ASSESSING THE ACCURACY OF PURCHASE QUANTITY DECISIONS

BRYAN LILLY, University of Wisconsin Oshkosh

This study investigates decision-making accuracy with respect to quantity aspects of purchase decisions. A set of experiments simulated a retail purchase context. These experiments gauged decision accuracy and tested factors that bias quantity judgments. Experiment participants tended to overbuy, but overall accuracy was high: 88 percent of decisions were within ten percent of the optimal quantity. Decision accuracy worsened in the presence of demand trends, increased variation in demand probabilities, and increased attention to non-diagnostic framing information. Subjects were largely unable to assess their decision accuracy. Results identify decision contexts that may produce poor decisions by professional buyers.

INTRODUCTION

Intuitively, how good are individuals with respect to quantity aspects of decisions, for example in purchase decisions where judgment involves deciding "how many" of an item to purchase? What decision processes bias these quantity decisions? Sophisticated decision-making tools are rarely employed when individuals make quantity decisions. Even when formal methods are used, for example when universities decide how many students to enroll, results are often adjusted to account for intuitive judgments. Thus, answering the decision accuracy questions above is needed to develop recommendations aimed at improving performance of decision makers and parties who transact with decision makers. To clarify, for this paper intuitive judgments are viewed as judgments based on informally mentally simulating plausible outcomes that would be related to different decision quantities (e.g., Simon 1976).

Unfortunately, assessing decision accuracy is difficult, largely because identifying an optimal decision quantity is elusive. Indeed, when everyday consumer decisions related to how much of an item to purchase are considered, an economic perspective suggests that *whatever* quantity is purchased is utility maximizing, thus any quantity decision is good. For example, in discussing the circular nature of definitions used for utility, Thaler notes, "If someone does something, no matter how odd it may seem, it must be utility maximizing" (1992, 2).

The Marketing Management Journal
Volume 11, Issue 2, pages 58-66

Helpfully, optimal quantities exist in certain contexts, such as professional settings where retail agents buy quantities of inventory that will later be resold. Optimal solutions for inventory replenishment decisions are based on Economic Order Quantity models and probabilistic Markov models (e.g., Pinder 1995), and recommendations have been developed to help practitioners use these models (Bassin 1990). Hence, to gauge decision accuracy, the approach followed here is to identify a context in which an optimal solution exists, and to then compare optimal decisions to intuitive decisions. A retail inventory purchase context is useful because retail purchase decisions are both common and important to managers. Also, many professional purchasing situations are similar to this retail context, for example raw materials purchasing or supplies purchasing, making this context relevant to a variety of non-marketing managers. Finally, the retail context is similar in many ways to consumer purchasing, for example by being affected by promotions (Blattberg, Eppen and Lieberman 1981; Gupta 1988). Thus, results obtained from the study of a retail context should provide at least some generalizable insights into the accuracy of purchase quantity decisions in contexts that lack clear normative solutions.

PURCHASE QUANTITY ABILITIES

Three studies have examined issues close to the quantity-accuracy question posed above. These studies yield mixed conclusions, yet collectively are quite useful in suggesting why quantity decisions deviate from optimal solutions. First, Meyer and Assunção (1990) examined stockpiling in the context of varying prices, where purchasers made sequential purchase quantity decisions to "stock-up" as protection from later price increases. Meyer and Assunção conducted experiments, comparing individual quantity decisions to optimal decisions while manipulating the trend and distribution of prices presented to subjects. Subjects comprised 35 students who progressed quickly through multiple simulations, each making a total of 270 decisions within 35 to 60 minutes. Meyer and Assunção found that subjects generally overstocked (i.e., stockpiled too much in comparison to optimal stockpiling). Understocking occurred in one special case, which was when a price trend pushed optimal quantities higher, and subjects increased purchase quantities, but at a rate that failed to adequately meet the trend.

Cripps and Meyer (1994) examined sequential purchases, using a simulation that challenged student subjects to make decisions replacing expensive production equipment. The focus of this study was on the timing of equipment replacement. A major finding was that subjects chose to replace production equipment at a rate that was slower than normative. The finding means that, across the set of decision periods, subjects purchased fewer-than-optimal new machines or tended to underbuy, a reversal from Meyer and Assunção's findings. The underbuying result was found to be consistent across a variety of manipulations. When considering the question of "how good" individuals are in making purchase replenishment decisions, the authors found that "subjects made purchases at performance lags that were 30 percent larger than those that were prescribed by optimal theory, which suggests that the underreplacement bias was both real and often quite sizeable" (Meyer and Assunção 1990, 309-310).

