Team decision-making and individual learning in the newsvendor problem: a laboratory investigation

Nico Laya, Nico.Laya@opencountry.co.nz, phone: +64 9 5891372, facsimile: +64 9 525 347

Open Country Dairy, PO Box 11 159, Ellerslie, Auckland 1542, New Zealand

Valery Pavlov, v.pavlov@auckland.ac.nz, phone: 64 9 923 8726, facsimile: 64 9 373 7430

The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

Abstract

This laboratory study investigates the impact of the decision-making hierarchy and heterogeneity of knowledge on team performance and on individual performance after the team is dissolved. Our experiment examines effects of both factors. Hierarchical teams perform similarly to individual decision-makers but flat teams outperform both by making fewer extremely poor decisions. Additionally, former members of flat teams make better decisions than former members of hierarchical teams and participants who were not involved in team decision-making. The effects of knowledge heterogeneity are more subtle. While experts do not seem to improve the performance of their teams, their former team members tend to get rid of their old biases more effectively than individuals and former members of homogeneous teams.

Keywords

behavioral operations; knowledge transfer; newsvendor; group learning

Acknowledgements

Nico Laya and Valery Pavlov gratefully acknowledge the support of the University of Auckland Faculty Research Development Fund (grant # 3700887)

1. Introduction

Many important business decisions are made in teams. Gordon (1992) reports that 82% of organisations with 100+ employees use teams to manage their activities. This industry practice goes along well with academic research findings. First, teams tend to make better decisions than individuals and, in particular, act more consistently with the reasoning underlying backwards induction (Bornstein, Kugler, and Ziegelmeyer 2004), learn better and play more strategically (Cooper & Kagel 2005), are less prone towards myopic loss aversion (Sutter 2007) and are better in information processing (Tindale 1989). Charness and Sutter (2012) summarise their literature review as follows: “The bottom line emerging from economic research on group decision-making is that groups are more likely to make choices that follow standard game-theoretic predictions, while individuals are more likely to be influenced by biases, cognitive limitations, and social considerations.” Second, significant knowledge transfer takes place in the process of decision-making and, as a result of learning from each other, former team members make better decisions individually afterwards (Maciejovsky et al., 2013).

These capacities of the team decision making can be particularly valuable for dealing with a variety of complex business problems that involve uncertainty and are collectively known as the *newsvendor problem* (Arrow et al., 1951; Edgeworth, 1888; Porteus, 2008). The problems range from a multi-billion dollar problem of investing into a promising technology to an everyday decision of how far in advance to leave home going to work in the morning. The travel time is random, affected by the weather, road accidents, etc., and, therefore, there is always a chance of arriving late and incurring related costs. However leaving home far in advance is costly too and to find the best time somewhere between “too late” and “too early” one should carefully consider both the costs and the odds. The most common business applications include overbooking, workforce scheduling, production of fashion apparel, and other problems that involve allocation of a perishable resource.

However, despite the problem being familiar to most people from their everyday experience, there is now substantial evidence that most people are unable to make decisions in the newsvendor problem optimally. Schweitzer and Cachon (2000) observe that the average earnings across all participants were only about 85% of the earnings they could make by making optimal decisions. Another key finding of the study is that the problem challenges decision-makers somewhat differently from other problems that involve uncertainty. As Bostian et al. (2008) put it, “The newsvendor problem has fascinated decision theorists because it generates regular deviations from optimal behavior that cannot be explained by risk or loss aversion.” Instead, the overall pattern can be best explained as anchoring and insufficient adjustment coupled with minimizing ex-post inventory error. The research efforts that focused on fixing the decision biases showed that the quality of decisions can be greatly improved, for instance, by using “standing orders” – making decisions for several periods ahead (Bolton and Katok, 2008). However, such interventions may not be always available in real life. As Bolton et al. (2012) note, “… business is typically a dynamic environment …” but when the problem changes from period to period the use of standing orders becomes rather questionable. For the same token, other potentially effective approaches such as using a Decision Support System or hiring a consultant may not be available, leaving the decision-makers to deal with the problems on their own.

Our study makes two contributions. First, the research concerning team performance in the newsvendor problem has been very scarce. Gavirneni & Xia (2009), the only study of team decision-making in this context that we are aware of, find that teams are subject to anchoring and insufficient adjustment similarly to individuals so that the overall performance does not improve. However, their study does not address the impact of the factors peculiar for team decision-making and known to be critically important to the team performance. In our experiment we try to bridge this gap and investigate the impact of two of these factors: the decision-making hierarchy (see Anderson and Brown 2010 for a review) and heterogeneity of knowledge in the team (Devine et al., 1999; Stewart, 2006; Stewart and Barrick, 2000). Second, Maciejovsky et al. (2013) demonstrate the effectiveness of teams for individual learning but, (i) their results pertain to Monty Hall and Wason selection problems, the cognitive tasks very different from the newsvendor problem, and, (ii) they do not investigate the impact of team structure and heterogeneity of knowledge. In our experiment we manipulate these factors and find that (i) flat teams are more effective in avoiding extremely poor decisions, (ii) knowledge transfer takes place between team members, and, (iii) playing on a team with an expert has the effect of eliminating pre-existing biases.

The rest of the paper is organized as follows. Section 2 positions the study and formulates the research hypotheses. Section 3 explains the experiment design and rationale of choices. Section 4 presents and discusses the results. Section 5 concludes the paper summarizing the implications and suggesting directions for future research.

2. Related literature and research hypotheses

The consequences of sub-optimal decisions can be quite detrimental to a company's success or even mere survival. Therefore, it is not surprising that companies have been using various means to improve the quality of decisions, most commonly Decision Support Systems (DSS), external consultants and teams. We do not review the studies of DSS and the use of external consultants noting only that neither is a substitute for regular managers and teams: (i) there is evidence that managers tend to doubt the quality of DSS recommendations in situations that involve uncertainty (Keizers et al., 2003; Wagner, 2002) and override them (Van Donselaar et al., 2010), and (ii) organisations tend to use the external expertise to solve big infrequent problems, dealing with operational problems themselves because they have (or believe they have) the required expertise themselves (Bryson, 1997).

2.1. Team decision-making and characteristics

With the size and significance of problems that keep growing, the utilisation of work teams has been ubiquitous in organisations, especially for organisations that have multiple divisions, multiple departments, and more employees (Devine et al., 1999). Lawler et al. (1995) report a significant increase of team utilisation in Fortune 1000 companies from 70% in 1987 to 91% in 1995. Such an increase occurs not only out of necessity but, referring to the sources cited in the Introduction, also because teams do tend to make better decisions than individuals.

