Naveen Jindal School of Management

2012-2-25

Managers and Students as Newsvendors

Gary Bolton, *et al.*

MANAGEMENT SCIENCE
Vol. 58, No. 12, December 2012, pp. 2225–2233
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)
http://dx.doi.org/10.1287/mnsc.1120.1550

Managers and Students as Newsvendors

Gary E. Bolton
Jindal School of Management, University of Texas at Dallas, Richardson, Texas 75080; and
Smeal College of Business, Pennsylvania State University, University Park, Pennsylvania 16802, gbolton@utdallas.edu

Axel Ockenfels
Department of Economics, University of Cologne, D-50923 Cologne, Germany, ockenfels@uni-koeln.de

Ulrich W. Thonemann
Department of Business Administration, University of Cologne, D-50923 Cologne, Germany,
ulrich.thonemann@uni-koeln.de

We compare how experienced procurement managers and students solve the newsvendor problem. We find that managers broadly exhibit the same kind of pull-to-center bias as students do. Also, managers use information and task training no better than students. The performance of managers is positively affected by the level of their education and their level in the organizational hierarchy. We discuss implications for theory and for how ordering might be improved in practice.

Key words: newsvendor game; laboratory experiment; manager behavior; information usage
History: Received May 10, 2011; accepted February 25, 2012, by Teck Ho, decision analysis. Published online in *Articles in Advance* July 18, 2012.

1. Introduction

The newsvendor problem is a foundational model for the study of inventory control and supply chain management. 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. The optimal policy for maximizing expected profits is well documented. It stipulates that the product be stocked up to the point where marginal cost and marginal expected revenue are equal (Arrow et al. 1951). The newsvendor problem applies to products with short selling seasons relative to production and distribution time (e.g., fashion apparel and promotion items), and it is a building block for more advanced inventory models.

Recently, attention has turned to the question about the extent to which managerial judgment agrees with the newsvendor model. In their seminal paper on this topic, Schweitzer and Cachon (2000) observe in a laboratory experiment with student subjects a *pull-to-center bias*—that is, a tendency of subjects to order between the expected profit-maximizing quantity and the mean demand. The investigators go on to show that this bias cannot be explained by conventional risk preferences (Schultz et al. 2007 exhibit a similar result). The pull-to-center bias has proven to be remarkably robust among undergraduates and MBA students. The bias persists for a variety of demand distributions (Benzion et al. 2008) with substantial task repetition and even when descriptive statistics on the performance are provided (Bolton and Katok 2008); more frequent feedback can actually degrade performance (Lurie and Swaminathan 2009, Katok et al. 2008).

Various explanations have been proposed for the pull-to-center bias. Schweitzer and Cachon (2000) argue that anchoring on the mean demand and minimization of the ex post inventory error can explain the effect. Recent models rely on adaptive learning (Bostian et al. 2008), computational errors (Su 2008), the overconfidence bias (Ren and Croson 2012), and investigation of the relationship between cognitive reflection or framing and newsvendor decisions (Moritz et al. 2011, Kremer et al. 2010). Ho et al. (2010) posit that there are psychological costs associated with leftovers and stockouts. Their model nests the standard model and ex post inventory error minimization as special cases, permitting a simultaneous, structural estimate of all three models. The result suggests that the psychological cost model explains newsvendor behavior and profits better than the others.

Yet if we had managers with procurement work experience solve the newsvendor problem, would they exhibit the same pull-to-center bias the students do? This is easily the most common question we hear about in this literature. The newsvendor problem is arguably the simplest stochastic inventory problem, devoted to the essential principles of rational inventory decisions. There is nothing in the formal theory to suggest that experienced managers should be more able than students to optimally solve the newsvendor problem. Still, because students typically have little experience with procurement, they have had less opportunity to think about the problem or to learn optimizing behavior. Also, procurement managers are subject to self-selection and market selection, whereas student subjects are typically selected on a first-come, first-served basis out of a pool looking to earn some money.

Although most newsvendor-related laboratory experiments were conducted with students, there are two notable exceptions. Katok et al. (2008) analyze a service-level agreement and the effect of the review period on order decisions. They conduct experiments with 76 students and 8 managers. In one treatment, students and managers face the same problem, which allows the authors a direct comparison of the behavior of managers and students. They find that managers and students do not behave significantly differently, but state that "the small sample size in the manager treatment may account for the lack of significance" (Katok et al. 2008, p. 617). Moritz et al. (2011) conduct three laboratory newsvendor studies, two with supply chain professionals and one with students. Their focus is on analyzing how cognitive reflection affects decision making. Although students and managers behaved similarly, a close comparison is difficult because students and professionals face different parameter values.

