from German retailers, 20% from German industry. Table 2 provides a breakdown by experience and position held at the time of the experiment.

Table 2. Breakdown of manager experience (number of subjects)

Position		Work Experience	
▪ Vice President	5	▪ 1 – 2 years	9
▪ Manager	14	▪ 3 – 4 years	14
▪ Team Leader	7	▪ 5 – 6 years	6
▪ Buyer	23	▪ 7 – 10 years	8
		▪ 11 – 15 years	6
		▪ 16 – 20 years	4
		▪ 21 – 25 years	1
		▪ 25 – 30 years	1

The second factor that distinguishes treatments is whether subjects are given **task training** just prior to playing the game. This factor has two levels:

- **Basic:** At the beginning of the game, subjects receive the briefing described in Subsection 2.1 but no further training.
- **Task Training**: In addition to the basic briefing, subjects receive a 60 minute video lecture on how to analyze demands and how to determine optimal order quantities. The video states that often there exists a tendency of people to order towards the mean demand and explains why that is wrong. It then explains how an optimal order quantity can be determined in settings such as the one of the experiment, where the optimal order quantity is simply the profit margin multiplied by the maximum demand. After the video, subjects are asked to determine the optimal order quantity for an example (Figure 4). The answers are collected. The video then continues and provides the correct answer. In the experiment, all subjects answered the question correctly. The training is provided before

the game, immediately before the basic information is presented.[2] All slides shown in the video were available during the video and during the experiment. We decided to use a video as opposed to presenting the slides to ensure consistency of the information provided to the subjects.

Figure 4. Question asked after task training

The selling price is 10 €/unit. The unit cost is 4 €/unit. All demands between 11 and 20 have a probability of 10 %. Demands between periods are independent.

a) How high is the profit margin?

b) How high is the optimal order quantity?

A key distinction between task training and the provision of expected profit information in Phase 3 of the experiment is that training explains the rationale behind the optimal order quantity calculation and differentiates it from the common anchoring mistake that people are prone to.

2.3 Laboratory protocol

The experiment was conducted in Germany. A total of 164 people participated. Each subject participated in exactly one session. A breakdown by treatment appears in Table 3. Cash was the only incentive offered. Subjects were recruited from two sources. Experienced managers were recruited from companies associated with the Department of Supply Chain Management of the University of Cologne. These sessions were conducted on site at the companies. Students were recruited from the subject pool of the Cologne Laboratory for Economic Research (CLER) with the help of the recruitment software ORSEE (Greiner 2004). These sessions were conducted at the University of Cologne.

The game software was created by the Laboratory for Economic Management and Auctions (LEMA) at the Smeal College of Business, Penn State University. Snapshots of a typical newsvendor computer screen appear in Appendix A.2. The screens displayed information about the outcomes in prior periods of the game, including demand realization, the order placed, and the resulting profit, as well as the current total profit accumulated since the start of the session. The experiment's software was web-based php application with a mySQL back end.

At the end of the session, subjects were paid their total game earnings (includes all 100 decisions) at a rate of 1 Euro per 44 taler for Freshmen and Graduates and 1 Euro per 11 taler for

[2] The video is available at http://www.wiso.uni-koeln.de/scmms/multimedialecture.zip.

Managers. Freshmen and Graduates received an additional 10 Euros if they received task training. This additional payment was not offered to Managers, because it would have resulted in total payments that would have been taxable under German law. Sessions lasted between 50 and 60 minutes, both for students and managers, plus 60 minutes in those sessions with training. Students received a participation fee of 2.50 Euros and actual average earnings were about 10 Euros for students without training, 20 Euros for students who received training and 27 Euros for Managers.[3]

3 Results

We organize the data analysis along the lines of the three broad research questions laid out in the Introduction. We refer to the order that maximizes expected profits as the **optimal order quantity** (= 75 units). The analysis focuses on two measures of newsvendor performance. The first is **order quantity** (or just 'order'). The second is **efficiency**, defined as 'the average profit that would have been achieved with the order quantity chosen by the subject' divided by 'the average profit achieved with an optimal order quantity'. Efficiency is an interesting additional performance indicator because it penalizes for order variability; for example, ordering 75 every period is a more efficient strategy than alternating orders of 100 and of 50.

The unit of analysis for all of our inferential statistical tests will be the individual subject. Table 3 provides a breakout of the number of subjects in each treatment. Table 3 also displays some summary statistics for the data. We will refer back to this table in the inferential analysis. For now, note the treatment nomenclature and subject sample sizes. The table also provides a 'thirty thousand feet' view of the data (the means in the table aggregate over all phases of the game). Even at this distance, one can see shifts with regard to experience and training.

[3] 10 Euro were about 14 US$ at the time of the experiments. Typically we target the size of payment to cover the average opportunity costs of the subjects. 10 Euro is probably an upper limit of the hourly costs for our student population. The opportunity cost for the managers is likely higher (perhaps not in all cases since the experiments were conducted during normal work hours). Given this (and tax law constraints), we increased manager payments by a factor of 4. We note that we saw no evidence of insufficient manager incentives. For instance, managers took as much time to conclude the experiments as students, performance both relative to the student population and across managers depended systematically on the information scenario (as we describe below), and the discussions in the debriefing phase with managers were characterized by a high degree of involvement, ambition and interest.