Teams differ along a number of dimensions such as purpose (Sundstrom et al., 1990), type of decision-making process (Devine et al., 1999), autonomy, group composition, etc. (Guzzo and Dickson, 1996; Stewart, 2006). An example could be an ad-hoc, geographically dispersed, selfmanaged cross-functional team made to cope with a certain production problem, having the authority to choose their leader but using voting to arrive to decisions. For the purposes of this study we focus on two dimensions that have been long recognized as critical factors determining the team performance: hierarchical structure of decision-making and expert knowledge (Humphrey et al., 2002).

2.2. The newsvendor problem

We mentioned earlier that we chose the newsvendor problem as a decision-making task for this study because it is ubiquitous in the business world. However, another reason is that it turned out to be a hard and rather unique problem, presenting non-trivial cognitive challenges for humans. For academic researchers, the results are challenging as well. First, the problem belongs to the "decision-making under uncertainty" category but, surprisingly, risk- and loss-aversion, biases that typically drive decisions under risk (Kahneman and Tversky, 1979; Tversky and Kahneman, 1974), cannot explain the results observed in the newsvendor problem (Schweitzer & Cachon 2000). Second, although the quality of decisions tends to improve over time, learning is slow and not necessarily leads to the optimal decision (Kahneman, 2003). These results motivated several follow-up experimental studies.

The search for interventions to enable better decisions went mainly in two directions – improve learning and provide training. With regard to learning, several studies (Bolton and Katok, 2008; Bostian et al., 2008; Lurie and Swaminathan, 2009) investigate standing orders and feedback frequency. A standing order, by forcing a decision-maker to experience the consequences of the same decision for several periods, effectively reduces the feedback variability and may be expected to facilitate learning. Indeed, a common finding of these studies is that lowering the feedback frequency improves learning. As Bolton and Katok (2008) explain: "… observe that the standard error of the average of 10 independent observations is smaller by a factor of the square root of 10 than the standard deviation of a single observation."

With regard to effectiveness of training, yet another striking observation from the Schweitzer and Cachon (2000) study is that although the experiment participants were second-year MBA students from MIT who covered the newsvendor problem in the first year of the program, in the experiment fewer than 10% of them were making decisions maximizing their monetary payoffs. Nevertheless, in-task training provided in (Bolton et al., 2012) right before the decision task had a strong positive effect on performance. (Moritz et al., 2013) present a case for a broader, out of task training. Using a Cognitive Reflection Test (Frederick, 2005) they find that people who score higher on the test also make higher profit in the experiment. The authors find that "… cognitive reflection is a better predictor of task outcome than other individual characteristics such as college major, years of experience, or managerial position". This result resonates with (Bolton et al., 2012) finding that graduate students and high-rank managers (but not those in lower ranks) tend to outperform undergraduates. The authors propose that "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." Overall, the existing experimental evidence suggests that analytical reasoning is superior to learning through experience.

2.3. Research hypotheses: team performance

Considering the overall conclusion of Charness and Sutter (2012) that teams show a better cognitive ability than individuals in a variety of tasks, and the results of Moritz et al. (2013) that link cognitive abilities to high performance in the newsvendor problem, one might expect teams to perform better than individuals in the newsvendor problem as well. However, this problem, again, appears to be an exception. Gavirneni and Xia (2009) tested the team performance in the newsvendor problem and found no improvement over individuals, although with a somewhat smaller number of particularly poor decisions. Allowing for the quite unusual specifics of decision-making in the newsvendor problem mentioned above, it is Gavirneni and Xia (2009) result that we take as our first hypothesis.

Hypothesis 1a. Teams perform similarly to individuals.

Next, the extant literature provides mixed evidence of the effect of team structure. (Beekun, 1989) argues that flat teams are more flexible in processing information and can work faster than hierarchical teams. The reason, as some later studies argue, is that critical views among members in a team might enable situations where alternatives generated can be questioned and eventually force every member in a team to use their knowledge and ability more optimally (Cohen, 1994; King, 2002; O'Donnell and O'Kelly, 1994). Nevertheless, (Anderson and Brown, 2010) conclude from their review of laboratory and field studies that the evidence is mixed. Therefore, we do not have strong evidence for the superiority of either flat or hierarchical groups and, therefore, choose a neutral claim as our null hypothesis.

Hypothesis 1b. Flat teams perform similarly to hierarchical teams.

Considering the possible effects of heterogeneity in teams, we use the individual performance in the newsvendor problem as the relevant dimension of heterogeneity. This choice is based on a finding that heterogeneity in teams will be more beneficial when the characteristics are more job related (O'Reilly III et al., 1998; Pelled et al., 1999; Webber and Donahue, 2001). The evidence regarding the effect of heterogeneity is mixed. Some studies find that it may well be detrimental to team performance (Campion et al., 2006a, 2006b) while some others report that it may be beneficial (Bantel and Jackson, 1989; Magjuka and Baldwin, 1991). Yet, turning specifically to studies of teams that perform knowledge work, those tend to find that heterogeneity is beneficial (Guzzo and Dickson, 1996; Jackson et al., 1995; Stewart, 2006). Even more importantly, groups are often able to determine the best member when making quantitative judgement and choose the views held by the most cognitively central member (Henry, 1995; Kameda et al., 1997). Summarizing these literatures, we state the following hypothesis.

Hypothesis 1c. Heterogeneous teams outperform homogeneous teams and individuals.

2.4. Research hypotheses: knowledge transfer

While the performance of a given team in a given task can be very important for an organization, learning that takes place inside the teams is critically important in the long run (Gavin 1993). Therefore other aspects peculiar to team decision-making that our study concerns are knowledge transfer (Argote and Ingram 2000) and cognitive skills (Singley 1989). In a recent study, Maciejovsky et al. (2013) observe significant knowledge transfer, finding that in two classic cognitive tasks, the Monty Hall problem and the Wason selection task, the very exposure to team decision-making improves the quality of decisions individuals make afterwards. Their result clearly supports an earlier view that "the group decision-making has a unique potential to debias …" (Sniezek, 1992) and we use it to state the following hypothesis.

Hypothesis 2a. Exposure to team decision-making results in a knowledge transfer and better individual decisions afterwards.

However, Maciejovsky et al. (2013) consider neither the role of decision-making hierarchy nor the impact of the heterogeneity of knowledge. In this regard, the literature indicates that both are important for knowledge transfer. Rulke & Galaskiewicz (2000) as well as Csaszar and Eggers, (2013) find that when some specific knowledge exists in a team, flat teams are effective in identifying the specific member who is more competent to complete the task. Identifying the most knowledgeable person is necessary but not sufficient for knowledge transfer to take place. After all, after teams identify the right person to handle the problem they delegate the task to that person so that other team members may not even be involved at all. However, (Somech, 2006) finds, specifically, that in heterogeneous teams flat structures are more beneficial as they can accommodate and encourage more cognitive processes of team reflection among members. Similarly, Michaelsen (2002) suggests that in order to enable effective team learning, distribution of knowledge is essential. We summarize these findings as the following hypotheses.