Finally, Cox and Summers examined whether individuals exhibit an intuitive regression-toward-the-mean effect in quantity decisions. Thirty-one experienced retail buyers were presented with historic data for retail handbags and estimated future sales. These forecasts were compared to regression-based forecast solutions. A major finding was that subjects did adjust handbag market share forecasts toward the mean market share provided by the historic data, but that this adjustment was much smaller than the adjustment that would be made by a regression model. That is, subjects were prone to overemphasize sales information from the current period and were not adequately adjusting forecasts to account for a likely regression toward the mean. Cox and Summers also found that "experienced retail buyers display similar biases and use the same types of heuristics as found for naïve subjects in previous psychological research on human prediction" (Cox and Summers 1987, 295).

These prior results suggest two mechanisms that may bias quantity decisions. The studies reported by Meyer and Assunção (1990) and Cox and Summers (1987) entail trend contexts, where pricing or forecast information rises or falls systematically. Quantity inaccuracies found in these studies may be due to a context of changing prices (Meyer and Assunção 1990) or a context that encourages overemphasis of recent information (Cox and Summers 1987), but suggest the anchoring-and-adjustment bias may be broadly relevant to quantity decision-making. Anchoring-and-adjusting is a process by which a decision-maker makes an initial evaluation and subsequently receives information, insufficiently adjusting, so that the decision outcome is overly entrenched in the initial evaluation (Russo and Schoemaker 1989), and has been applied to other purchase quantity research questions (Wansink, Kent and Hoch 1998). The second theoretical issue advanced by these earlier studies is a sunk-investment effect. The underbuying found by Cripps and Meyer (1994) was noted in a context where subjects made decisions to spend hundreds of thousands of dollars repetitively on equipment purchases. This context may prompt subjects to focus predominantly on costs, causing a bias similar to what is termed the sunk-cost or sunk-investment bias (Arkes and Blumer 1985), which has been found to bias marketers' decisions in other decision-making areas, such as product development (Schmidt and Calantone 1998). Thus, the underbuying found by Cripps and Meyer may be due to the cost-sensitive context of their experiment and may provide only a portion of the answer to whether individual decision makers are inherently good at making quantity decisions.

METHODOLOGY

An experimental Visual Basic simulation was developed. The simulation presented subjects with recurring inventory replenishment decisions. Each subject completed a practice round followed by two rounds of the simulation, with six decision periods per round. In each experiment, subjects were provided with specific cost-penalties of overbuying, in the form of inventory carrying costs, and cost-penalties of underbuying, in the form of lost profit that would occur if demand exceeded inventory. Subjects' quantity decisions were compared to a linear programming solution that determined the optimal order quantity given the distribution of demand and the overbuying and underbuying penalties.

Figure 1 depicts the main screen presented to subjects and is taken from Experiment 2. The top portion of the screen shows subjects a four-period distribution of possible demands. To encourage subjects to mentally simulate outcomes related to purchase quantities, the testing mechanism permitted subjects to click on possible demands (high, medium, or low) for upcoming periods, and these demands then populated the "actual demand" spreadsheet cells (shown in the bottom portion of Figure 1). Moving the slider below the spreadsheet caused potential order quantities to be inserted into "planned receipts" cells. Thus, subjects could informally test different demand-order combinations to evaluate potential overstocking and stock-out conditions. All subject clicks were recorded, as were response times and answers to post-experimental retrospective questions.

Experiment 1: Description and Results

Subjects comprised six retail buyers and twenty-seven students. Students were business students taking coursework that included a component of professional purchasing and were advised that a question pertaining to the study would be asked on their next exam. A $20 award was offered to the student making the overall most accurate quantity decisions.

Demand forecasts were manipulated. High, medium, and low forecast amounts were displayed in the spreadsheet (see Figure 1), and their respective probabilities were presented in the text box above the spreadsheet. All demand levels ranged from 700 to 800 units per period (this range was modified during later experiments). An example scenario would be high demand of 800 units with a 25 percent likelihood of occurring, medium demand of 750 units with a 40 percent likelihood of occurring, and low demand of 720 units with a 35 percent likelihood of occurring. Probabilities were expressed in multiples of five to avoid confusing subjects, which may occur when subjects are presented with more complicated probability data (Friedman and Kelley 1998).