These two studies provide an indication that managers and students behave similarly.[1] However, to the best of our knowledge, the present study is the first to use an experimental design that allows a systematic and structured comparison of manager and student decision making in the newsvendor problem. Unlike previous research, we use a subject pool of managers with extensive experience in newsvendor-type problems, such as ordering promotion items that are sold over a short period without the option to reorder during the selling season. Within the manager subject pool, we furthermore analyze how experience and position in the hierarchy affect performance; within the student subject pool, we differentiate between freshmen and graduates. Also, we analyze a variety of settings to obtain robust results: we test whether students and managers respond differently to historical demand information (the type of demand information that is available to managers in practice), to information about the demand distribution (the type of demand information usually provided in laboratory experiments), to information on the expected profit-maximizing order quantity (the type of information provided by decision support systems), and to training (the type of analytical information provided in business schools).

As we will show in the following sections, this combination of subject pools and information scenarios allows us to speak to the robustness of our main findings that (1) managers broadly exhibit the same kind of pull-to-center bias as students, (2) all subject groups use information and task training to improve their decision making (although no subject group manages to fully overcome the pull-to-center bias), and (3) managers do not utilize information or task training better than students.

2. Design of the Experiment

The experiment has six treatments. Within each treatment, subjects play the newsvendor game with phased analytical information, as explained in §2.1. Across treatments, we manipulate the experience subjects have, as well as whether they get task training, as explained in §2.2. The experimental protocol is reviewed in §2.3.

2.1. The Newsvendor Game and Provision of Information

All treatments involve a 100-period newsvendor game with the same demand and economic parameters. We chose a parameterization where expected profits are optimized at an order quantity of 75% of the maximum possible period demand, so that inventory is optimally stocked above average demand.

Table 1 provides an overview of the activities within a treatment. Subjects in all treatments received the same basic preplay briefing on the structure of the game (item 2 of Table 1): the unit purchase price is three talers (with an exchange rate of 11 talers = 1 euro), it can be sold to customers for 12 talers per unit, and unsold inventory remaining at the end of the period is "worthless." (See the subject directions in Online Appendix A.1. The online appendix is available at http://ockenfels.uni-koeln.de/bot_appendix.html.) 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 included examples to illustrate profit calculation.

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. The nature of the information to be given was not discussed prior to its introduction.

[1] A number of studies have compared subject pool performance on other kinds of tasks (see reviews by Ball and Cech 1996, Croson 2010, Fréchette 2012). A majority of the studies comparing student and professional performance find no or only small differences. However, there is some indication that games posed in a real-world context, as opposed to the more abstract language common in many experiments, can give professionals an edge (e.g., Cooper et al. 1999). Newsvendor experiments are naturally and commonly posed in a business context (e.g., "profit," "retailer," "price"). We will continue to do so here, providing a new test of the context hypothesis.

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 and Figure 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 is uncorrelated (Figure 2).
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 (Figure 3).
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.

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 (see Figure 1) and are told that demands in the experiment are generated by the same approach that was used for generating the sample demand. They are *not* told what distribution generated the draws.

Phase 2: Demand-distribution information. After 40 periods, subjects receive the demand-distribution information in Figure 2. With this information, there is sufficient information available to compute the expected profit-maximizing order quantity (Arrow et al. 1951).

Phase 3: Expected profit information. After another 40 periods (so, prior to period 81), subjects are presented with 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 yields the maximum expected profit. Although profit information is redundant—in the sense that it could already be derived from the information we provided in Phase 2—the derivation is not likely transparent to many subjects.

2.2. Treatment Manipulations: Subject Groups and Training

The experiment has a fully crossed 3×2 design for a total of six treatments. The first factor that distinguishes treatments is subject group; the second factor is task training. We used three subject groups:

• **Freshmen:** Freshmen business students who have not yet had classroom education in operations management.

• **Graduates:** Master's degree students who have taken at least one operations management course and

Figure 1 Sample Demands Given to Subjects Before They Placed Their First Order (All Treatments)