Hypothesis 2b. Knowledge transfer is stronger in flat teams than in hierarchical ones.

Hypothesis 2c. Knowledge transfer is stronger when at least one of team members is an expert in both flat and hierarchical teams.

3. The experiment

3.1. Decision making task: the Newsvendor Problem

In the experiment we use the most basic version of the newsvendor problem, the same used in the aforementioned experiments. The newsvendor needs to decide how many units, q, of the product to have in stock by the beginning of the selling season. The per unit selling price on the market, *p*, and the production cost, *c*, are known constants (*p>c*). The customer demand, *D*, is a random variable drawn from a known distribution $F(x)$. The realization of *D* is unknown at the time when the decision must be made and the lead time is longer than the selling season. All units unsold by the end of the season perish (there is neither cost of disposal nor a salvage opportunity). The newsvendor's profit realized during the selling season is, therefore, a random variable, $\pi(q, D) = p\min(q, D) - cq$, so that the newsvendor's problem is to find q that maximizes the expected profit $E[\pi(q, D)] = p\left(\int_0^q xf(x)dx + \bar{F}(q)\right) - cq$. The optimal number of units is implicitly defined by the first-order condition known as the Critical Ratio or Critical Fractile: $F(q^*) = \frac{p-c}{p}$. In our experiment we used the uniform distribution but, to prevent simple copying of decisions from other team members, we varied the parameters of the problem such that the problem used at the team stage is sufficiently different from the problem used during the individual decision-making (see details in the description of the experiment below). This way, ordering similar quantities in both instances was not even possible. We used Veconlab (Holt, 2015) online implementation of the newsvendor problem as the decision-making interface.

3.2. Participants and incentives

Participants were students of the [*the name of the university concealed for the review purposes*] from a variety of majors, mostly business school undergraduates. Overall, 122 subjects participated in the study. To recruit participants we used ORSEE (Greiner 2004). All sessions were conducted in a computer laboratory dedicated and equipped for running decision-making experiments.

Participants were paid a \$5 [*country concealed for review purposes*] "show-up" fee and an additional amount of money proportional to the profit they made acting as newsvendors during the experiment. Those participants who earned negative profits were paid \$5. The average payment, including the show-up fee, was around \$20.

3.3. The design of the experiment

The design of the experiment has much in common with that of (Maciejovsky et al., 2013), including the number of teams in each condition. In terms of the number of factors, their levels and controls, our design corresponds, approximately, to their Study 2 and Study 3 combined. The design is mixed, both between- and within-subject. Between-subject, we use a 2x2 full factorial design with one additional treatment in which subjects act individually. The two factors manipulated are the team structure (“flat” vs. “hierarchical”) and heterogeneity of knowledge in teams (“regular” vs. “expert”). Similarly to Gavirneni and Xia (2009), we choose face-to-face communication mode for teams and three-member teams. Note that (Maciejovsky et al., 2013) use two-member teams but given our focus on flat vs. hierarchical teams we consider it important to control for a possibility of flat teams acting informally as hierarchical teams due to personality traits (e.g., gender, see (Sczesny et al., 2004), (Ivanova-Stenzel and Kübler, 2011) and references therein) and, we believe, three-member teams are less susceptible to emergence of an informal hierarchy. Each subject participated in only one experimental session. Within-subjects, each participant dealt with 30 repetitions of the newsvendor problem. In periods 1 to 10 (Part 1) and 21 to 30 (Part 3) decisions were made individually; communication between participants was not allowed. In periods 21 to 30 (Part 2) participants played in teams (except in the control treatment).

Table 1. Between-subject part of the experiment design

Factors		
	Homogeneous teams	Teams with experts
Flat	10 teams	7 teams
Hierarchical	10 teams	9 teams

The difference between flat and hierarchical teams is that in a hierarchical team one team member is assigned a role of decision-maker and the other two members are advisors. For designated decision-makers the conversion coefficient from experimental tokens to dollars was 40:1 while for advisors it was 60:1. This difference in the conversion coefficients was explained in the experiment instructions and also announced aloud before each session. Since sessions were conducted on different days at different times, it was possible that in some sessions, participants were, for example, mainly freshman or mainly accounting students. To partially block biases that could be potentially introduced due to sessions conducted at different days and time of day, in each session we used both flat and hierarchical teams. Assignment of participants during each session to different teams and conditions was random except for assignment of experts to teams (explained below).

To study the effect of heterogeneity of knowledge we manipulate another factor: expertise. In the “expert” (or “seeds”) condition each team had one member (an expert) who knew how to find the optimal number of units. To prepare experts we provided an additional set of instructions. We did not specify the number of units that was optimal for the market conditions used in the experiment but only explained the reasoning that allows finding the optimal quantity. Applying the reasoning is a cognitive task on its own and we expected that not everyone would be able to cope with it successfully. Therefore, in the “expert” condition subjects arriving to the lab were initially separated in two equal groups. One group had access to regular instructions only while the other group had an extra set of instructions. Out of those who received extra instructions only about 50% were able to make optimal decisions when playing individually in the first part and they became “seeds” in teams formed for the second part. The rest were allowed to participate individually but we are not using their data in the analysis. To control for the effects of providing the information about the relative expertise of the team members (Bonner et al., 2002) we do not allow team members to see profits made by others and we do not explain the very existence of “experts” to the participants, including themselves (note that there was no deception involved). Lastly, to reduce variance in the expertise of the designated decision-makers in the “expert” condition we assigned experts to be the team leaders. This is not an obvious choice, of course. In real life, hidden experts are not likely to be always put in a “best shot” position but, (i) for a small sample size extra experiment a tighter control appears more beneficial than randomization and (ii), whether this causes a strong bias can be revealed later at the analysis stage by testing the corresponding interaction term.

Similarly to Bolton and Katok (2008), we use 10-period standing orders. Participants had to choose the number of units for periods 1 to 10 (Part 1), then another quantity for periods 11 to 20 (Part 2), then, again, for periods 21 to 30 (Part 3). Each 10-period part lasted for 10 minutes.

Participants were required not to type anything on computers in the first 5 minutes and then enter their decision 10 times in the remaining 5 minutes. That is, during the experiment each participant made only three inventory decisions and had 5 minutes to arrive at each decision. This particular timing was chosen because (i) inter-team communication requires time and giving a team 5 minutes to arrive to their decisions was in line with previous team experiments (e.g., Sutter 2005) and, (ii) allocating the same time for all parts allowed controlling for time pressures.