Table 3. Summary statistics, by treatment (n = number of subjects)

Treatment	n	Mean Order	Mean Efficiency
Freshmen Basic (FB)	32	54	0.854
Freshmen Trained (FT)	22	65	0.932
Managers Basic (MB)	26	56	0.865
Managers Trained (MT)	23	62	0.905
Grads Basic (GB)	31	57	0.883
Grads Trained (GT)	30	70	0.964

3.1 What experience leads to the best use of historical information? Basic treatments.

As discussed in the Introduction, previous research finds that student newsvendors tend to anchor their orders away from the optimal order towards average demand. Our focus in this subsection will be on the extent to which this pattern of bias extends to procurement managers in the basic treatments of the experiment.

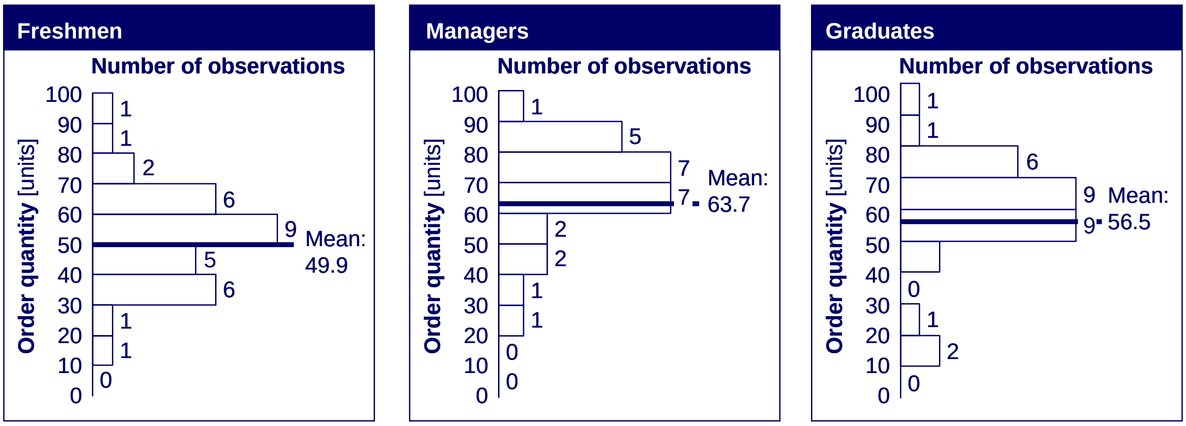
Observation 1. While all groups exhibit an anchoring bias in Period 1, Managers bias is significantly less than that of Freshmen or Graduates. Averaged over all of Phase 1, however, there is no significant difference in order quantities across groups.

In Period 1, the only information subjects have to base a decision on is a description of the game (including economic parameters) and historical information on previous demand draws; they do not yet have any experience or feedback resulting from their own decisions. Figure 5 displays the Period 1 decisions. All three groups exhibit the same qualitative pattern of anchoring bias with the mean order below the optimal order quantity in period 1 ordering; in all three cases, a one sample Wilcoxon test comparing with a fixed order quantity of 75 is highly significant (two tailed $p < 0.002$). The mean initial order for Freshmen, however, is almost exactly the average demand of 50, while for Graduates it is weakly significantly higher (two-tailed Wilcoxon $p = 0.093$). The average initial order for Managers is strongly significantly higher than for students (two-tailed Wilcoxon $p = 0.049$ compared with Graduates, $p = 0.002$ compared with Freshmen). Inferential tests on the efficiency of initial orders yield comparable results. So in Period 1, Managers outperform the students.

That said, the performance difference is short lived. Already in period 2, the average order of Managers is below that of either student group. And averaged over all periods in Phase 1:

Managers order an average of 53.0 compared to 51.9 for Graduates and 49.2 for Freshmen. These differences are not significant at any standard level (the same is true for efficiency comparisons). These results might indicate that the Period 1 ordering difference is not robust. But they are also consistent with the search behavior that feedback on demand draws triggers; see Observation 5 in Subsection 3.2, here we note that the Period 1 demand draw was below average, 44, and average Manager orders moved lower in Period 2 (visible in Figure 6).

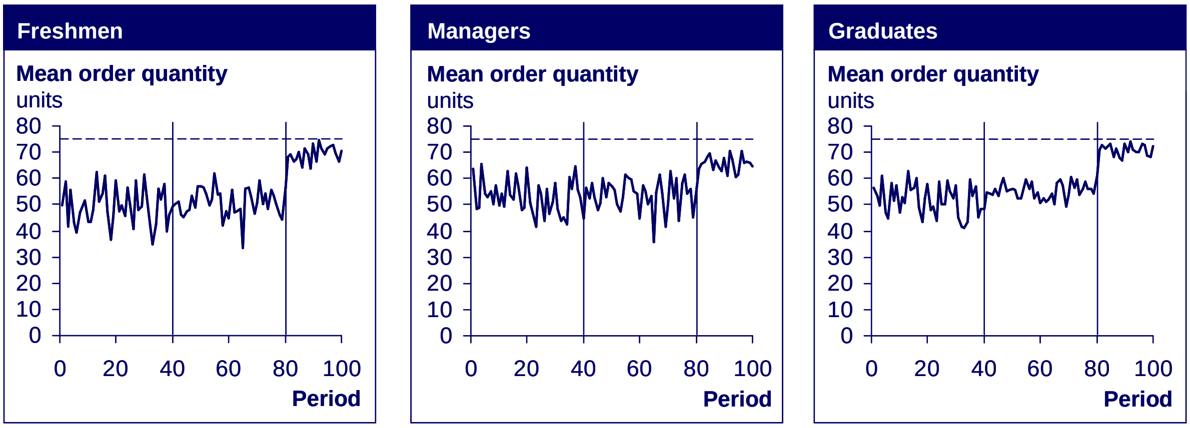
Figure 5. Distribution of Period 1 order quantities, basic treatments



Observation 2. Within phases, none of the three subject groups show improvement, that is, the period has no significant effect on order quantities or efficiency.