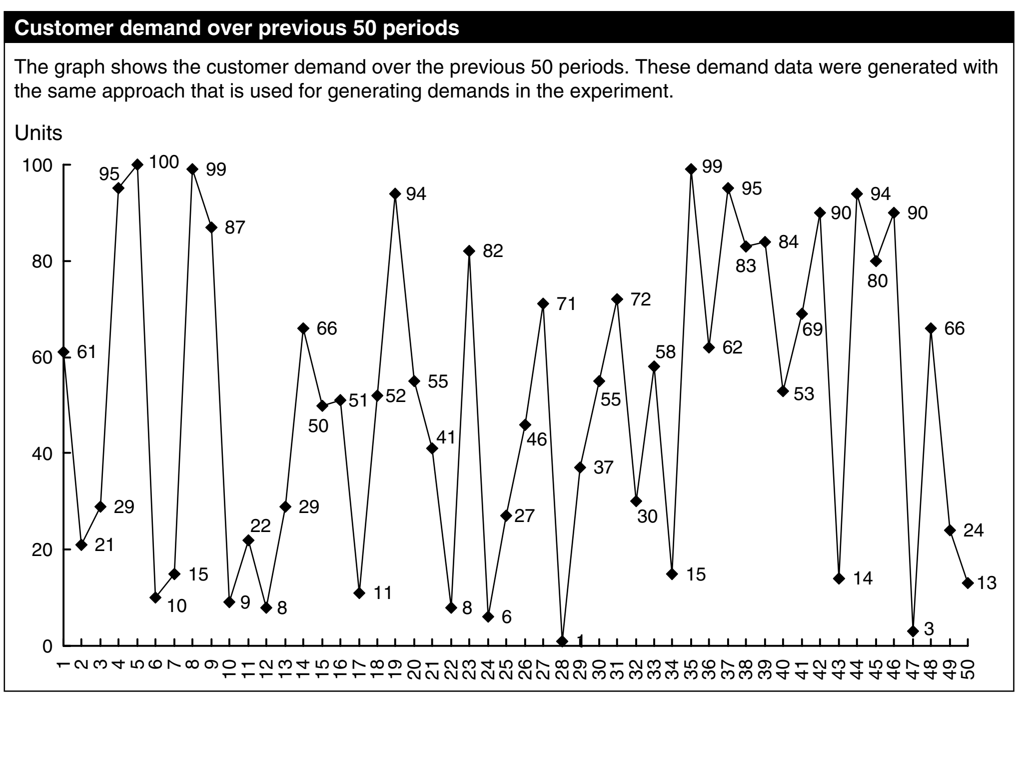


Figure 2 Information Provided to Subjects After Period 40

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.

Note. Translated from German.

have completed their undergraduate studies in business administration.

• **Managers:** Purchasing managers who have worked for at least one year making newsvendor-type decisions on a daily basis.

Freshmen and graduate students were recruited from the Faculty of Management, Economics and Social Sciences of the University of Cologne. Because students were recruited from preexperience study programs, few, if any, student subjects have newsvendor-type work experience.

Managers were recruited from companies associated with the Department of Supply Chain Management of the University of Cologne. The managers all order products that involve one-time ordering (as opposed to regular reordering) for a single selling period with no or few replenishments during the selling period. That is, as in the newsvendor problem, the products lose considerable value if unsold at the end period, and there is significant uncertainty in demand with known costs and prices at the time of decision making. Table 2 provides a breakdown by length of experience and position held at the time of the experiment.

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 §2.1 but no further training.

• **Task training:** In addition to the basic briefing, subjects receive a 60-minute video lecture immediately before the game. The lecture explains in detail the rationale behind the optimal order quantity calculation and informs the subjects that people often have a tendency to order toward the mean demand and explains why that is wrong.

2.3. Laboratory Protocol

Each subject participated in exactly one session, and cash was the only incentive offered. All subjects played without having any information about the performance of other subjects. Communication between subjects was forbidden and none was observed.

At the beginning of each period, subjects were reminded of the unit cost and unit revenue of the product. After each period, subjects were shown the demand of the period, the profit of the period, and average profit achieved so far. Online Appendix A.2 displays snapshots of a typical newsvendor computer screen as it appears both at the beginning and end of a period.