Table 2. Within-subject part of the experiment design

Part	Periods	Timing	Decision-making	Market parameters
Training		5 + 5 min	individual	Low-margin p=9, c=3, U(50,150), period fixed cost = 150, q*=117
Part 1	1 to 10	same as previous	individual	High-margin p=12, c=3, U(0,100), period fixed cost = 0, q*=75
Part 2	11 to 20	same as previous	team	Low-margin p=12, c=9, U(100,200), period fixed cost = 300, q*=125
Part 3	21 to 30	same as previous	Individual	Same as in Part 1

We use a high-profit setting of the newsvendor problem in Parts 1 and 3 (individual decision-making) and a low-profit setting for Part 2 (team decision-making). Had we used the same setting in Parts 2 and 3 we would not be able to disentangle a simple replication of someone else's decisions (without understanding of the underlying reasoning) and a decision made by applying reasoning learned from the former team members. The specific market parameters are similar to those in (Schweitzer and Cachon, 2000) and (Bolton and Katok, 2008).

3.4. Performance measures

All experimental studies we are aware of analyse order quantities and sometimes (e.g., Bolton & Katok 2008) also consider the efficiency of decisions (defined as the ratio of the expected profit for a given order quantity to the maximum expected profit). Both measures are important to analyse. The decision efficiency gives a clear understanding of how much money decision-makers are leaving on the table whereas order quantities are better for understanding biases that affect the decisions. Thus, for example, two order quantities, one smaller than the optimal quantity but a certain amount and the other larger by the same amount, are equivalent in terms of their expected profits but the biases behind these decisions may be very different and can be overlooked if the expected profit is the only performance measure used. In the analysis below we focus primarily on order quantities.

4. Results and discussion

4.1. The data set and considerations of the small sample-size

Overall, 122 people participated in the experiment; 108 took part in one of four team treatments (30+30+21+27 = 108) and 14 in the individual. Considering the number of observations per treatment is relatively small there is a danger that the results may be largely driven by even one abnormal observation. To mitigate the potentially detrimental effects of abnormal observations we used a two-fold approach. First, we screened the data for observations that could have resulted from hardware or software malfunctioning or participants not following the protocol of the experiment (e.g., different quantities entered by a participant in different periods within a stage), and did not find any. Second, similarly to (Maciejovsky et al., 2013) our approach to the data analysis was to partition the data and analyse the subsets most appropriate for a given research question (explained in more detail further). This way, a few influential observations would be less likely to affect all results but only some of them. Additionally, we used both regression models and non-parametric tests that are robust to outliers.

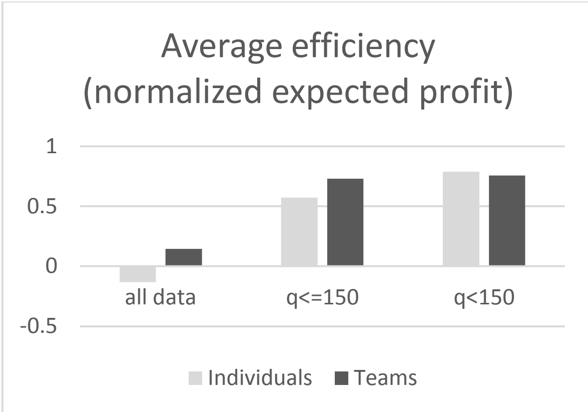
The variables in the dataset included the order quantities and the profits made at different stages, group IDs, as well as various dummy variables indicating whether an observation was on a team or an individual, flat or hierarchical team, teams with embedded experts (“seeds”) or without, etc.

4.2. Team performance

4.2.1. Are teams different from individuals?

Figure 1 shows the average efficiency of decisions in the individual and team conditions in Part 2. Calculated over the complete data set, the average efficiency is extremely poor, only 15% for teams and even negative for individuals, (–13%). The reason for this, in both cases, is the orders at or above the mean demand because such orders, respectively, result in zero and negative expected profits. Note that such orders are not uncommon in newsvendor experiments with individual decision-makers. In fact, more than 50% of orders fall outside of the "pull-to-center" (PTC) zone (Lau et al., 2014). In our experiment, we expect a larger proportion of decision blunders because of its within-subject dimension. Some participants might have simply ignored some information about the change in the market conditions from Part 1 to Part 2. Therefore, we compare teams and individuals using both the complete data set and the subset restricted to order quantities resulting in positive expected profits ($q2<150$) The reason we have chosen $q=150$ as the boundary is that $q>150$ results in the negative expected profit while "a rational person would not want to lose money." In both cases, despite Figure 1 shows the means values are somewhat different, we find no statistically significant differences (the Wilcoxon test gives $W = 274$, p-value = 0.64 for the complete data set and $W = 91.5$, p-value = 0.62 for the "rational" subset). That is, the data support Hypothesis 1a.

Figure 1. In the full data set, the average efficiency is close to zero because of many orders with q>150, especially in the individual condition. When only "rational", q<150, decisions are considered, the difference in the performance disappears.



4.2.2. Does the team structure matter?

To test whether the team structure affects the team performance we compare the expected profits resulting from the decisions made by hierarchical teams and flat teams. As before, we run three tests: using the complete data set, the subsets of "rational" and "irrational" decisions. The only difference that turns out to be statistically significant, albeit weakly, is in the case of irrational decisions (Wilcoxon tests of three comparisons: W = 137, p-value = 0.45, W = 63, p-value = 0.29, W = 4, p-value = 0.09). Next, we fit three 2-variable linear models to check whether our manipulations with the embedded experts and the team structure (represented by dummy variables *SEEDS* and *FT*), affect the teams' order quantities. The results presented in Table 1 show that the team structure did affect the team decisions but only in the case of truly bad decisions (the p-value is 0.06 but we consider the effect significant, since the sample size is very small). It is worth to re-iterate that the effect is not present in the model fitted to the complete data set and we only discovered it because we partitioned the data. The advantage of having many poor decisions in the data set is that if there is enough of them one can analyse them separately rather than treat them as outliers.

We conclude that the data refutes Hypothesis 1b.

Table 3. The team structure matters only for truly poor decisions. Flat perform better than hierarchical. Note also that the availability of an expert on the team has no effect.

Q2

	All data	“Rational”	“Irrational”
SEEDS	4.37	-1.35	3.60
	p = 0.44	p = 0.75	p = 0.44
FT	-1.83	-1.37	-9.64
	p = 0.75	p = 0.73	p = 0.06*
Constant	140.61	134.46	166.16
	p = 0.00***	p = 0.00***	p = 0.00***
N	36	26	10
Adj. R-squared	-0.04	-0.08	0.26

$^{***}p < .01$; $^{**}p < .05$; $^{*}p < .1$;

4.2.3. Does heterogeneity of knowledge matter?

Referring to Table 3, the coefficient for SEEDS is not statistically significant in any of the models and so, we conclude that, as far as team performance is concerned, the heterogeneity of knowledge does not matter. That is, we refute Hypothesis 1c.