A second finding of previous newsvendor studies is that learning-by-doing, as measured by changes in average orders towards the optimal order quantity, is markedly slow (see Introduction). This is true for all three groups sampled in the present study, as can be observed in Figure 6, and within all three information phases.

Figure 6. Mean order quantities by period, basic treatments



We ran regression analyses to quantify the effect of learning by doing on mean orders ($\overline{O_t}$) and mean efficiencies ($\overline{E_t}$) per period (Table 4). We denote the time period by t (t = 1, 2, …, 100). To control for subjects we ran a fixed effects model using indicator variables S_i for each subject i. For the phases, we use indicator variables $Phase_j$ for phase j (j = 2, 3). The regressions confirm that no group shows any particular learning-by-doing proclivity; the only significant, non-negligible coefficient for t appears in the regression of the mean orders of Managers and this coefficient is small and negative, indicating a modest tendency for Manager orders to move away from the optimal order quantity as the number of repetitions increases. We conclude that experience gained by repetition is of no significant value.

Table 4. Effect learning by doing on mean orders and mean efficiencies, basic treatments: Fixed effect regression [two-tailed *p*-value]

				R^2_{Adj}
$\overline{O}_t^{FB} = \sum_{i=1}^{32} d_i S_i$	$+\ 0.011 \cdot t$	$+\ 1.107 \cdot Phase_2$	$+\ \mathbf{19.35} \cdot Phase_3$	0.309
	[0.663]	[0.367]	[0.000]	
$\overline{O}_t^{MB} = \sum_{i=1}^{26} d_i S_i$	$-\ \mathbf{0.068} \cdot t$	$+\ \mathbf{3.617} \cdot Phase_2$	$+\ \mathbf{17.41} \cdot Phase_3$	0.206
	[0.034]	[0.016]	[0.000]	
$\overline{O}_t^{GB} = \sum_{i=1}^{31} d_i S_i$	$-\ 0.038 \cdot t$	$+\ \mathbf{5.043} \cdot Phase_2$	$+\ \mathbf{21.67} \cdot Phase_3$	0.364
	[0.120]	[0.000]	[0.000]	
$\overline{E}_t^{FB} = \sum_{i=1}^{32} d_i S_i$	$+\ 0.000 \cdot t$	$+\ \mathbf{0.029} \cdot Phase_2$	$+\ \mathbf{0.139} \cdot Phase_3$	0.238
	[0.409]	[0.010]	[0.000]	
$\overline{E}_t^{MB} = \sum_{i=1}^{26} d_i S_i$	$+\ 0.000 \cdot t$	$+\ 0.027 \cdot Phase_2$	$+\ \mathbf{0.111} \cdot Phase_3$	0.160
	[0.870]	[0.051]	[0.000]	
$\overline{E}_t^{GB} = \sum_{i=1}^{31} d_i S_i$	$+\ \mathbf{0.000} \cdot t$	$+\ \mathbf{0.075} \cdot Phase_2$	$+\ \mathbf{0.169} \cdot Phase_3$	0.302
	[0.046]	[0.000]	[0.000]	

To summarize the observations in this subsection, Managers largely exhibit the same qualitative pattern of ordering bias as do students: They anchor between the optimal order and average demand and they exhibit little longitudinal adjustment in this behavior. However, Managers outperform students in the opening period of play, with only historical information for guidance and prior to receiving any feedback on their own performance.

3.2 What experience leads to the best use of analytical information?

The information provided in Phase 2 (demand distribution) is sufficient to identify the optimal order quantity, and the information provided in Phase 3 (expected profit) may further support decision making in transforming demand information into the directly relevant expected profit information. In this subsection we focus on how well this information is, in fact, used and which group uses it best in the basic treatments of the experiment.

Figure 7. Average order and efficiency, by treatment and by phase, basic treatments.

