Incorporating Bounded Rationality into Dynamic Inventory Control: Experiments and Theory

ABSTRACT

In the traditional dynamic inventory control theory, decision makers are assumed to optimize their expected payoffs considering the full time horizon of planning. We use a series of laboratory experiments to study the impact of human decision making on inventory management. The study focuses on the classical multi-period inventory setting, and considers the single-period counterpart, i.e., the newsvendor model, to benchmark empirical behaviors resulting from dynamic controls. In the experiments, we manipulate the inventory control option that is available to human subjects: order quantity, price, or both. One major result we find is that, contrary to the optimization theory, more decision freedom does not necessarily lead to better inventory performance. We refer to this phenomenon as the "Optimizer's Paradox". To explain the behavioral departures, we develop a descriptive model based upon the probabilistic choice framework and incorporate decision makers' limited forward-looking capabilities in dealing with inventory dynamics.

Keywords: Inventory Management, Bounded Rationality, Dynamic Decision-making, Inventory Control, Probabilistic Choice, Behavioral Operations Management

1. INTRODUCTION

Traditional operations management theories assume optimization behavior. However, research in behavioral economics and social sciences has recognized the cognitive limitations in human decision-making (Simon 1955), and helped develop theories of "bounded rationality" to explain the resulting behavioral departures (Simon 1982, Rubinstein 1998). While human errors are often treated as random noises in theoretical modeling, sometimes they exhibit systematic and nontrivial patterns. For instance, Schweitzer and Cachon (2000) document that empirical newsvendor decisions by human subjects fall between the average demand and the optimal solution. Numerous behavioral studies since then attempt to understand the so-called "pull-to-center" effect. Other research discusses the impact of bounded rationality under different operations management contexts such as the economic order quantity (EOQ) problem (Chen and Wu 2019). We refer to Katok et al. 2018 for a thorough review of this fast growing field of behavioral operations management.

In this study, we examine empirical decision making under a dynamic operating system in which agents manage the inventory of a single product over a finite number of periods. The demand of the product is stochastic and price-dependent, and its inventories can be controlled by adjusting the selling price and/or replenishment amounts. We are interested in this particular setting for several reasons. First, it is one of the corner stone scenarios of inventory theory with well-understood analytical solutions. Predictions from models assuming perfect rationality can serve as normative benchmarks for empirical behaviors. Second, while many operations management studies focus on ordering decisions, in industries such as retailing, pricing and stocking decisions are often made jointly. We believe it is more practical and relevant to explore decision-making process given different control options. From the management perspective, we are primarily concerned with the impact of suboptimal behaviors on inventory performance because of potential inefficiency it costs.

Bounded rationality has been shown to be interdependent on the complexity of a decision task (Johnson and Paney 1985, Ho and Weigelt 1996). Intuitively, the level of human errors is related to how difficult the problem is and how complicated the environment to interact with. Therefore, we use "decision complexity" as the main treatment variable in experimental design. Under our baseline condition, decision makers face a multi-dimensional decision task, deciding the price and order quantity jointly over multiple periods. We manipulate the decision complexity by restricting the inventory control options that are available: stocking decisions only (with selling price fixed) versus pricing decision only (with replenishment amount fixed). The scenario of our interest also highlights a conundrum in the literature. On one hand, a number of analytical studies tout the need and economic benefits in integrating pricing and replenishment planning (Thowsen 1975, Federgruen and Heching 1999, Pertuzzi and Dada 1999). On the other hand, we observe quite different practices where pricing and ordering decisions are made separately. It is well known that Walmart seldom offers promotions and locks in its retail prices for relatively long period. Under its "Every Day Low Price" business model, managers of Walmart make less frequent price adjustments. In fact, Feng and Gallego (1995) provide more evidence that, even in industries that adopt dynamic pricing policies, many companies still restrict the number of price changes during the selling horizon.