4.3. Knowledge transfer

To test whether knowledge transfer took place and, if this was the case, which factors were at play, we fit several linear models. To isolate the effect of interest we used different subsets of the data (Maciejovsky et al., 2013, p.1263 and p.1266). The dependent variable in all models was the order quantity that participants placed in Part 3 (*Q3*), when they all acted individually. Table 4 presents the results. The leftmost column contains the list of regressors: the quantities ordered in Part 1 and Part 2 (*Q1* and *Q2*), the superior and the inferior (*Q1SUP* and *Q1INF*, based on the expected profit) order quantities placed in Part 1 by the team members, profits made in Part 1 and Part 2 (*PROFITP1* and *PROFITP2*), and, lastly, dummy variables for teams, experts and flat teams (correspondingly *TEAM*, *SEEDS*, *FT*). The numbers in the table are the coefficients and their p-values. The rectangles marked with “n/a” indicate regressors that were omitted because they were either kept constant or could not be defined in a given model. For example, *Q1SUP* and *Q1INF* are undefined for subjects who played in Part 2 individually and therefore could not be used in a complete model. Another example is the dummy variable *SEEDS*. It is redundant in Models 5 and 6 because in one of the subsets the dummy variable *SEEDS* was always equal to zero and in another it was always equal to 1. Similarly, the *TEAM* variable is redundant in models 3 to 6 because its value was equal to 1 in the subsets used to fit them. All models have been fit using a stepwise regression to identify the best subset of the regressors. The overarching idea of the analysis is that pre-existing biases are captured by *Q1* and by contrasting the changes in correlation between *Q3* and *Q1* one can understand the effect of different factors on learning.

Model 1 has been fit to the whole data set, including observations on subjects who participated in the individual condition. The results show that what happens during the team round has a strong impact on what happens after: although *PROFITP2* is not significant, *Q2* and TEAM are weakly significant, and *FT* is significant at p=0.03. Just as importantly, the strongest predictor of *Q3* appears to be *Q1*. At this point, it is worth noting that *FT* and *Q1* are significant in almost every other model (details follow).

Model 2 has been fit to a subset of data that contains observations only on people who made their decisions individually in all parts. The main takeaway from this model is that *Q1* is highly significant and large. Noting also that Adj. R-squared = 70% we conclude that *Q1* explains most of the variation in *Q3*. In other words, there is very little, if any, learning in the individual condition.

Model 3 investigates whether there is a knowledge transfer from top performing subjects to those who performed not as strongly. The data set for Model 3 only contains records that belong to people who were not the best among the team members. In order to test whether they learned from the top-performing team member(s) we introduce *Q1SUP* into the model. Its coefficient in Model 3 is statistically significant and large, 0.64. The coefficient of *Q1* is statistically significant as well, suggesting that Part 1 biases have not vanished completely, but it is much smaller than the coefficient of *Q1SUP*. That is, subjects who performed poorly in Part 1 made in Part 3 decisions that were much closer to those of top-performers.

Model 4 is the "mirror image" of Model 3, in the following sense. Its data set includes only records that belong to subjects made the best Part 1 decisions among their Part 2 team members. Also, instead of *Q1SUP* we introduce *Q1INF* into the model to test if subjects who performed better in Part 1 "learned" from their underperforming peers. The coefficient of *Q1INF* proved insignificant while the coefficient of Q1 is not only significant but is also the second-largest across all models, 0.71. Therefore, we conclude that top-performers are not affected by poor performers and maintain their original performance level stronger than other team members.

Models 5 and 6 "zoom" deeper into the data set of poor performing subjects. We further split the data subset containing poor Stage 1 performers into two subsets, one for teams with seeds and one without. Model 5, fitted into the subset without seeds, basically replicates the key results of Model 3: both *Q1* and *Q1SUP* are significant. Model 6 shows something very unique, never seen in other models: *Q1* is not significant.

Table 4. Decision biases and knowledge transfer

	Q3					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	All data	Indiv. only	Teams only			
			Inf. Dec.	Sup. Dec.	Inferior decisions	
					w/o seeds	w/ seeds
Q1	0.44 p = 0.000***	0.96 p = 0.001***	0.18 p = 0.01***	0.71 p = 0.02**	0.30 p = 0.001***	
Q2	0.25 p = 0.09*	0.36 p = 0.09*				-0.20 p = 0.07*
PROFITP1	-0.01 p = 0.05**					
PROFITP2						
TEAM	6.46 p = 0.08*		n/a			
Q1SUP		n/a	0.64 p = 0.02**		0.74 p = 0.004***	
Q1INF			n/a		n/a	
SEEDS				4.34	n/a	

				p = 0.09*		
FT	4.45		4.59	5.28		7.74
	p = 0.03**		p = 0.06*	p = 0.04**		p = 0.05**
Constant	3.44	-52.66	8.16	-81.03	-4.47	97.66
	p = 0.92	p = 0.21	p = 0.68	p = 0.24	p = 0.82	p = 0.000***
N	122	14	69	39	38	31
Adj. R-squared	0.20	0.70	0.16	0.23	0.35	0.14

***p < .01; **p < .05; *p < .1

Thus, Model 5 strongly supports our Hypothesis 2b – subjects who underperformed in Part 1 learn from their better performing peers. Moving on to Model 6 results, we immediately note that the coefficient for flat teams is significant. However, it is more important that the coefficient for *Q1* is not statistically significant and, again, it is the only model when this is the case. Our interpretation of this result is that playing on a team with an expert removes pre-existing biases.

Overall, we find that the data supports Hypotheses 2a, 2b and 2c - teams do make their members smarter but we also find, regarding 2c, that the exposure to working with experts makes a dramatic change to the players' biases.

5. Summary and future research

We conduct a laboratory experiment investigating team decision-making in a task that involves uncertainty and has countless applications in real life – the newsvendor problem. Specifically, we study the impact of (i) the type of decision-making hierarchy and (ii) the heterogeneity of expertise among team members, on (i) the team performance and (ii) knowledge transfer. Our study relates to literatures on team decision-making, learning in teams, and decision-making under uncertainty.

With regard to team performance *per se*, our data show that teams generally do not outperform individual decision-makers. Moreover, we observe that even teams with experts do not perform better than regular teams. We find this rather surprising. First, because it goes against the evidence from earlier studies showing that team members can often identify the most knowledgeable among them and side with her/his opinion. Second, because in hierarchical teams we placed experts to positions of designated decision-makers and they could simply implement their own decisions, completely ignoring suggestions of other team members in advisor positions. The only difference we find is between the performances of flat and hierarchical teams when it comes to “decision blunders”: highly detrimental mistakes made by flat teams were not as bad as mistakes made by hierarchical teams.