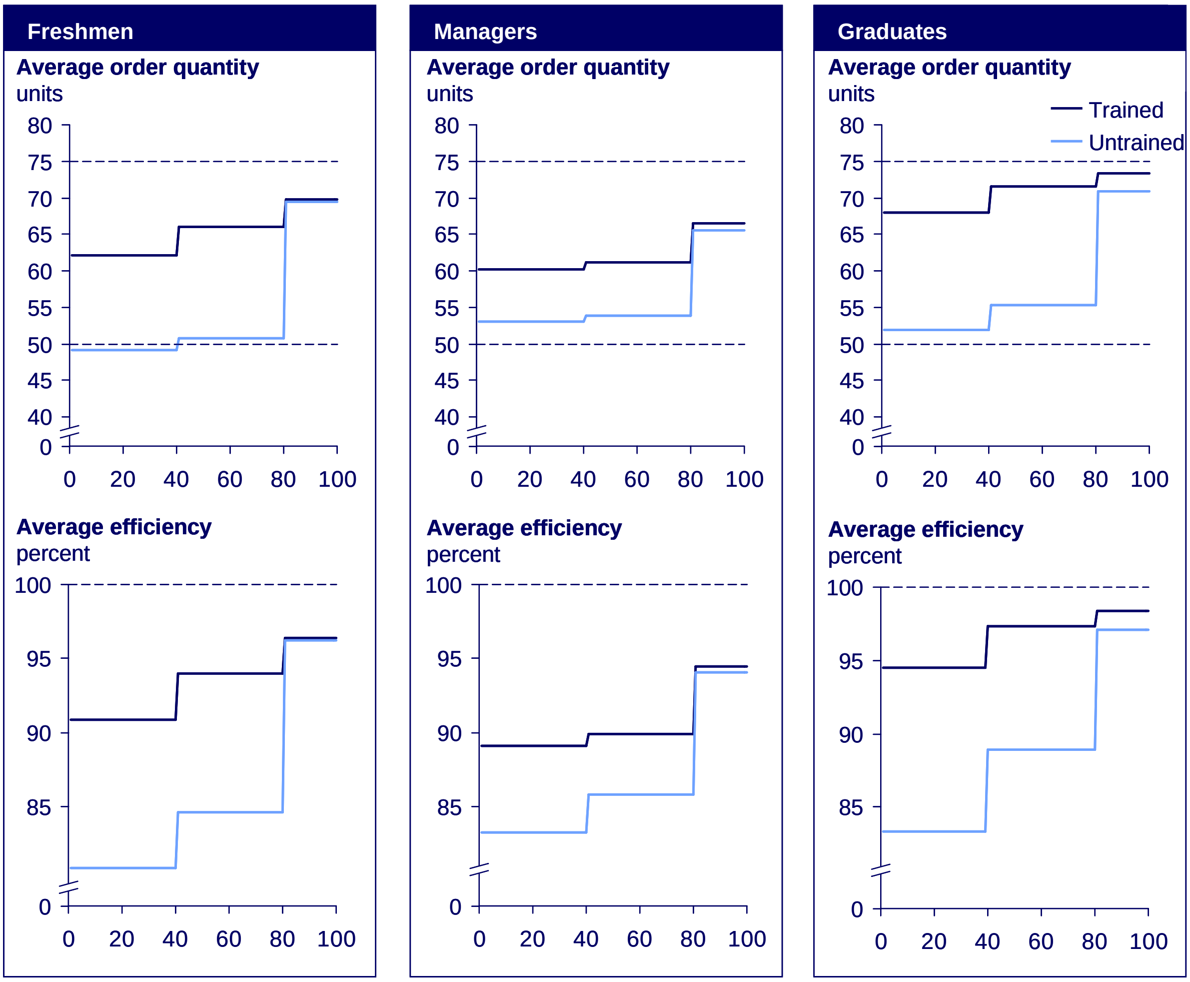


Table 5. Changes in mean orders and mean efficiencies between Phases 1 and 2 and between Phases 2 and 3 [Wilcoxon 2-tailed p-values], basic treatments

	Demand Information		Expected profit information	
	$\Delta\overline{O}_{1,2}$	$\Delta\overline{E}_{1,2}$	$\Delta\overline{O}_{2,3}$	$\Delta\overline{E}_{2,3}$
Freshmen	1.6	**0.037**	**18.6**	**0.116**
	[0.210]	[0.004]	[0.000]	[0.000]
Managers	0.8	0.025	**11.7**	**0.083**
	[0.861]	[0.269]	[0.000]	[0.000]
Graduates	**3.5**	**0.056**	**15.5**	**0.080**
	[0.021]	[0.000]	[0.000]	[0.000]

Observation 3: While Managers do not much profit from demand information, students, especially Graduates, increase mean order quantities and efficiencies.

From Figure 7, the introduction of demand information moves average order quantities modestly towards the optimal order quantity. Table 5 shows that these movements are highest for Graduates, less than half the amount for Freshmen, and less than a quarter for Managers.[4] Moreover, the same table shows that, for both student groups, efficiency rises significantly with the addition of demand information, while there is a smaller and insignificant effect for Managers. Increases in efficiency are due to both modest increases in mean orders and decreases in search (as shown below, Observation 5).

Observation 4: Expected profit information strongly improves mean order quantities and efficiencies for all subject groups, yet again more so for students.

Returning to Figure 7, we see that the introduction of expected profit information in addition to demand information improves performance for all three groups. Table 5 shows that the effect is substantially larger than the effect of demand information both in terms of order quantities and efficiencies, and for all three groups. Applying two-tailed Wilcoxon tests, students generally gain more from adding profit information than Managers (p = 0.005 and 0.071 for order resp. efficiency improvement of Managers compared to Freshmen, and p = 0.081 and 0.665 compared to Graduates). Because, theoretically, profit information can be derived from the demand information, the Phase 3 effects also indicate that some subjects have difficulty transforming demand information into the relevant profit information by themselves. That said, the next observation shows that the addition of analytical information does at least partially affect the approach that all groups take to the data:

Observation 5: The amount of search behavior falls as information is added in Phases 2 and 3, consistent with the new information inducing a more analytical approach to ordering. The reduction in search behavior is largest for Graduates and smallest for Managers.

Most newsvendor learning-by-doing patterns involve search processes characterized by a substantial variability within a subject's orders across time (as opposed to optimal ordering behavior which proscribes a constant order; see Bolton and Katok 2008). Table 6 presents an

[4] Only the Graduates move is significant in Table 6. There is one difference in statistical significance if one takes into account possible repetition effects as shown in Table 4: Here, the Managers' move also becomes significant, indicating that phase information may 'compensate' for moving away from the optimal order within phases.

analysis of how search behavior changes with the information phase. The regressions take as the dependent variable the standard deviation in orders calculated over 4-period blocks for each subject. Standard deviation is a proxy measure for the amount of individual search. The phase coefficients measure the reduction in search induced by the addition of new information. Observe from the table that the reductions are all strongly significant.