Figure 3 Expected Profit Information Provided to Subjects After Period 80

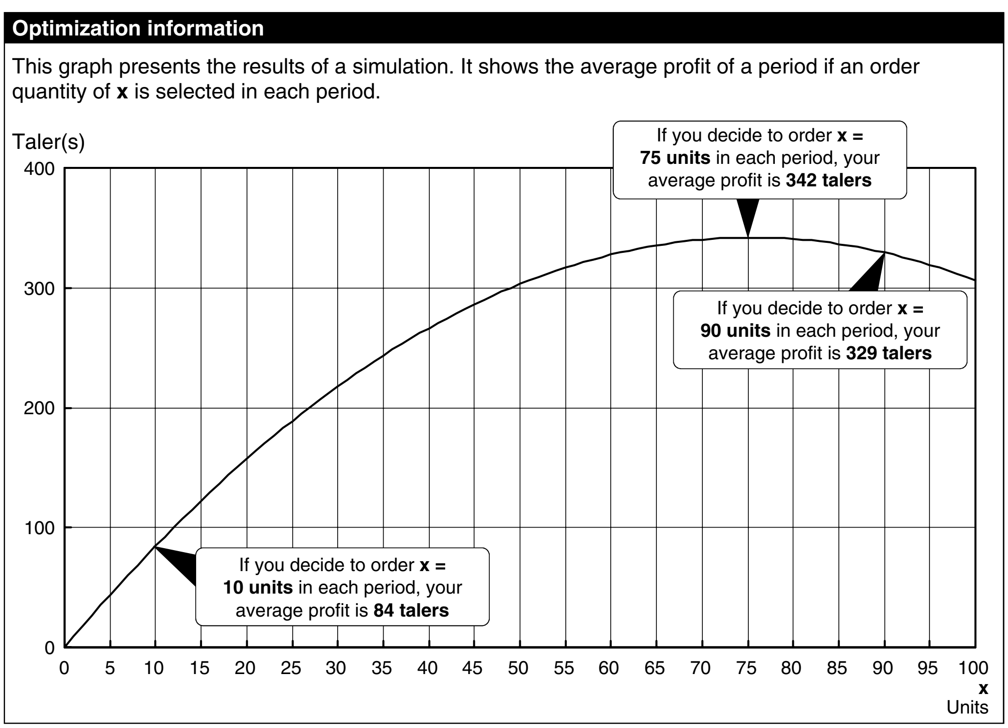


Table 2 Breakdown of Manager Experience (Number of Subjects)

Position	
• Vice president	5
• Manager	14
• Team leader	7
• Buyer	23
Work experience	
• 1–2 years	9
• 3–4 years	14
• 5–6 years	6
• 7–10 years	8
• 11–15 years	6
• 16–20 years	4
• 21–25 years	1
• 26–30 years	1

All sessions lasted about 60 minutes, plus an additional 60 minutes in those sessions with training. At the end of the session, subjects were privately paid their game earnings based on the profits achieved over all 100 decisions. Freshmen and graduates received one euro per 44 talers, and managers received one euro per 11 talers. 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. Students received a participation fee of 2.50 euros. Actual average total earnings were about 10 euros for students without training, 20 euros for students who received training, and 27 euros for managers. The payment level for students is comparable with that in other newsvendor experiments (e.g., Schweitzer and Cachon 2000, Bolton and Katok 2008).[2] After the experiment, managers were asked for some biographical data.

3. Results

In this section we compare the order behavior between managers and students and analyze how different subject pools utilize different demand information. The *null hypothesis* is that managers and students exhibit similar order patterns. Subsection 3.1 deals with the ordering behavior in the basic treatments and §3.2 with behavior under task training. In addition, §3.3 briefly shows—in line with previous research with student populations—that there is little in the way of learning-by-doing *within* the experiment phases. The unit of analysis for all of our inferential statistical tests will be the individual subject. Unless otherwise stated, all hypothesis tests are two-tailed Mann–Whitney for one sample tests or Wilcoxon for two sample tests.

3.1. Order Decisions in the Basic Treatments

Table 3 provides a breakout of the mean order quantity (aggregated over all phases of the experiment) and the number of subjects in each treatment. All three mean orders are well below the optimum of 75, and thus the hypothesis that subjects are playing optimally can be rejected ($p < 0.001$ for all subject groups). In fact, *all* 89 subjects' mean orders are below the optimum.

Phase 1: Sample demands. We now analyze ordering separately for the experiment phases. At the beginning of the experiment, we provided sample demand information (see Figure 1). This type of demand information is typically available to managers when they make order decisions. Figure 4 shows the average order quantities by phase (solid line). Looking at the data for the basic treatments reveals that all three subject groups exhibit a strong pull-to-center bias toward the mean demand (=50.5) during Phase 1. Also observe that the mean orders in Phase 1 are similar between subject groups. In fact, the entire distribution of orders is similar (see Online Appendix A.3). In Phase 1, no mean order is at or above the optimal order quantity of 75 (the maximum mean orders are 71.4, 70.4, and 70.5 for freshmen, graduates, and managers, respectively). That is, all subjects exhibit to one degree or another the pull-to-center bias.