Speaking of knowledge transfer, we observe substantial learning from better performing peers (as opposed to copying). Individuals who make inferior decisions in the beginning, prior to the team decision-making, make better decisions later, correlated with previous decisions made by their better performing team partners whereas the latter are not affected by the former. This might seem an obvious result but, considering that teams performed no better than individuals, the grounds for the observed knowledge transfer are not clear. Interestingly, we observe complete debiasing in teams with experts. In regular teams, individuals who make inferior decisions in the beginning make better decisions later but these decisions tend to be positively correlated. However, such correlation disappears (becomes statistically insignificant) for individuals who played with an expert on the team.

These results, we believe, can offer certain managerial insights. First, “decision blunders” happen with teams just like with individuals. However, flat teams appear to be better in avoiding extreme mistakes. In our experiment such mistakes occur, supposedly, because the task significantly changes at the team stage. That is, flat teams may be better at handling situations when several task parameters change at once. We must note that we have only a very limited amount of data and this result is not very reliable. Second, although teams perform no better than individuals, temporary teams may still be very useful, provided they are flat, because of significant knowledge transfer helping individuals to make better decisions individually in the future.

Finally, our findings bring some questions that appear interesting for future research. First, similarly to our data, about 30% of decisions in the past newsvendor experiments are “decision blunders”, quantities falling on the wrong side of the demand distribution (Lau et al. 2014). They are the ones hurting profitability the most and, therefore, the potential of flat teams to avoid them calls for further investigation. It is possible that flat teams generally better utilize the relevant information. Second, we do not have a good understanding of why teams with (proven) experts performed not as we expected. On one hand, our data show that teams with experts performed no differently from teams without. One might argue that experts failed due to lack of any formal credentials legitimizing their status (Brandts et al. 2015) and, evidently, for their inability to demonstrate superiority of their solutions otherwise (Laughlin and Ellis, 1986). Yet, at the same, experts still greatly affected their team members. We find these two results hard to reconcile and warranting an in-depth investigation.

References

Anderson, C., Brown, C.E., 2010. The functions and dysfunctions of hierarchy. Research in Organizational Behavior 30, 55–89. doi:10.1016/j.riob.2010.08.002

Arrow, K.J., Harris, T., Marschak, J., 1951. Optimal Inventory Policy. Econometrica 19, 250–272. doi:10.2307/1906813

Bantel, K.A., Jackson, S.E., 1989. Top management and innovations in banking: does the composition of the top team make a difference? Strategic Management Journal 10, 107–124.

Beekun, R.I., 1989. Assessing the effectiveness of sociotechnical interventions: antidote or fad? Human Relations 42, 877–897.

Bolton, G.E., Ockenfels, A., Thonemann, U.W., 2012. Managers and Students as Newsvendors. Management Science 58, 2225–2233. doi:10.1287/mnsc.1120.1550

Bolton, G., Katok, E., 2008. Learning by Doing in the Newsvendor Problem: A Laboratory Investigation of the Role of Experience and Feedback. Manufacturing & Service Operations Management 10, 519–538. doi:10.1287/msom.1060.0190

Bonner, B.L., Baumann, M.R., Dalal, R.S., 2002. The effects of member expertise on group decision-making and performance. Organizational Behavior and Human Decision Processes 88, 719–736. doi:10.1016/S0749-5978(02)00010-9

Bornstein, G., Kugler, T., Ziegelmeyer, A., 2004. Individual and group decisions in the centipede game: Are groups more “rational” players? Journal of Experimental Social Psychology 40, 599–605. doi:10.1016/j.jesp.2003.11.003

Bostian, A.A., Holt, C.A., Smith, A.M., 2008. Newsvendor “pull-to-center” effect: Adaptive learning in a laboratory experiment. Manufacturing & Service Operations Management 10, 590–608.

Bryson, J.R., 1997. Business service firms, service space and the management of change. Entrepreneurship & Regional Development 9, 93–112.

Campion, M.A., Medsker, G.J., Higgs, A.C., 2006a. Relations between work group characteristics and effectiveness: Implications for designing effective work groups. Personnel psychology 46, 823–847.

Campion, M.A., Papper, E.M., Medsker, G.J., 2006b. Relations between work team characteristics and effectiveness: A replication and extension. Personnel psychology 49, 429–452.

Charness, G., Sutter, M., 2012. Groups make better self-interested decisions. The Journal of Economic Perspectives 26, 157–176.

Cohen, E.G., 1994. Restructuring the classroom: Conditions for productive small groups. Review of educational research 64, 1–35.

Cooper, D.J., Kagel, J.H., 2005. Are Two Heads Better Than One? Team versus Individual Play in Signaling Games. American Economic Review 95, 477–509.

Csaszar, F.A., Eggers, J.P., 2013. Organizational Decision Making: An Information Aggregation View [WWW Document]. URL http://pubsonline.informs.org/doi/ref/10.1287/mnsc.1120.1698 (accessed 11.17.13).

Devine, D.J., Clayton, L.D., Philips, J.L., Dunford, B.B., Melner, S.B., 1999. Teams in Organizations Prevalence, Characteristics, and Effectiveness. Small Group Research 30, 678–711. doi:10.1177/104649649903000602

Edgeworth, F.Y., 1888. The Mathematical Theory of Banking. Journal of the Royal Statistical Society 51, 113–127.

Frederick, S., 2005. Cognitive Reflection and Decision Making. The Journal of Economic Perspectives 19, 25–42.

Gavirneni, S., Xia, Y., 2009. Anchor Selection and Group Dynamics in Newsvendor Decisions—A Note. Decision Analysis 6, 87–97. doi:10.1287/deca.1090.0136

Gordon, J., 1992. Work teams: How far have they come? TRAINING-NEW YORK THEN MINNEAPOLIS- 29, 59–59.

Guzzo, R.A., Dickson, M.W., 1996. Teams in organizations: Recent research on performance and effectiveness. Annual review of psychology 47, 307–338.

Henry, R.A., 1995. Improving Group Judgment Accuracy: Information Sharing and Determining the Best Member. Organizational Behavior and Human Decision Processes 62, 190–197. doi:10.1006/obhd.1995.1042

Holt, C.A., 2015. Veconlab Experiment Selection Menu [WWW Document]. URL http://veconlab.econ.virginia.edu/admin.htm (accessed 5.20.15).

Humphrey, S.E., Hollenbeck, J.R., Meyer, C.J., Ilgen, D.R., 2002. Hierarchical team decision making, in: Research in Personnel and Human Resources Management, Research in Personnel and Human Resources Management. Emerald Group Publishing Limited, pp. 175–213.

Ivanova-Stenzel, R., Kübler, D., 2011. Gender differences in team work and team competition. Journal of Economic Psychology 32, 797–808. doi:10.1016/j.joep.2011.05.011

Jackson, S.E., May, K.E., Whitney, K., 1995. Understanding the dynamics of diversity in decision-making teams. Team effectiveness and decision making in organizations 204, 261.

Kahneman, D., 2003. Maps of Bounded Rationality: Psychology for Behavioral Economics. The American Economic Review 93, 1449–1475.