Moreover, when dynamic responses from decision makers are required, degrees of complexity can increase. For example, using the Beer Distribution Game that simulates a supply chain process, Sterman (1989) showed that subjects tend to underestimate delayed feedback in the system due to lead times, triggering costly ordering cycles of the "bullwhip" effect. In order to identify and isolate the effect of dynamics on empirical behaviors, we also include the single-period newsvendor model as another test bed in the study. In the experiments, we observe that subjects tend to underprice the product; and when restrained from setting the price, they tend to understock. This results in significant efficiency loss under the dynamic inventory control setting. In addition, we find that having more options to manage inventory does not always lead to better profit outcomes. It is contradictory to what the optimization theory implies. An intuitive explanation for such an observation is that more decision freedom can lead to more rooms for decision makers to err. Particularly in a dynamic system, mistakes can propagate and be more difficult to correct. We call this phenomenon as the *Optimizer's Paradox*.

We propose a descriptive model to explain the observed behaviors. The model is motivated by two key insights from the experiments. The first is that decision makers make mistakes when they evaluate potential choices. To model this tendency, we apply the framework of probabilistic choice (Luce 1959, McFadden 1976: decision makers do not always choose the best payoff choice but select more attractive decisions more frequently. Second, decision makers have imperfect forward-looking abilities when dealing with dynamics. In the traditional dynamic control theory (Bertsekas 1976), agents are usually assumed to perfectly anticipate the impact of their current decisions on future ones (at least in expectation), and to reason backwards in time from the end of the planning horizon. However, past experimental studies demonstrate that human subjects have difficulties in managing dynamic systems without delays (see Busemeyer 1999 for a review). Under the model we propose, decisions makers do not always evaluate decisions over the whole time horizon while computing utilities. Instead, depending on their ability to look forward, they may consider limited number of periods ahead and weigh them with varying importance. The standard dynamic control theory is therefore a special case of our behavioral model when the level of human errors approaches zero and the number of periods considered by the decision maker approaches the true horizon.

The behavioral model predicts that, with the same degree of bounded rationality, decisions and corresponding profit performance would vary by the availability of different inventory control options. These predictions are consistent with our lab observations and help shed light on the design of inventory management system. Moreover, estimations of the model lend more supports to the limited forward-looking behaviors under the dynamic setting. We note that the proposed model is solely motivated to describe human decision-making behavior. Results may certainly change if Artificial Intelligent software or other decision support systems are adopted. However, there are business cases where no more sophisticated tools other than spreadsheets are available to managers, or the relevant information can only be obtained via human expertise and reasoning. Our study highlights, from the behavioral perspective, the potential benefit of restricting decision flexibility as a necessary management strategy. The proposed behavioral model provides a framework on how a behaviorally optimal business plan should be constructed (e.g. how to determine the frequencies for dynamic pricing).

The rest of the paper is organized as follows. Section 2 discusses the relevant literatures. Section 3 describes experimental settings and reports observational results. Section 4 presents a general behavioral modeling framework and its implications. Section 5 discusses behavioral estimations from the proposed model. We offer concluding remarks and discuss directions for future studies in Section 6.

2. LITERATURE REVIEW

Two streams of literatures are relevant to the research reported. The first is the area of inventory management theories and the second is behavioral economics, and its applications to operations management.

One simple but fundamental building block of inventory theory is the newsvendor model, in which a decision maker determines the order quantity for selling a perishable product before some random demand realizes. Traditionally, market parameters such as selling price and demand function are exogenous, and thus the objective is to minimize the expected costs under the single-period model without carrying over inventories. The optimal newsvendor solution involves the computation of the *critical fractile* that balances the overage and underage costs. Porteus (2002) provides a detailed review of such models. Pertuzzi and Dada (1999) extend the newsvendor framework to analyze a scenario where a decision maker sets a selling price and an ordering quantity simultaneously, and the stochastic demand is dependent on the pricing decision either in an additive or multiplicative form.