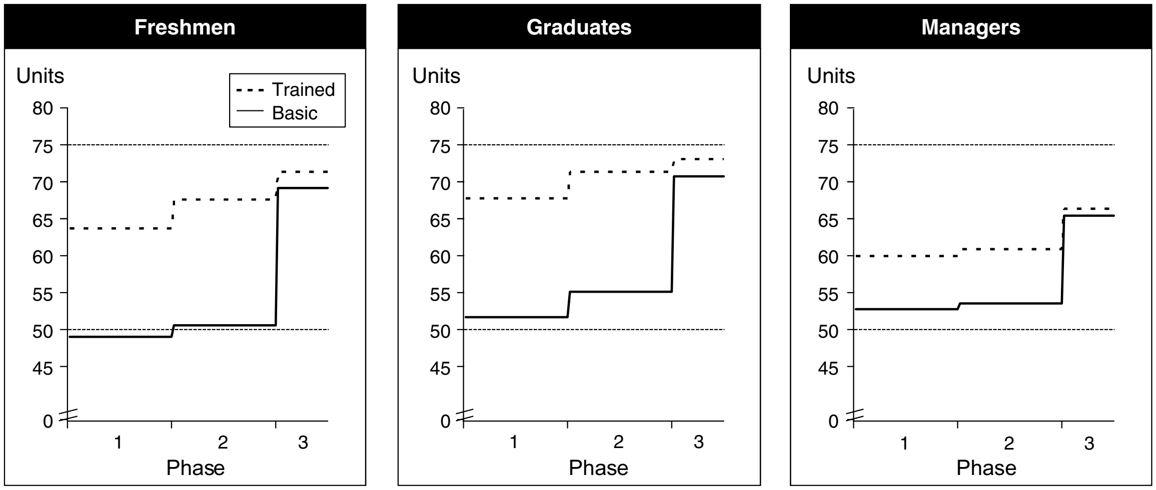
In Phase 1, managers ordered an average of 53.0 compared with 51.9 for graduates and 49.2 for freshmen; all three averages are significantly different from the optimum, 75 ($p < 0.001$). Pairwise tests find no significant differences in average orders between managers and students (managers versus graduates, $p = 0.569$; managers versus freshmen, $p = 0.106$). Thus, *under sample demand information, all three subject groups exhibit the same kind of pull-to-center bias. There is no significant difference in the average orders placed by managers and students*.

Table 3 Summary Statistics in Basic Treatments, by Subject Population

Treatment	n	Mean order
Freshmen basic (FB)	32	54
Graduates basic (GB)	31	57
Managers basic (MB)	26	56

[2] At the time of the experiment, 10 euros were equivalent to about US$14. Typically, we target the size of payment to cover the average opportunity cost of the subjects, and 10 euros is probably an upper limit for our student population. The opportunity cost for the managers is likely higher (perhaps not in all cases, because the experiments were conducted during normal work hours). Given this, and tax law constraints, we increased the exchange rate for managers by a factor of 4. We note that we saw no evidence of insufficient manager incentives. For instance, managers took as much time to conduct the experiments as students; performance both relative to the student population and across managers depended systematically on the information scenario (as we describe in §§3.1 and 3.2); and the discussions in the debriefing phase with managers were characterized by a high degree of involvement, ambition, and interest (but did not involve revelation of individual earnings or performance).

Figure 4 Average Orders by Treatment and by Phase—Basic and Trained Treatments



Phase 2: Demand-Distribution Information. At the beginning of Phase 2, we provided demand-distribution information (see Figure 2). This type of information is used in essentially all laboratory experiments on the newsvendor problem (see references in §1). Because unit cost and unit revenue are also known, demand-distribution information is sufficient to identify the optimal order quantity. However, all subject groups hardly change their order behavior after receiving demand-distribution information.

Over all periods in Phase 2, managers ordered an average of 53.9 compared with 55.4 for graduates and 50.8 for freshmen; all three averages are significantly different from 75 ($p < 0.001$), showing that neither subject group orders according to the newsvendor model. Pairwise tests find no significant differences in average orders between managers and students (managers versus graduates, $p = 0.522$; managers versus freshmen, $p = 0.352$), which indicates that managers and students have a similar pull-to-center bias under demand-distribution information.

From Figure 4, the movements toward the optimal quantity from Phase 1 to Phase 2 in the basic treatment are small for all three groups; only the average graduate movement is statistically significant; $p = 0.021$ (within-sample test). The results of Phase 2 are thus essentially the same as the results of Phase 1: *All three subject groups exhibit the same kind of pull-to-center bias. There is no significant difference in the average orders placed by managers and students.*

Phase 3: Expected Profit Information. One possible explanation for the pull-to-center bias in the newsvendor game is that subjects have difficulties computing expected profit-maximizing quantities from cost, revenue, and demand information. To analyze how providing expected profit information affects order behavior, we provided a graph at the beginning of Phase 3 that shows how the expected profit depends on the order quantity (Figure 3). Figure 4 shows that the introduction of expected profit information moves the orders of all three groups to the optimal order quantity of 75.

Over all periods in Phase 3, managers ordered an average of 65.6 compared with 70.9 for graduates and 69.4 for freshmen. The movement of these averages over the respective Phase 2 averages is significant ($p < 0.001$ in all three cases; within-sample tests). Still, for managers and freshmen, the averages are significantly different from 75 ($p < 0.001$), and for graduates, they are weakly significantly different from 75 ($p = 0.073$). Pairwise tests find no significant differences in average orders between managers and freshmen ($p = 0.162$) but significant differences in average orders between managers and graduates ($p = 0.016$).

Thus, the results of Phase 3 can be summarized as follows: *Under expected profit information, all three subject groups exhibit a pull-to-center bias. There is no significant difference in the average orders placed by managers and freshmen, but graduates perform significantly better than managers.*

3.2. Use of Task Training

An alternative to providing decision support via expected profit information is to train employees in order management. Our task training furnishes our subjects with both the rationale and a method for computing the expected profit-maximizing order quantity. The summary statistics are provided in Table 4.