Table 6. Changes in search behavior across phases, basic treatments:
Regression results and adjusted R^2

$$D_k^{FB} = \sum_{i=1}^{32} d_i S_i + \mathbf{0.040} \cdot k - \mathbf{7.11} \cdot Phase_2 - \mathbf{11.97} \cdot Phase_3 \qquad 0.224$$

[0.079] [0.000] [0.000]

$$D_k^{MB} = \sum_{i=1}^{26} d_i S_i - 0.014 \cdot k - \mathbf{5.21} \cdot Phase_2 - \mathbf{7.37} \cdot Phase_3 \qquad 0.154$$

[0.649] [0.001] [0.000]

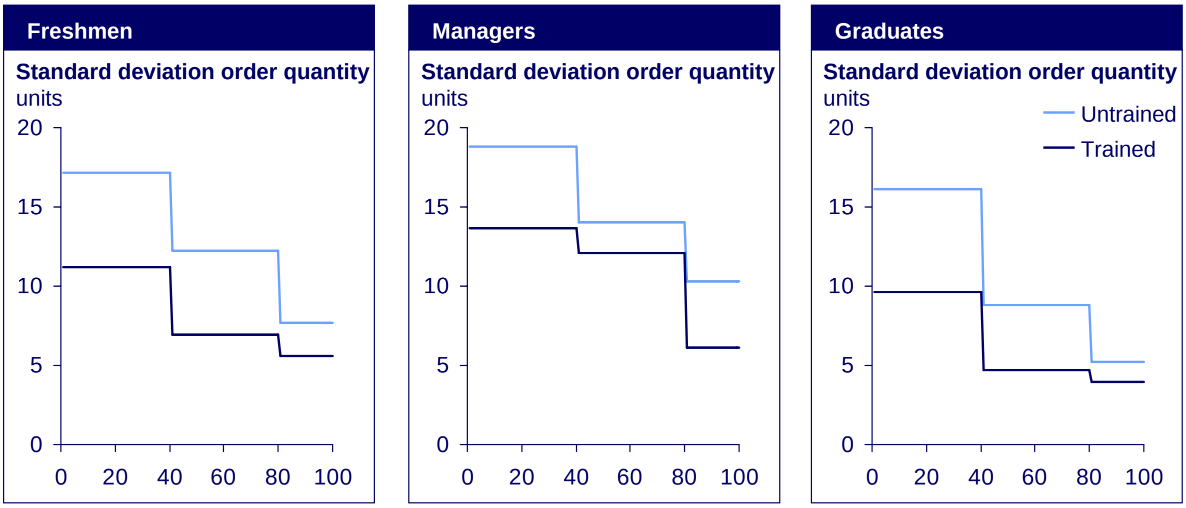
$$D_k^{GB} = \sum_{i=1}^{31} d_i S_i + \mathbf{0.044} \cdot k - \mathbf{9.13} \cdot Phase_2 - \mathbf{12.88} \cdot Phase_3 \qquad 0.284$$

[0.056] [0.000] [0.000]

where k = 4, 8, 16,…, 100 and D_k is the standard deviation of a subject i calculated over periods k-3, k-2, k-1, k.

The regression results in Table 6 also suggest that reductions in search behavior are larger for Graduates than for Managers, with Freshmen in between. To check this further, we ran Wilcoxon tests comparing individual standard deviation across groups; For Phase 1, there is no significant difference in standard deviations across groups (two-tailed $p > 0.39$ for all comparisons). From Phase 1 to Phase 2 the within subject drop in standard deviation is greater for Graduates than for Managers or Freshmen (two-tailed $p = 0.006$ and 0.063). From Phase 1 to Phase 3, the within subject drop in standard deviation for Graduate is greater than that for Managers (two-tailed $p = 0.012$). Other comparisons are not significant at any standard level. Figure 8 confirms that the average standard deviation decreases as information is added in Phases 2 and 3. In sum, information about the demand distribution helps students to reduce (inefficient) search in the face of fluctuating demand. Managers, however, are more like to continue relying on historical demand information and do not gain to the same extent from the analytical information as do students.

Figure 8. Mean standard deviation of order quantities for each phase, basic treatments



Observation 6. Students and in particular Graduates make the best overall use of analytical information. As a result, Graduates outperform Managers in Phase 3.

All three groups gain from the provision of analytical information, whether it be demand information or, additionally, expected profit information. However, as shown in Figure 7, students respond more strongly to both demand and profit information. In particular, when aggregating over the information effects, the improvements of Freshmen and Graduates from Phase 1 to Phase 3 regarding order quantities are significantly larger than the corresponding improvements of Managers (two-tailed Wilcoxon test, p = 0.017 for Freshman and p = 0.020 for Graduates; the corresponding values for efficiencies are p = 0.033 for Freshmen and p = 0.182 for Graduates).

As a result, while Managers initially perform better, both student groups perform better in Phase 3. Applying a two tailed Wilcoxon test, this is significant for Graduates (p = 0.016 for order quantities and p = 0.010 for efficiencies), but not significant for Freshmen (p = 0.164 for order quantities and p = 0.180 for efficiencies). It is important to note, however, that for all three groups, even in Phase 3, order quantities still average significantly less than 75 (Freshmen p = 0.002, Managers p = 0.000, Graduates p = 0.021, all two-tailed.)

3.3 What experience leads to the best use of task training?

Task training explains both the rationale for picking and method of computing the expected profit maximizing order quantity. The observation in this subsection concerns the influence of training on each of the experience groups. With training, as without, there is no significant learning-by-doing effect. Regressions similar to those presented in Table 4 exhibit an insignificant period coefficient ($p > 0.27$ in all cases).