Another large category of inventory control theory considers leftove to be carried and shortages to be backlogged over multiple periods dynamically. Previous research shows that there exists an optimal stationary strategy for such settings in general. More specifically, a base stock policy should be administered to bring the inventory level (or the inventory position when lead-time is positive) up to a predetermined optimal stock level *S* in each period (Porteus 1990, Federgruen 1993). Furthermore, if the ordering/production cost function is linear, the optimal policy is a myopic one, in which *S* is the solution to the single-period newsvendor problem (Veinott 1965). Similarly, many studies attempt to address the need to integrate pricing and replenishment planning. Under certain conditions, a base stock list price policy is proven to be optimal (Thowsen 1975 and Federgruen and Heching 1999). Given such policy, if the inventory level is below *S*, orders should be placed to bring the inventory to the base stock level and the list price should be charged. If the inventory level is above S, nothing is ordered and the price should be discounted.

Classic works in inventory control theory illustrate the value of optimal decision making. However, behavioral studies point out inconsistencies between predicted and empirical decisions under various contexts. Bendoly et al. (2006), Gina and Pisano

(2008), and more recently, Katok et. al (2018) provide reviews of behavioral research in the operations management literature. Schweitzer and Cachon (2000) are the first to examine the newsvendor decisions experimentally. They find that subjects systematically order too many low-profit-margin products and too few high-profit-margin ones. In other words, ordering decisions are biased toward the mean of the random demand. Su (2008) applies the discrete choice framework to model the newsvendor's bounded rationality. Such a model is capable to explain the "pull-to-center" bias observed in many experiments. Lim and Ho (2007) test the performance of various contracts experimentally in which subjects act as a price-setting newsvendor. They show that a quantal response equilibrium model helps explain the observed behavioral interactions. We refer readers to the Handbook of Behavioral Operations Management (Chapter 10 by Becker-Peth and Thonemann 2018) for a more detailed discussion of recent research on the single-period inventory decisions using experiments or behavioral modeling.

In the dynamic operations settings, sub-optimal performance other than solutions suggested by dynamic programming has long been documented. Rapoport 1966 and 1967 find that subjects in stochastic multistage decision tasks could only plan a few steps ahead, as compared with the optimal model that assumes an unlimited planning horizon. Rapoport (1975) proposes an approach to model these anomalies by adding information processing constraints on the planning process, or by including subjective utilities into the objective function. Bearden et al. (2008) study a dynamic pricing problem in the laboratory. They observe that decision makers employ strategies of the same form as the optimal policy; however, they exhibit systematic biases that lead to less revenue than they could achieve if a simple heuristic is used. In the context of dynamic inventory control with delays, Sterman (1989), Diehl (1992) and Diehl and Sterman (1995) show that decision makers have difficulty in controlling systems with lagged and indirect feedback. Subjects generally fail to appreciate time delays between action and response and therefore misperceive or ignore feedback processes. Such misperceptions of system dynamics cannot be fully corrected by training or communication (Wu and Katok 2007). Studies in this area often adopt the experimental platform of the beer distribution game.

Perera et al. (2020) present a most recent review of behavioral research on ordering and inventory decisions. They classify the literature according to the decision context: single-period, multi-period and the beer game. The survey notes that the beer game can be a noisy environment to study the behavioral impact of dynamics since there exists multiple players and lead times. They also point out that research on multi-period ordering decisions is in a great need to build new theory of inventory management and to offer practical guidance. In this study, we attempt to bridge the literature gap by contrasting empirical decisions under the multi-period setting with those under the single-period one. To rule out the known effect of delayed feedback on behaviors, we control replenishments to be instantaneous (i.e., no lead time) in the experiments. This design also helps create clean environment for us to focus on the issue of decision complexity.

3. EXPERIMENTS WITH HUMAN SUBJECTS

3.1 Experimental Design and Implementation

The general inventory management setting that we used considers a single decision maker who plays the role of a retailer. The retailer buys units of one fictional product from an automated supplier without any lead-time, and sells the product to meet the stochastic customer demand (*D*). In designing treatments, we first control whether or not leftover inventory and unmet demand can be carried over to the next period (*Dynamic* setting versus *Newsvendor* setting). We next vary the type of inventory control options that a subject can use in the game: both the price and the order quantity (*PQ*), the selling price only (*Ponly*), or the stocking quantity only (*Qonly*). This leads to a 2 x 3 full factorial design of experiment. In all six treatments, the game lasts for 36 periods and subjects make ordering decisions in integer every period. Table 1 summarizes our experimental design, including the sample sizes we collected for each treatment.