Observe from Figure 4 that for Phases 1 and 2, all three groups improve performance in the trained treatment versus the basic treatment (dashed versus solid line). The difference of mean orders between basic and trained treatments is significant for Phases 1 and 2 (for students at $p < 0.001$ for both phases, and for managers at $p = 0.064$ for Phase 1 and $p = 0.032$

Table 4 Summary Statistics for Treatments with Training, by Subject Population

Treatment	n	Mean order
Freshmen trained (FT)	22	65
Graduates trained (GT)	30	70
Managers trained (MT)	23	62

for Phase 2). There is no significant difference at any standard level for any subject group for Phase 3. Still, all subject groups reach their maximal performance in Phase 3 with training. On average, trained freshmen ordered 69.7, trained managers ordered 66.6, and trained graduates ordered 73.3. Phase 3 orders are still significantly different from 75 for all three groups ($p \leq 0.03$ for all three subject groups).

Summing up, compared to the treatment without training, *in Phases* 1 *and* 2 *task training significantly improves the performance for all three subject groups, with graduates performing significantly better than managers. In Phase* 3, *task training does not significantly improve the performances of any subject groups, and average orders of all subject groups stay significantly below the optimum of* 75.

3.3. Learning-by-Doing

The strong effects of introducing different kinds of information and training *between* our experiment phases are not confounded by learning-by-doing *within* phases. Table 5 shows the result of regression analyses to quantify the effect of learning-by-doing on mean orders ($\bar{O}_t$). We denoted the time period by t ($t = 1, 2, \ldots, 100$). To control for subjects, we ran a fixed-effects model using indicator variables S_i for each subject i. For the phases, we used 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 nonzero coefficient for t appears in the regression of the mean orders of managers, where it is negative, indicating a tendency for manager orders to move away from the optimal order quantity as the number of repetitions increases. So, *experience gained through play has no significant positive effect on ordering.*

Table 5 Effect of Learning-by-Doing on Mean Orders and Mean Efficiencies—Basic Treatments: Fixed-Effects Regression

$$\bar{O}_t^{\mathrm{FB}} = \sum_{i=1}^{32} d_i S_i + \underset{[0.663]}{0.011t} + \underset{[0.367]}{1.107\ Phase_2} + \underset{[0.000]}{\mathbf{19.35}\ Phase_3}$$

$$\bar{O}_t^{\mathrm{GB}} = \sum_{i=1}^{31} d_i S_i + \underset{[0.120]}{0.038t} + \underset{[0.000]}{\mathbf{5.043}\ Phase_2} + \underset{[0.000]}{\mathbf{21.67}\ Phase_3}$$

$$\bar{O}_t^{\mathrm{MB}} = \sum_{i=1}^{26} d_i S_i + \underset{[0.034]}{\mathbf{0.068}t} + \underset{[0.016]}{\mathbf{3.617}\ Phase_2} + \underset{[0.000]}{\mathbf{17.41}\ Phase_3}$$

Note. Two-tailed p-values are in brackets.

This result is consistent with what has been observed in student-only studies (see §1 for references). We just note here that with training, too, there is no significant learning-by-doing effect. Regressions similar to those presented previously exhibit an insignificant period coefficient ($p > 0.27$ in all cases).

4. Implications and Conclusion

Our main finding is that, in searching for the profit-maximizing stock, manager decisions exhibit the same pull-to-center bias as do student decisions. The finding is supported by our observation that managers are no better than students in utilizing information relevant to rational inventory stocking. We also find that classroom experience and on-the-spot training improve the performance of managers and students alike.

Our data speak to some of the theoretical treatments of newsvendor behavior. Bostian et al. (2008) fit newsvendor data to a reinforcement adaptive learning model (Camerer and Ho 1999). We observe across-phase learning effects, attributable to adding information, that cannot be explained by reinforcement adaptation models. Ho et al. (2010) describe a model of reference-dependent preferences in which newsvendors incur psychological (in addition to the economic) costs to being over- and understocked. Su (2008) shows that computational errors can explain the pull-to-center bias. Our findings regarding the success of on-the-spot training and providing expected profit information are consistent with debiasing psychological and reducing computational errors.