Results were as follows. A main question motivating this study was to evaluate decision accuracy. Subject quantity decisions were surprisingly accurate. Overbuying decisions occurred more frequently than underbuying decisions ($p<0.01$). However, 88 percent of all decisions were within 10 percent of the optimal decision, and 93 percent of the decisions were within 30 percent of the optimal decision (see Figure 2). Four other results are notable. First, the pattern of professional buyer subject results was very similar to results obtained from student subjects. Second, across-subject variation was minimal, thus results are not attributable to a small number of outlier subjects especially prone to overbuying. Third, performance was relatively stable across the two simulations. Specifically, decisions made during the starting and ending periods were not unusual, and the six decisions made during the second-round were not generally

FIGURE 1
Primary Screen Viewed by Subjects

45% chance each week HIGH demand
35% chance each week MEDIUM demand
20% chance each week LOW demand

Profit:
$4/unit

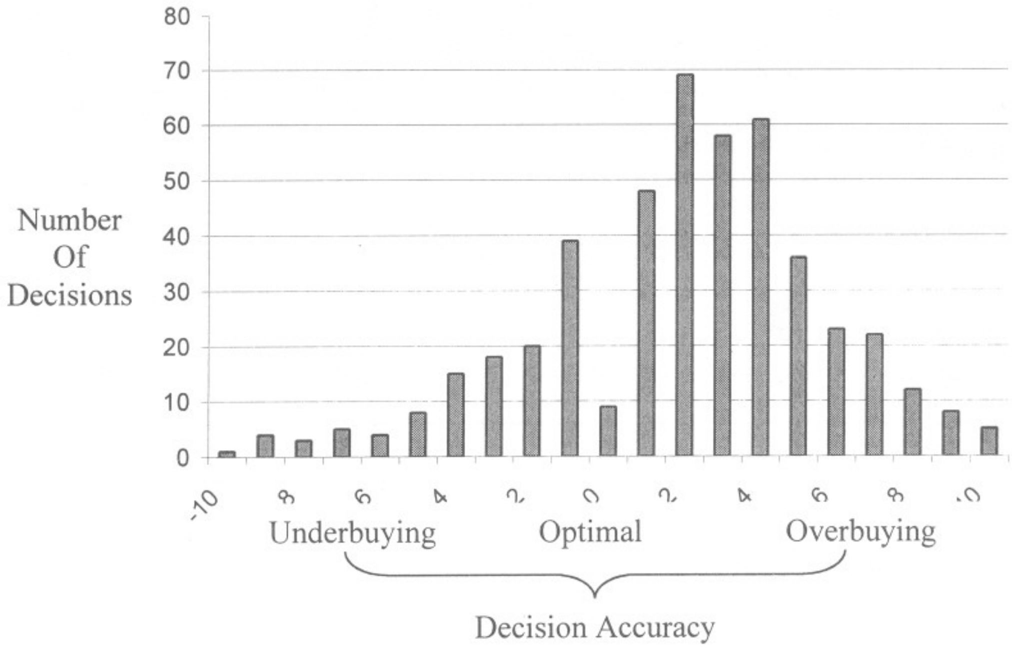
Holding cost:
$1/unit

High demand	800	832	865	893				
Medium demand	780	812	841	872				
Low demand	756	788	821	850				
Week	1	2	3	4	5	6	7	8
Starting inventory	39	16	16					
Planned receipts	757	812	0					
Demand	780	812	841					
Ending inventory	16	16	-825					

ORDER

FIGURE 2
Experiment 1 Frequency of Overbuy and Underbuy Decisions.

Figure 2: Experiment 1 Frequency of Overbuy and Underbuy Decisions. The center bar in the histogram represents decisions that were exactly optimal, thus the frequency of these decisions is low compared to the other bars, which reflect two-percent intervals.



more accurate than the six decisions made during the first-round. And fourth, comparing actual decisions to a post-experimental question, the correlation between actual performance and self-assessed performance across subjects was 0.021, $p>0.20$ (subjects were asked to gauge their overbuying/underbuying). Thus, it seems intuitively difficult for decision makers to assess themselves in terms of the extent to which they overbuy.

Experiment 2: Description and Results

Forty-nine students participated and were offered the same incentives provided to students participating in Experiment 1. The only changes made for Experiment 2 was that the 700-800 range of possible demands was removed, and a linear trend of 30 units-per-period was added to forecasted demand levels. Subjects faced both rising and falling trends with conditions counterbalanced across subjects. By comparing subject orders to the trend of 30 units per week, subjects were assessed in terms of whether they lagged the trend, as would be suggested by an anchoring and adjustment effect.

To assess whether an anchoring-and-adjustment process biased quantity decisions, subjects' responsiveness to the trend was evaluated. In the rising demand scenario, subjects ordered an average of 27.6 extra units during each successive decision. In the declining demand scenario, subject orders dropped an average of 26.4 units during each successive decision. As expected based on an anchoring and adjustment rationale, subjects responded to the 30 unit/week increasing and decreasing trends, but responses did not fully reach the trend amount ($p<0.05$ comparing the actual versus optimal response to trends for both rising declining comparisons).