Table 7. Differences in mean orders and mean efficiencies between basic (B) and trained subjects (T) [two-sided p-values of Wicoxon test]

	Phase 1		**Phase 2**		**Phase 3**	
	$\Delta\overline{O}_{B,T}$	$\Delta\overline{E}_{B,T}$	$\Delta\overline{O}_{B,T}$	$\Delta\overline{E}_{B,T}$	$\Delta\overline{O}_{B,T}$	$\Delta\overline{E}_{B,T}$
Freshmen	**12.85**	**0.100**	**15.22**	**0.094**	0.30	0.002
	[0.000]	[0.000]	[0.000]	[0.000]	[0.329]	[0.197]
Managers	**7.25**	**0.058**	**7.29**	**0.041**	0.99	0.004
	[0.064]	[0.024]	[0.032]	[0.070]	[0.547]	[0.147]
Graduates	**16.10**	**0.112**	**16.1**	**0.084**	2.24	0.014
	[0.000]	[0.000]	[0.000]	[0.000]	[0.573]	[0.194]

Observation 7. During Phases 1 and 2, training improves performance significantly for all three groups. Training has little effect on Phase 3 performance for any group. As before, the effects are largest for students, and in particular for Graduates.

Observe from Figure 7 and Table 7 that, for Phases 1 and 2, all three groups significantly improve on both performance measures. In Phase 3, training does not significantly improve performance beyond what expected profit information does with untrained subjects. Still, all subject groups reached their respective maximal performance with maximal information, namely in Phase 3 and with previous training. For instance, trained Graduates in Phase 3 perform almost optimally with average orders of 73.3 units and an average efficiency of 98.4 %. The average trained Manager order in Phase 3 is 66.6 units with an average efficiency of 94.4 %. While Phase 3-performance of Graduates and Managers turn out not to be statistically different (two-tailed Wilcoxon test $p = 0.124$ for order and $p = 0.164$ for efficiencies), overall performance across all three phases is ($p = 0.004$ for orders and $p = 0.002$ for efficiency).

Demand and expected profit information have qualitatively similar effects on performance of trained and untrained subjects. However, since training already provides most of the relevant information, orders and efficiencies levels are much higher from the very beginning. As a result, the room for improvement is more limited for trained subjects, and so improvements from demand and expected profit information are only sporadically significant.[5]

4 Summary and discussion

We observe a pattern of anchoring bias in ordering behavior in all three groups we study. Experienced managers as well as master-level graduate students who have done coursework in operations management and freshmen undergraduate students who have not, tend to make orders that are biased toward average demand away from the expected profit maximizing order quantity. Within information phases, we observe little adjustment in these orders from period-to-period. As analytical information about the demand distribution and expected profits are introduced, orders adjust towards the expected profit maximizing quantity, although not all the way, and the search behavior associated with learning-by-doing diminishes somewhat.

We find that out-of-task experience affects the way information is (or is not) utilized. Managerial experience was most valuable in the very beginning of the experiment, when the only information available about the underlying stochastic model is historical demand draws. In Period 1 of the game, managers exhibit somewhat less anchoring bias than either student group although this advantage is short-lived. Managers, like the students, engage in search behavior dependent on the individual performance feedback they get as the task is repeated. With the introduction of information identifying the underlying demand distribution or additionally showing the expected profit associated with individual order quantities, the student groups catch up and the Graduates eventually outperform Managers for the basic treatments, without training. In all comparisons with the student groups and for both, demand and profit information, Graduates best utilized analytical information and Managers gained least from the information with respect to both order quantities and efficiencies. The experience that had the biggest effect on ordering behavior in our experiment was task training. It lifted the performance of all three groups in all three phases of the experiment. Although again, Graduates profited significantly

[5] Graduates gain significantly more than Managers from expected profit information both in terms of orders and efficiency. Because the impact of training on search behavior is already partly captured by our efficiency measure and, in fact, analogous to the effect on efficiency, we drop the corresponding results here.

more than Managers, the latter moving least towards optimality and efficiency in response to training.

One major implication of our findings has to do with the pros and cons of particular kinds of experience for solving procurement problems. That managers initially exhibit less bias than students (with only historical data on demand available for guidance) suggests that work experience leads to an intuitive feel for the solution to the newsvendor problem. That graduate students with an operations background are better at leveraging analytical information about the stochastic process behind the newsvendor problem than managers or freshmen undergraduates with no operations background suggests that classroom education provides important insight into the process behind the newsvendor solution. Consistent with this observation, training has a big effect on performance, particularly when it is coupled with an operations management background as in the case of the graduate students. However, the fact that Graduates perform better with the addition of training than without suggests that time lags play an important role in the effectiveness of classroom education. And the fact that theoretically redundant expected profit information strongly significantly improves performance across all subject pools suggests that overcoming the computational problems involved in the newsvendor problem is a challenge even with education and training. This observation also points to the potential value of decision support systems that take the cognitive limitations of decision makers into account.

We mentioned in the Introduction that previous studies comparing managers and students had found that the framing of the experiment was critical to performance. While our experiment is model-based in the sense that we took the incumbent model of procurement as our guide for what the important elements of the newsvendor problem are (and what the solution is), it was couched in natural terms; specifically, in terms of product stocking, uncertain demand for that product and payoffs determined by costs and revenues. Yet managers performed less than optimally. This might indicate that inventory procurement in the field is being conducted suboptimally. On the other hand, it is unlikely that the parameters of the problem exactly matched the terms that our managers were familiar with from their field experience. So it might be that managers are conducting field procurement optimally for the special inventory problem they face, but do not understand enough about the more general underlying process to successfully translate their success to other parameterizations of the problem. It might also be that some inventory managers in the field are not motivated by a concern of profit maximization, as assumed for the newsvendor solution, but are rather guided by aspirations or goals on, say, out-of-stock quotas (see Brown and Tang 2006). (Under none of these hypotheses would there be a reason to believe that Managers would have an advantage in finding the optimal, profit maximizing order of our newsvendor problem.)