Table 1: Summary of Experimental Design.

Treatments	Restrictions on Inventory Control Option	Inventory Settings	
		Dynamic	Newsvendor
PQ	None	27	21
Ponly	Fixed replenishment with $q = 6$	23	19
Qonly	List price with $p = 13$	29	23

In the experiment, we use a random price-dependent demand function in an additive form to simulate the customer demand per period. Namely, $D = (15 - p) + \varepsilon$, where ε is a random integer that follows a uniform distribution from 0 to 8. The same sequence of random demand draws is used in all experimental sessions. The inventory holding and shortage costs are set at $c_h = 2$ and $c_b = 4$ per unit, respectively, and the transfer price that the retailer pays for each unit ordered is fixed at $w = 8$. We choose the above parameter values so that the resulting critical fractile in the newsvendor setting is at 50%. This allows us to tease out the "pull-to-the-center" effect identified in the literature, and thus help isolate behaviors induced by decision restrictions and/or dynamics that are of our interest. The chosen parameter settings also keep certain features of the Beer Game, such as the uniform demand function and the relationship between the backlog and inventory costs (i.e. $c_b = 2 \times c_h$).

The experimental scenarios discussed above have been well analyzed by classical models in inventory management. Under the newsvendor setting, we apply results from Pertuzzi and Dada 1999 and obtain the optimal price and order quantity to be $p^* = 13$ and $q^* = 6$, respectively. According to Thowsen (1975) and Federgruen and Heching (1999), a *stationary base stock list price policy* is optimal under our dynamic setting. Under such a policy, if the inventory level is below a predetermined stock level s, orders should be placed to bring the inventory to the base stock level and a fixed list price should be charged. If the inventory level is above s, nothing is ordered and the price should be discounted. Given the specific parameters used in our experiments, the optimal list price is $p^* = 13$ and the optimal order-up-to level is $s^* = 7$.

In treatments where we restrict the inventory control options, the retail price per round is fixed at 13 in Qonly; and replenishment per round is set to be 6 in Ponly. Note that the pricing decision is at its optimal level under the Qonly treatments. Subjects are expected to place orders of 6 all the time without deviations in the newsvendor setting. Under the dynamic setting, they are supposed to bring their inventory levels up to *7* every period, resulting in an average order quantity of 6 (i.e. mean of the uniform demands given $p = 13$) and a decision variance equal to that of the random demand. As for Ponly treatments, $q = 6$ per period is optimal in the newsvendor model, and the normative theory predicts subjects to respond with a fixed price of 13. Under the dynamic setting, however, $q = 6$ is suboptimal since subjects should follow the base stock policy. In fact, the corresponding decision making scenario becomes similar to the dynamic pricing problem: the price should be determined based upon the inventory level instead of a fixed one. We purposely design this particular treatment for several reasons. First, since optimal strategies may not be always available in practice, we are interested in understanding how decision makers deal with such a situation. Second, we can use observations from this treatment as a stress test for predictions by our proposed behavioral model.

The game is implemented in a software platform, called MUMS developed by the formal HP Laboratories. Each experimental session proceeds as follows. Written instructions are posted on a website at least three days before a scheduled experimental session (sample instructions are available upon request). Subjects are required to pass a web-based quiz before they can take part in the study. Upon their arrivals to the lab, we provide a demonstration of the game software. Subjects are then given the opportunity to practice for 4 rounds and ask any questions before the paid game starts. We inform subjects that the game lasts for no more than 50 periods while it stops at the 36^{th} period to avoid any end-of-game effect (Sterman 1989). After completing the experiment, subjects are paid in cash according to their own accumulative profits. Communication among participants in any form is strictly prohibited during the experiment.