Experiment 3: Description and Results

Experiment 3 was designed to test two additional quantity-bias mechanisms, a framing bias motivated by the sunk-investment effect discussed in the review of Cripps and Myer's (1994) study, and a cue-compatibility bias. The cue-compatibility bias has been investigated in other decision contexts. For example, in the study of preference reversals, subjects have been found to respond most readily to information-cues that are compatible in format to the format of decisions (Slovic and Lichtenstein 1968; Tversky and Slovic 1990). In the present case, the optimal quantity is an amount rooted in future demand, and future demand contains possible demand amounts and probabilities. The decision outcome is an amount, expressed in units, which is more compatible with the forecast amount rather than the probability that an amount will actually occur.

Sixty-three students participated in exchange for extra credit and a \$40 incentive was offered to the individual having the most accurate overall performance. A 2x2 experimental design was employed. The first manipulation was a between-subject framing manipulation. During each decision round, the top portion of the main screen advised subjects of either the maximum potential costs (or revenues) that could be realized given projected demand and the prior two decisions. Pilot testing revealed that this manipulation significantly affected the degree to which subjects reported a motivation to focus on potential costs or potential revenues. The second manipulation was a within-subjects manipulation of demand-variance. The two simulations entailed a combination of low variance of demand amounts coupled with high variance of demand probabilities, and then high variance of demand amounts coupled with low variance of demand probabilities (counterbalanced across subjects in the order of presentation). The cue-compatibility bias suggests that subjects can incorporate variance in amount changes more easily than variance in probability changes, and hence that decision-accuracy might remain higher as demand levels change, but become worse as probabilities become more varied.

Results were as follows. A t-test was used to compare decisions of subjects assigned to the cost frame to decisions of subjects assigned to the revenue frame. All twelve decisions for each subject were aggregated so that one score was calculated for each subject. Mean overbuying among the 32 subjects in the revenue frame was 51.4 units per period. In contrast, mean overbuying among the 31 subjects in the cost frame was 41.7. This cost-revenue framing effect difference was significant ($p < 0.05$). Further testing was conducted to evaluate the impact of changing demand amounts and probabilities on decision-accuracy. Each quantity decision was compared to the optimal decision and the absolute value of the difference or error was recorded. For demand probabilities, the mean error made in the low variance condition was 42.8 units and the mean error made in the high variance condition was 52.5. Here the difference is statistically significant ($p<0.05$). This result indicates that errors become worse as demand probability variance increases, indicating that subjects failed to adequately incorporate probability information. The error difference was not significant when comparing low versus high variance of demand amounts, indicating that subjects were more able to incorporate changes in demand amount information. Thus, a bias emerged that supports the notion that cue-compatibility is an additional mechanism that affects quantity decisions.

DISCUSSION AND IMPLICATIONS

The results of this study add to the understanding of individual abilities and biases with respect to quantity decisions. Table 1 summarizes the main points discovered as a result of this research. A main finding from this set of experiments is that individual quantity decision-making intuitions are generally quite good. This finding is a contrast to findings reported by Cripps and Meyer (1994). Across three experiments in the present study, subjects tended to overbuy more than underbuy, but over 85 percent of the decisions were within 10 percent of the optimal decision and very few decisions diverged by more than 30 percent of the optimal decision.

In evaluating mechanisms that might bias quantity decisions, several types of biases can now be more clearly generalized to the purchase quantity arena. The anchoring and adjusting bias indicates that decision-makers do adjust quantity decisions in response to diagnostic data, but that the response is less than optimal. The cue-compatibility finding indicates that bias mechanisms exist where subjects essentially do not adjust in response to changes in diagnostic information. And finally, the framing effect bias indicates that subjects do adjust quantity decisions in response to non-diagnostic information.

Another interesting finding was that individuals are largely unable to correctly assess the extent to which they overbuy. The question of whether decision makers are good in sensing their quantity biases was not explored in the prior studies, and poor self assessment suggests that decision makers are unlikely to improve their own decision making skills without external interventions. In terms of using simulations as a tool for understanding decision-making, one weakness inherent in simulations is that multiple decisions are treated as independent pieces of data. Particularly in simulations involving a large number of decisions and between-subject analysis, subjects in a simulation may make one general decision akin to an ordering policy during the start of the simulation, and then simply execute this decision repeatedly. To hedge against this problem, the simulations used here involved only six decisions per simulation iteration, and some analysis was done at the subject level so that decisions were aggregated. A key benefit in using simulations is the ability to unobtrusively measure variables. For example, response times and degree of scenario testing (reflecting mental simulation) were recorded and evaluated in the present study, and such variables are important to examine when testing decision-making process issues.