Table 8. Effect of years of working experience and position on Managers' mean orders and mean efficiencies: Regression [two-tailed *p*-value]

	R^2_{Adj}
$\overline{O}_i^{MB} = 49.35 + 0.883 \cdot Phase_2 + \mathbf{12.62} \cdot Phase_3 - \mathbf{0.424} \cdot Years + \mathbf{2.437} \cdot Position$ [0.725] [0.000] [0.029] [0.020]	0.325
$\overline{O}_i^{MT} = 61.78 + 0.926 \cdot Phase_2 + 6.365 \cdot Phase_3 - \mathbf{1.059} \cdot Years + \mathbf{5.588} \cdot Position$ [0.780] [0.058] [0.000] [0.016]	0.205
$\overline{E}_i^{MB} = 0.811 + 0.025 \cdot Phase_2 + \mathbf{0.108} \cdot Phase_3 - \mathbf{0.004} \cdot Years + \mathbf{0.019} \cdot Position$ [0.188] [0.000] [0.005] [0.017]	0.368
$\overline{E}_i^{MT} = 0.900 + 0.008 \cdot Phase_2 + \mathbf{0.054} \cdot Phase_3 - \mathbf{0.008} \cdot Years + \mathbf{0.046} \cdot Position$ [0.731] [0.032] [0.000] [0.008]	0.234

Position = 1 if simple employee (5 observations in the basic treatment and 18 observations in the trained treatment), = 2 if team manager (6/1), = 3 if head of department (10/4), = 4 if board member (5/0). The Pearson correlations between position and years are 0.134 (p = 0.513) in the basic treatment and 0.627 (p = 0.001) in the trained treatment.

While an altogether new investigation would need to be designed to separate these hypotheses, the regressions in Table 8 suggest that looking for years of working experience on the job and position in the firm is a promising starting point. In particular, our data indicates that for Managers, more *Years* of experience significantly *reduces* performance, while higher *Position* significantly *improves* performance as measured by deviation from the optimal order quantity and efficiency. That is, more years of work experience appear to make managers more resistant to utilizing the kind of analytical information needed in the newsvendor problem (controlling for position). This suggests that the standard formulation of the newsvendor problem, with its sole focus on profit maximization calculus, misses important aspects of everyday routine inventory decisions. Yet at the same time analytical skills related to the newsvendor model appear to be helpful in inventory careers (controlling for number of years with work experience), probably suggesting that the standard approach is better suited to the larger, strategic picture of inventory decision making.[6] It would be interesting to extend experimental and theoretical research to an environment that includes more ‘artifactual’ features in real world procurement to see if these features are truly superfluous to the problem or if they aggravate or possibly mitigate the anchoring bias we observe (ex., Schulz et al., 2007).

References

Arrow, K. J., T. Harris, J. Marschak (1951). Optimal inventory policy. *Econometrica* 19(3) 250-272.

Benzion, U., Y. Cohen, R. Peled, and T. Shavit (2008). Decision-making and the newsvendor problem - an experimental study. *Journal of the Operational Research Society*, 59, 1281-1287.

Bolton, G. E. and E. Katok (2008). Learning-by-doing in the newsvendor problem: A laboratory investigation of the role of the experience. *Manufacturing and Service Operations Management*, 10, 519-538.

Bostian, A., C. Holt, and A. Smith (forthcoming). The newsvendor "pull-to-center effect:" Adaptive learning in a laboratory experiment. *Manufacturing and Service Operations Management*.

Brown, A. O. and C. S. Tang (2006). The impact of alternative performance measures on single-period inventory policy. *Journal of Industrial and Management Optimization*, 2(3), 297-318.

Burns, P. (1985). Experience and decision making: A comparison of students and businessmen in a simulated progressive auction. In V. Smith, ed., *Research in Experimental Economics Vol. 3*, Greenwich: JAI Press, 139–157.

Cooper, D., J. Kagel, W. Lo, and Q. Gu (1999). Gaming against managers in incentive systems: Experimental results with Chinese students and Chinese managers. *American Economic Review*, 89(4), 781-804.

Corbett, C. J. and J. C. Fransoo (2007). Entrepreneurs and newsvendors: Do small businesses follow the newsvendor logic when making inventory decisions? Working paper, UCLA.

Croson, D., R. Croson, and Y. Ren (2008). How to manage an overconfident newsvendor. Working paper, University of Texas at Dallas.

[6] The regression in Table 8 do not fully control for the selection effect inherent in less skilled managers being less likely to be promoted. But the extent to which this is true strengthens our interpretation that promotion is positively correlated with analytical skills.

Greiner, B. (2004). The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics. University of Cologne, Working Paper Series in Economics 10.

Kremer, M., and S. Minner (2008). The human element in inventory decision making under uncertainty: A review of experimental evidence in the newsvendor model, *Zeitschrift für Betriebswirtschaftslehre*, 4, 83-97.

Kremer, M., S. Minner, and L. N. Van Wassenhove (2007). Anchoring and regret in the newsvendor problem - the impact of task complexity and framing. Working paper, University of Mannheim.