Observe also that our findings provide a justification for why none of the newsvendor models that we are aware of is explicitly restricted to student subject pools, although the motivating data almost exclusively come from student subjects. That said, however, we emphasize that our results do not contradict the notion that job experience enables managers to do the job more effectively than students who have no such experience. Yet our results imply that the reason managers are better in the real world, if they are, is not because they are better at analytical thinking regarding the basic concepts behind inventory ordering, as modeled in the newsvendor problem. The version of the newsvendor problem we put forward is arguably one of the simplest to optimize against (uniform and stationary demand distribution and all other parameters are fixed and certain). The experiment induces a straightforward profit-maximization objective. That we observed managers (and students) generally maximize expected profits once they are explicitly told the profit-maximizing order quantity would strongly suggest they adopted this goal. True, in the field, managers likely confront different demand distributions and probably even different objective functions than those in our test version of

Table 6 Effect of Years of Working Experience and Position on Managers' Mean Orders: Random-Effects Maximum Likelihood Estimation

	Log$L$
$\bar{O}_i^M = 63.76 + 0.828 \cdot Phase_2 + \mathbf{9.713} \cdot Phase_3 - 0.216 \cdot WorkYrs + \mathbf{2.958} \cdot Position - \mathbf{0.476} \cdot Age + \mathbf{4.998} \cdot Grad + \mathbf{10.06} \cdot Trained$ [0.000] [0.562] [0.000] [0.394] [0.033] [0.008] [0.059] [0.001]	−503.0

Notes. Position $=1$ if simple employee (5 observations in the basic treatment/18 in the trained treatment), 2 if team manager (6/1), 3 if head of department (10/4), and 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. *Grad* $= 1$ if manager has a university diploma (see §2.2). Random effects are clustered on individual managers. A fixed-effects version of the model is rejected by the data. *Age* data exhibit multicollinearity with *WorkYrs* ($R^2 = 0.312$). Residuals from regressing *Age* on *WorkYrs* were substituted. Neither *Grad* nor *WorkYrs* show a strong relationship to *Position* ($R^2 = 0.084$ and 0.037, respectively). The two managers that did not give us education information were dropped from the estimate. Two-tailed p-values are shown in brackets.

the newsvendor problem. But if a manager has a better grasp of the basic principles underlying rational stocking than do students, he or she should be able to solve our laboratory version of the problem and to utilize the provided information better than the students. The fact that we observe the same pattern and magnitude of pull-to-center bias among procurement managers as observed among students is strong evidence that managers are unlikely to apply these basic principles in the more complex and dynamic work environment.[3]

It is possible that managers unconsciously apply these principles, overcoming the pull-to-center bias in their work environment through some sort of adaptive learning process. But we should probably not take a lot of comfort in this possibility. For one, adaptation in the face of variable feedback (of the sort received in the newsvendor problem) is a slow process (Kahneman 2003). It often fails to result in the optimal solution (Erev and Roth 1998; Bostian et al. 2008 and Ockenfels and Selten 2012 show this for the newsvendor problem). Finally, even if adaptive learning does eventually lead to the optimal solution, business is typically a dynamic environment—every time there is a dynamic bump, an adaptive learner who fails to grasp the underlying principles for rational judgment and suffers from pull-to-center bias will need to start the debiasing process again. From this we conclude that the pull-to-center bias is likely to be prevalent in inventory manager decisions.

Given this conclusion, what might improve the efficiency of field judgments of procurement managers in the field? Our experiment suggests two answers: experience and cognition. With regard to experience, our findings offer a number of clues as to what might help. In Phases 2 and 3 of all experiments, graduates tended to outperform managers and freshmen. This suggests that classroom education provides important insight into the process behind the newsvendor solution. Consistent with this observation, training has a strongly positive 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, too, play an important role in the effectiveness of classroom education. And the fact that theoretically redundant expected profit information 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.

Further support for the influence of "the right" experience, as well as a hint of the potential importance of cognition, is offered by the breakdown of manager performance in Table 6. The table examines the manager data for the influence of work experience (*WorkYrs*), position (*Position*), age (*Age*), and education (*Grad*) on ordering behavior, controlling for phase information and training. Given our findings that students, particularly graduates, use information and training better than managers overall, we might expect those managers with graduate education to perform better than managers without this educational experience. In fact, this is the case, albeit the effect is weakly statistically significant given a two-sided test. Another hypothesis might be that the accumulated experience from a greater number of work years or greater age might improve manager performance. In fact, both the work years and age coefficients have a negative value, the former insignificantly and the latter significantly decreasing performance. Higher organizational position, however, is associated with higher performance.[4]

[3] One potential reason why managers fail to order optimally in the laboratory is that they rely on decision support tools in the field that are not available in the lab. In any case, it would be a promising complementary task to go beyond theory and laboratory studies and to very closely study stocking practices in particular companies (for the purposes of hypothesis building regarding objectives, demand scenarios, the role of decision support systems, parameter dynamics, etc.).

[4] Substituting dummy variables for the *Position* variable, one dummy for each hierarchy value greater than 1, we find the estimates for the new coefficients to all be positive and monotonically increasing as we move up the hierarchy; only the coefficient for hierarchy level 4, however, is significant at standard levels ($p = 0.028$).