Kahneman, D., Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk. Econometrica 47, 263–292.

Kameda, T., Ohtsubo, Y., Takezawa, M., 1997. Centrality in sociocognitive networks and social influence: An illustration in a group decision-making context. Journal of Personality and Social Psychology 73, 296–309. doi:10.1037/0022-3514.73.2.296

Keizers, J.M., Bertrand, J., Wessels, J., 2003. Diagnosing order planning performance at a navy maintenance and repair organization, using logistic regression. Production and Operations Management 12, 445–463.

King, A., 2002. Structuring peer interaction to promote high-level cognitive processing. Theory into Practice 41, 33–39.

Laughlin, P.R., Ellis, A.L., 1986. Demonstrability and social combination processes on mathematical intellective tasks. Journal of Experimental Social Psychology 22, 177–189. doi:10.1016/0022-1031(86)90022-3

Lau, N., Hasija, S., Bearden, J.N., 2014. Newsvendor pull-to-center reconsidered. Decision Support Systems, Integrating Information Systems and Operations Management 58, 68–73. doi:10.1016/j.dss.2012.12.041

Lawler, E.E., Mohrman, S.A., Ledford, G.E., Participation, A. for Q. and, 1995. Creating high performance organizations: practices and results of employee involvement and Total Quality Management in Fortune 1000 companies. Jossey-Bass.

Lurie, N.H., Swaminathan, J.M., 2009. Is timely information always better? The effect of feedback frequency on decision making. Organizational Behavior and Human Decision Processes 108, 315–329. doi:10.1016/j.obhdp.2008.05.005

Maciejovsky, B., Sutter, M., Budescu, D.V., Bernau, P., 2013. Teams Make You Smarter: How Exposure to Teams Improves Individual Decisions in Probability and Reasoning Tasks. Management Science 59, 1255–1270. doi:10.1287/mnsc.1120.1668

Magjuka, R.J., Baldwin, T.T., 1991. Team-based employee involvement programs: effects of design and administration. Personnel Psychology 44, 793–812.

Michaelsen, L.K., 2002. Getting started with team learning. Team learning: A transformative use of small groups. Westport, CT: Greenwood.

Moritz, B.B., Hill, A.V., Donohue, K.L., 2013. Individual differences in the newsvendor problem: Behavior and cognitive reflection. Journal of Operations Management 31, 72–85. doi:10.1016/j.jom.2012.11.006

O'Donnell, A.M., O'Kelly, J., 1994. Learning from peers: Beyond the rhetoric of positive results. Educational Psychology Review 6, 321–349.

O'Reilly III, C.A., Williams, K.Y., Barsade, S., 1998. Group demography and innovation: Does diversity help?

Pelled, L.H., Eisenhardt, K.M., Xin, K.R., 1999. Exploring the black box: An analysis of work group diversity, conflict and performance. Administrative science quarterly 44, 1–28.

Porteus, E., 2008. The Newsvendor Problem, in: Building Intuition: Insights from Basic Operations Management Models and Principles. Springer Science & Business Media, pp. 115–134.

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

Sczesny, S., Bosak, J., Neff, D., Schyns, B., 2004. Gender Stereotypes and the Attribution of Leadership Traits: A Cross-Cultural Comparison. Sex Roles 51, 631–645. doi:10.1007/s11199-004-0715-0

Sniezek, J.A., 1992. Groups under uncertainty: An examination of confidence in group decision making. Organizational Behavior and Human Decision Processes 52, 124–155. doi:10.1016/0749-5978(92)90048-C

Somech, A., 2006. The effects of leadership style and team process on performance and innovation in functionally heterogeneous teams. Journal of management 32, 132–157.
Stewart, G.L., 2006. A meta-analytic review of relationships between team design features and team performance. Journal of Management 32, 29–55.
Stewart, G.L., Barrick, M.R., 2000. Team structure and performance: Assessing the mediating role of intrateam process and the moderating role of task type. Academy of management Journal 43, 135–148.
Sundstrom, E., de Meuse, K.P., Futrell, D., 1990. Work teams: Applications and effectiveness. American Psychologist 45, 120–133. doi:10.1037/0003-066X.45.2.120
Sutter, M., 2007. Are teams prone to myopic loss aversion? An experimental study on individual versus team investment behavior. Economics Letters 97, 128–132.
Tindale, R.S., 1989. Group vs individual information processing: The effects of outcome feedback on decision making. Organizational Behavior and Human Decision Processes 44, 454–473. doi:10.1016/0749-5978(89)90019-8
Tversky, A., Kahneman, D., 1974. Judgment under Uncertainty: Heuristics and Biases. Science 185, 1124–1131. doi:10.1126/science.185.4157.1124
Van Donselaar, K.H., Gaur, V., Van Woensel, T., Broekmeulen, R.A.C.M., Fransoo, J.C., 2010. Ordering behavior in retail stores and implications for automated replenishment. Management Science 56, 766–784.
Wagner, H.M., 2002. And then there were none. Operations Research 217–226.
Webber, S.S., Donahue, L.M., 2001. Impact of highly and less job-related diversity on work group cohesion and performance: A meta-analysis. Journal of management 27, 141–162.

Appendix

Experiment Instructions

Standard version

All subjects, including future experts, were provided the following set of instructions:

General Information

You are about to participate in a decision-making experiment. The purpose of this study is to understand how people make decisions in a particular situation similar to those that arise in business environments. Your earnings will depend on your decision-making performance, and an element of chance. If you follow these instructions carefully, we expect that your average earnings in this experiment will be around $25. Our experience however, suggests that the exact amount can differ quite substantially among participants, from $5 to $45. The amount paid to you will not be disclosed to other participants.

Your participation in the experiment is contingent to your agreeing to participate and provide the informed consent (two copies of the Participant Consent Form are available on your table). If you do not consent to the conditions of your participation explained in the document you received, including the informed consent, please quietly leave the room without disturbing other people. Otherwise, please sign both copies and hand in one copy to the person administering the experiment. You may keep the other copy for your records.

From this point and until the end of the experiment any communication (including Internet chats, email, SMS, etc.) unauthorized by the person administering the experiment is prohibited. At any moment, if you have any questions, please raise your hand and wait until the person administering the experiment comes to help you.

Description of the Experiment

Your "company" produces a single (fictional) product, the widget. The widgets can be sold to the customers. You role is to decide how many widgets the company should make before each of 30 selling seasons (periods or rounds). As the result of this activity your company can accumulate some profit expressed in experimental currency, "dollars". The total amount of profit earned by your company will be converted into [*country redacted for review purposes*] dollars at the exchange rate of, for example, 1000 experimental dollars per $1 [*country redacted for review purposes*] dollar and, upon completion of the experiment, you will be paid this amount in [*country name concealed for review purposes*] dollars plus a $5 participation fee. Different exchange rates apply at different stages of the game but every change in the exchange rate will be announced.