Lurie, N. H. and J. M. Swaminathan (2007). Is timely information always better? The effect of feedback frequency on performance and knowledge acquisition. Working paper, Georgia Institute of Technology.

Moritz, B. (2008), Behavior, cognition and individual differences in the newsvendor problem, presentation at the 3rd Annual Behavioral Operations Conference, July 23-25, 2008, Edmonton, Canada.

Schultz, K. L., J. O. McClain, L. W. Robinson, and J. Thomas (2007). The use of framing in inventory decisions. Working paper, Cornell University.

Schweitzer, M. E. and G. P. Cachon (2000). Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science 46* (3), 404-420.

Siegel, S. and D. L. Harnett (1964). Bargaining behavior: A comparison between mature industrial personnel and college students. *Operations Research*, 12(2), 334-343.

Su, X. (forthcoming). Bounded rationality in newsvendor models. *Manufacturing and Service Operations Management*.

Appendix A. Materials used in the experiment

A.1 Instructions given to subjects prior to the play of the newsvendor game (translated from German).

Briefing

You are a retailer who buys and sells a single generic product. In each period of the game you will order the product from an external supplier at a purchase price of

Purchase Price = 3 TALER/unit.

You will sell the product to customers at a sales price of

Sales Price = 12 TALER/unit.

You will play 100 periods with identical activities:

1. Determination of purchase order quantity: At the beginning of each period you determine the order quantity. You can choose the order quantity freely. When you place the order, you do not know the customer demand of the period.

2. Generation of customer demand: The computer generates a customer demand. The graph on the following page shows the demands of the 50 previous periods.

3. Calculation of profit: The demand is filled and the profit of the period is computed. There are two different cases:

> a. Customer demand is less than or equal to order quantity: All customer demand can be filled. The profit is
>
> Profit = Sales Price * Customer Demand – Purchase Price * Order Quantity
>
> Profit = 12 TALER/unit * Customer Demand – 3 TALER/unit * Order Quantity
>
> Example: If the customer demand is 60 units and you ordered 80 units, each demand can be filled and the profit is
>
> Profit = 12 TALER/unit * 60 units – 3 TALER/unit * 80 units = 480 TALER
>
> b. Customer demand is greater than order quantity: Only customer demands up to the order quantity can be filled. The profit is
>
> Profit = Sales Price * Order Quantity – Purchase Price * Order Quantity
>
> Profit = 12 TALER/unit * Order Quantity – 3 TALER/unit * Order Quantity
>
> Example: If the customer is 60 units and you ordered 40 units, only 40 units can be filled and the profit is
>
> Profit = 12 TALER/unit * 40 units – 3 TALER/unit * 40 units = 360 TALER

4. Clearance of left over inventory: If your order quantity was greater than the demand, the remaining stock is disposed at no cost. In other words, remaining inventory is worthless and not carried over to following periods.

The roadmap of the game is as follows:

1. Playing of 40 periods

2. Short break and information of the player

3. Playing of another 40 periods

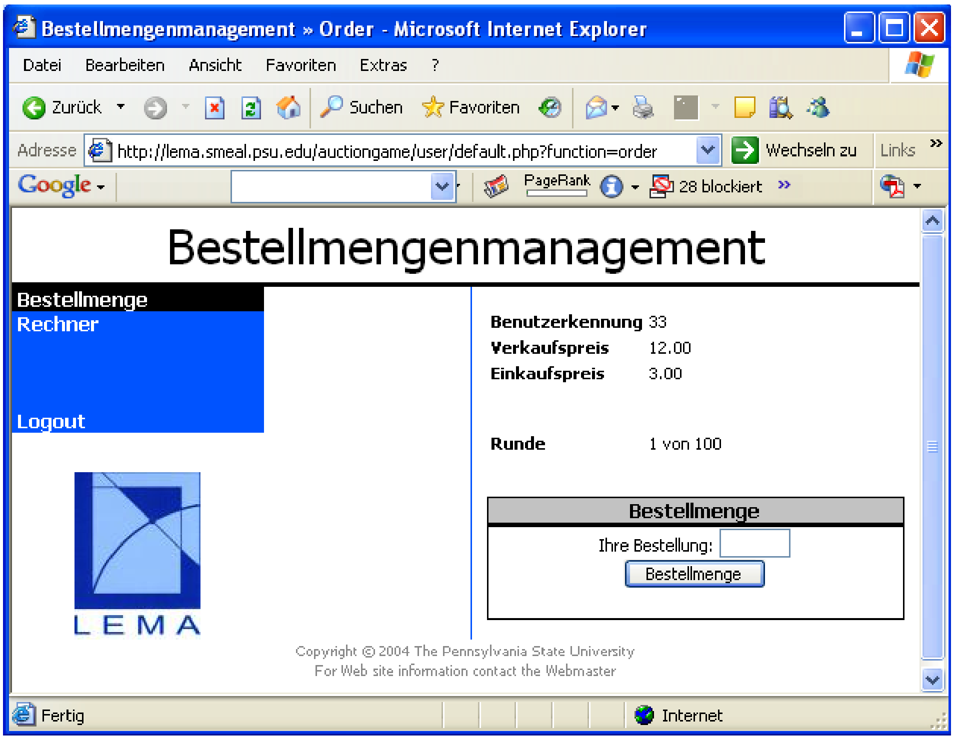
4. Short break and information of the player

5. Playing of another 20 periods

6. Calculation of average profit per period and compensation of participant: The exchange rate is 1 Euro = 11 TALER

A.2 Computer interface screen shots

Input screen



Feedback screen