Decision-makers should be made aware of their tendencies to overpurchase, to under-respond to trends and to purchase probability cues, and to over-respond to non-diagnostic information. The results reported here suggest that an important avenue for future work would be to design feedback mechanisms that would help decision-makers gauge their purchasing abilities. Another beneficial area of pursuit may be to develop intervention techniques that could improve intuitive judgments. In terms of the pragmatic value associated with improving intuitive quantity decision accuracy, the retail buyers who participated in Experiment 1 discussed software tools that can provide professional buyers with optimal solutions. However, they also noted that retail buyers most typically use output generated by these tools as a starting point, making intuitive adjustments. Thus, understanding intuitive quantity-biases may be quite relevant as a step toward improving retailer purchase performance.

TABLE 1
Summary of Experimental Findings

Experiment	Phenomenon studied	Results
One	General accuracy	Overbuying dominated underbuying. 93% of quantity decisions were within 30% of the optimal decision and 88% of decisions were within 10% of the optimal decision.
One	Self-assessment	The correlation between overbuying and self-assessed overbuying was 0.021. Subjects in general were unable to accurately assess their overbuying.
Two	Anchoring and adjustment	Subjects under-responded to diagnostic information. Upward and downward trends in demand volume were provided. Subjects adjusted in response to trends, but adjustments were insufficient compared to optimal adjustments.
Three	Cue compatibility	Subjects ignored some diagnostic information. Subjects responded to high variance in demand amounts, but failed to respond to high variation in demand probabilities.
Three	Framing effect	Subjects responded to non-diagnostic information. Overbuying significantly increased (decreased) as subjects focused on revenue (cost) aspects of the decision.

REFERENCES

Arkes, Hal, and Catherine Blumer (1985), "The Psychology of Sunk Costs," *Organizational Behavior and Human Decision Processes*, 35(1), 124-140.

Bassin, William (1990), "A Technique for Applying EOQ Models to Retail Cycle Stock Inventories," *Journal of Small Business Management*, 28(1), 48-55.

Blattberg, Robert C., Gary D. Eppen, and Joshua Lieberman (1981), "A Theoretical and Empirical Evaluation of Price Deals for Consumer Nondurables," *Journal of Marketing*, 45 (Winter), 116-29.

Cox, Anthony, and John O. Summers (1987), "Heuristics and Biases in the Intuitive Projection of Retail Sales," *Journal of Marketing Research*, 24(August), 290-297.

Cripps, John D., and Robert J. Meyer (1994), "Heuristics and Biases in Timing the Replacement of Durable Products," *Journal of Consumer Research*, 21(September), 304-318.

Friedman, Dan, and H. Kelley (1998), "Learning to Forecast Price," Working Paper #394, Department of Economics, University of California Santa Cruz.

Gupta, Sunil (1988), "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research*, 25 (November), 342-355..

Meyer, Robert J., and Joao Assunção (1990), "The Optimality of Consumer Stockpiling Strategies," *Marketing Science*, 9(Winter), 18-41.

Pinder, Jonathan P. (1995), "An Approximation of a Markov Decision Process for Resource Planning," *Journal of the Operational Research Society*, 46(7), 819-830.

Russo, J. Edward, and Paul Schoemaker (1989), *Decision Traps*, New York: Fireside.

Schmidt, Jeffrey B., and Roger J. Calantone (1998), "Are Really New Product Development Projects Harder to Shut Down?" *The Journal of Product Innovation Management*, 15(2), 111-123.

Simon, Herbert A. (1976), *Administrative Behavior: A Study in Decision-Making Processes in Administrative Organizations* (3rd Edition), New York: The Free Press.

Slovic, Paul, and Sarah Lichtenstein (1968), "Relative Importance of Probabilities and Payoffs in Risk Taking," *Journal of Experimental Psychology Monograph*, 78 (November), 1-18.

Thaler, Richard H. (1992), *The Winner's Curse: Paradoxes and Anomalies of Economic Life,* Princeton, NJ: Princeton University Press.

Tversky, Amos, and Paul Slovic (1990), "The Causes of Preference Reversal," *American Economic Review*, 80 (1), 204-217.

Wansink, Brian, Robert J. Kent, and Stephen J. Hoch (1998), "An Anchoring and Adjustment Model of Purchase Quantity Decisions," *Journal of Marketing Research*, 35 (February), 71-81.