The profit made by your company is determined as follows:

1. There are 30 selling periods.
2. The game proceeds in 10-period sets and time allotted for each set is 10 minutes. Earnings will be calculated based on the profit made within this time frame.

3. You will need to make one decision before each of 10-period set, the number of units to hold in stock in the beginning of each period in the 10-period set (it will be the same number in all 10 periods).
4. The customer demand for your widget in a given period is random, and not known to you at the time you need to make a decision. The likelihood of each demand realization will be specified on your decision screen. It may say, for example, demand can be any number from 0 to 100, equally likely (or, equivalently, it may say the demand distribution is uniform from 0 to 100).
5. Each period demand realization is independent of everything else.
6. Market price is fixed. This may be contrary to some “supply-demand” models used in the economics courses, but in this case the market price depends neither on the number of customers willing to buy the widgets nor on the number of widgets supplied to the market. For example, whether you produced 17 units while 25 customers wanted to buy them or you ordered 91 units while only 14 customers wanted to buy them, the price at which units are sold are the same.
7. Units left over at the end of a period can’t be carried over to the next periods and have to be disposed with no salvage value.
8. Profit made in each period will be accumulated. This will then result in the total profit. Positive profit made in a period will increase the total profit and negative profit (a loss) will decrease it.
9. The market conditions, such as the cost of a unit to buy, the selling price, and the distribution from which demand is drawn may be **different** in different 10-period sets. At any point they will be shown on the screen of your computer. Please **carefully** examine the market conditions before making your decisions. If the exchange rate (conversion coefficient from the experimental currency to NDZ) changes, this will be announced by a person administering the experiment.

[The section “Team rounds” was used in team treatments]

Team rounds

An important feature of this experimental session is that in periods 11 to 20 you will play as part of a team. All participants will be split in groups of three and you will stay on the team with the same individuals until period 21. The quantity each team decides to produce in a given period will be the same for all three companies that belong to the team members. The demand realizations will be the same for these companies and so the profits made by each company will be the same.

In this session teams are hierarchical. One player of each team is a designated decision-maker, having access to the computer interface and making decisions on behalf of the team. Other two team members play advisory roles. They can discuss which quantity they think is the best to order. They do not need to come to consensus and may advise different quantities to the decision-maker. The decision-maker is free to order any quantity, regardless whether it was advised by the other team members or not.

Please also familiarise with the examples on the next page to make sure you correctly understood the above. After you finish, please put the instructions on a side so the experimenter could see that you are ready to start the experimental part.

Please familiarise with the examples on the next page to make sure you correctly understood the above. After you finish, please put the instructions on a side so the experimenter could see that you are ready to start the experimental part.

Examples

You were provided with the following information:

1. In period 1 through 20, the market price is 12 $ and the cost to purchase one unit is 3 $. In each period, market demand can be any integer number from 1 to 300, each equally likely. Suppose in period 4 you decided to order 90 units when the realised demand was only 50 units. Therefore you sold 50 and had to dispose 40 units left over. As a result, you profit in this period was:
   12*50 – 3*90 = 330 $

That is, in period 4 you earned 330 experimental dollars and, if, suppose, the exchange rate is 1000:1, you earned $0.33 [*the country of currency concealed for review purposes*].

2. If it happened that demand was high (say 92, 91, 95, …) for few periods in a row. How likely is that demand will be high demand in the next period?

No, the likelihood of any demand realization does not depend on what happened in the past.

Quiz

1. In period 21 through 30, everything is the same except the cost per unit which has now become 9 $. Suppose in period 25 your company ordered 190 units but the realised demand was only 130 units. Question: How much profit did your company make in this period?
2. It happened that for several periods demand was higher than the number of units your company had in stock. Question: How likely is that the customers may become dissatisfied and a smaller number of them will come in the next periods?

Expert version

Subjects randomly selected to receive the explanations of how to use the marginal analysis to find the optimal order quantity were provided, in addition to the standard set of instructions, the following description of the method (there was no quiz after this part):

Additional decision support information

The market environment simulated in this experiment challenges the decision-maker with a rather demanding cognitive task. Finding the quantity maximizing monetary payoff of the decision-maker requires several steps:

1. Making sense of what it means "maximizing monetary payoffs". The profit realized in any period depends on an element of chance. Its realization is not under the decision-maker's control and while ordering a larger quantity increases the maximum possible profit realization it also increases the maximum possible loss. What is under the decision-maker's control is the *expected* profit and it is this objective that companies normally

pursue in such situations. The expected profit is a function of one variable, the order quantity.

2. In principle, finding the optimal order quantity is a relatively straightforward task once one can write it down using the standard “Revenue – Cost”. Then, making use of a powerful mathematical fact, one can find the optimal quantity by finding the point where “Marginal revenue = Marginal cost” (this is equivalent to setting the derivative of the profit function to zero). The difficulty on this path is that one needs to specify not “Revenue” but “Expected Revenue” (the same applies to the cost but in this case there is no uncertainty related to cost and the expected cost is simply the production cost). Deriving the expected revenue as a function of the order quantity analytically is possible in the problem presented in the experiment. However, this requires familiarity with calculus, specifically with integration, and most people, even the professional inventory managers are not likely to do that immediately.
3. In principle, one can learn what the profit function looks like if there is a possibility to try different quantities over and over and plot the average the profits resulting from different quantities. However, a person experienced with this type of problems would realize that there is so much variability that almost nothing can be learned even after 100 trials, let alone 10.
4. At this point, despite it may seem there are no means readily available for the decision-maker to find the best order quantity, the decision-maker can still find it by revisiting the analytic approach. While it is true that finding the expected profit function requires integration, it is not required for finding the point where “Expected Marginal Revenue = Expected Marginal Cost”. One can specify both sides of this equation directly. First, as it was mentioned before, there is no uncertainty about cost and so the expected marginal cost is simply the marginal (i.e. per unit) cost. Second, suppose there are N-1 units of the product in stock and consider the *change* in the expected revenue if one adds another unit that can be sold only after all previous N-1 are sold. In other words, this Nth unit will be sold only if demand is at least N.

- For example, suppose the demand is such that each number from 201 to 250 is equally likely. Then the probability that demand will be higher than 240 units is (250-240)/50 = 10/50 = 0.2. Therefore, if, suppose, the market price is 30, the expected marginal revenue of the 40^{th} unit is 30*0.2 = 6

Suppose now that the per unit production cost is 4. Since 6>4 this means that the 40^{th} unit is worth having in stock. One can try now a higher quantity, say 50, calculate the expected marginal revenue, compare it with the per unit marginal cost and decide whether 50 is too much or too little, until the exact quantity is found when "expected marginal revenue = expected marginal cost".