One natural explanation for these results would be that those managers with higher education and those with higher positions are the managers with higher analytical skills, the kinds of skills that plausibly account for graduates' better performance. Managers with more years of work experience (controlling for position) might rely less on analytical reasoning but more on routine-based decision making (see Selten 1978 for an early description of different levels of decision making and when they apply). Another hypothesis consistent with these findings (not mutually exclusive with experience) would be that cognitive ability is a major factor in better newsvendor performance; for a first step in this direction of investigation, see the recent paper by Moritz et al. (2011).

Finally, one important implication of our results has to do with the use of experimental methods. In the research community, we have frequent discussions on whether the effects found in laboratory experiments with student subjects translate to manager subjects. The results of our study show that, at least for the newsvendor model, the direction of effects is the same for students and managers, but that the magnitude of the effects can be different. This seems to suggest that the use of student subjects is justified for analyzing the direction of the effects in manager populations but that manager subjects should be used for estimating the magnitude of effects in manager populations.

Acknowledgments

The authors thank Elena Katok for developing and sharing the inventory software with them; Felix Lamouroux for competent research assistance; Mirko Kremer, Michael Becker-Peth, and Torsten Gully for a close read of the paper; and the participants in numerous seminars in the United States and Germany for their useful comments. The first author gratefully acknowledges the support of the National Science Foundation. The second author gratefully acknowledges the support of the National Science Foundation through the Leibniz program. The authors gratefully acknowledge the support of the Deutsche Forschungsgemeinschaft through the research group "Design and Behavior."

References

Arrow KJ, Harris T, Marschak J (1951) Optimal inventory policy. *Econometrica* 19(3):250–272.

Ball S, Cech P (1996) Subject pool choice and treatment effects in economic laboratory research. Isaac RM, ed. *Research in Experimental Economics*, Vol. 6 (JAI Press, Greenwich, CT), 239–292.

Benzion U, Cohen Y, Peled R, Shavit T (2008) Decision-making and the newsvendor problem—An experimental study. *J. Oper. Res. Soc.* 59(9):1281–1287.

Bolton GE, Katok E (2008) Learning by doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback. *Manufacturing Service Oper. Management* 10(3):519–538.

Bostian AA, Holt CA, Smith AM (2008) Newsvendor "pull-to-center" effect: Adaptive learning in a laboratory experiment. *Manufacturing Service Oper. Management* 10(4):590–608.

Camerer CF, Ho T-H (1999) Experience-weighted attraction learning in normal form games. *Econometrica* 67(4):827–874.

Cooper DJ, Kagel JH, Lo W, Gu QL (1999) Gaming against managers in incentive systems: Experimental results with Chinese students and Chinese managers. *Amer. Econom. Rev.* 89(4): 781–804.

Croson RTA (2010) The use of students as participants in experimental research. BDOM Network. Accessed July 5, 2012, http://www.ombehavior.com/.

Erev I, Roth AE (1998) Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *Amer. Econom. Rev.* 88(4):848–881.

Fréchette GR (2012) Laboratory experiments: Professionals versus students. Frechette G, Schotter A, eds. *The Methods of Modern Experimental Economics* (Oxford University Press, Oxford, UK).

Ho T-H, Lim N, Cui TH (2010) Reference dependence in multilocation newsvendor models: A structural analysis. *Management Sci.* 56(11):1891–1910.

Kahneman D (2003) Maps of bounded rationality: Psychology for behavioral economics. *Amer. Econom. Rev.* 93(5):1449–1475.

Katok E, Thomas D, Davis A (2008) Inventory service-level agreements as coordination mechanisms: The effect of review periods. *Manufacturing Service Oper. Management* 10(4):609–624.

Kremer M, Minner S, Van Wassenhove LN (2010) Do random errors explain newsvendor behavior? *Manufacturing Service Oper. Management* 12(4):673–681.

Lurie NH, Swaminathan JM (2009) Is timely information always better? The effect of feedback frequency on decision making. *Organ. Behav. Human Decision Processes* 108(2):315–329.

Moritz B, Hill AV, Donohue K (2011) Individual differences in the newsvendor problem: Behavior and cognitive reflection. Working paper, University of Minnesota, Minneapolis.

Ockenfels A, Selten R (2012) Impulse balance in the newsvendor game. Working paper, University of Cologne, Cologne, Germany.

Ren Y, Croson RTA (2012) Explaining biased newsvendor orders: An experimental study. Working paper, University of Texas at Dallas, Richardson.

Schultz KL, McClain JO, Robinson LW, Thomas J (2007) The use of framing in inventory decisions. Working paper, Cornell University, Ithaca, NY.

Schweitzer ME, Cachon GP (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Sci.* 46(3):404–420.

Selten R (1978) The chain store paradox. *Theory Decision* 9(2): 127–159.

Su X (2008) Bounded rationality in newsvendor models. *Manufacturing Service Oper. Management* 10(4):566–589.