All sessions were conducted at a large university in the United States. Participants are business school undergraduates, mostly juniors and seniors, recruited from several large introductory courses. Students were offered extra-credits for their participation and cash incentives for their performance in the experiment. To control for possible loss, we offer each subject an endowment of $5. Any loss incurred during the game is deducted from the $5. Subjects who lost the entire cash endowment at the end of the 36^{th} period were excluded from our data analysis (these subjects incurred losses mostly due to input errors such as p=0 in the game). Each lab session lasts for around 45 minutes, and the average earnings per subject is $12.

3.2 Observational Results

We treat each subject as an independent observation to compute three behavioral measures: the average and standard deviation of subjects' decisions made during the 36-period game, and their average profit per period realized. We report the medians of all samples in each treatment in Table 2. We use the Wilcoxon Signed Rank test for comparisons with theoretical benchmarks, and the Mann-Whitney U-test for across treatment comparisons.

Table 2: Experimental Results Summary.

Inventory Settings		Newsvendor				Dynamic			
Variables		Ponly	Qonly	PQ	Theory	Ponly	Qonly	PQ	Theory
Price	Avg.	12.81	*13*	12.19	13	12.67	*13*	12.56	13
	Std.	0.81	*0*	1.27	0	1.98	*0*	1.33	0
Order	Avg.	*6*	5.81	6.53	6	*6*	5.64	6.00	6
	Std.	*0*	0.94	1.44	0	*0*	2.72	3.04	2.78
Profit per period		4.25	4.78	2.06	8.89	14.86	20.39	16.5	20.75

Note: italicized numbers in the table are parameters fixed in the experiment.

Experimental Result 1: More options in inventory control do not necessarily lead to better profit performance.

We begin with a discussion of the most prominent treatment effect on profit comparisons. Under the newsvendor setting, subjects in PQ, where inventory can be controlled by both pricing and ordering decisions, receive a median profit of 2.06 per round, significantly lower than profits in Ponly (4.25) and Qonly (4.78) treatments (p-values < 0.01). The profit difference between Ponly and Qonly is not statistically significant (p-value = 0.3121). This result suggests that more options for inventory management in a single-period setting actually lead to worse performance empirically. Under the dynamic setting, Qonly results in higher profit than PQ (20.39 vs. 16.5, p-value < 0.001), and PQ outperforms Ponly significantly (16.5 vs. 14.86, p-value = 0.0323). Recall that the ordering policy in Ponly is suboptimal under the dynamic setting. The above result implies that such system errors can be more costly than human errors.

Experimental Result 2: Subjects tend to underprice in general; and when the selling price is fixed, they tend to understock.

We next explore behavioral causes for the observed efficiency loss. First, prices set by subjects in all treatments are significantly below the optimal price of 13 (all p-values <0.05). More specifically, under the newsvendor setting, the median price in Ponly is not statistically different from that in PQ (12.81 vs. 12.19, p-value = 0.1757). Standard deviations of the pricing decision goes up significantly from Ponly to PQ (p-value = 0.0105). Under the dynamic setting, similar levels of price are found in Ponly and PQ as well (12.67 vs. 12.56, p-value = 0.11), yet higher decision variance is associated with Ponly (p-value < 0.001). This result indicates that subjects indeed adjust the selling price more dynamically when the ordering decision is set to be suboptimal.

Examining the stocking decision, we discover that the median order quantity and its standard deviations in PQ are higher than those in Qonly under both the single-period and multi-period settings (all p-values < 0.01). The increased decision variations in PQ can be a consequence of subjects' effort in exploring over a two-dimensional decision space.

To better understand how subjects view the relationship between the two inventory control options, we plot the order quantities against prices by each individual in Figure 1 for Qonly and PQ along with the optimal benchmarks. Note that under the dynamic setting, the *order-up-to* levels (computed as the starting inventory plus the order placed) are used for comparisons. Visual inspection of the graph shows that when price is set optimally, decision makers tend to under-stock. The median order quantity in Qonly of the newsvendor setting is 5.81 (versus $q^* = 6$), and that in the dynamic setting is 6.44 (versus $s^* = 7$) – both significantly below the respective theoretical predictions (p-values < 0.05). In PQ of both inventory settings, underpricing behavior are also confirmed as large portions of observations are distributed to the left of $p = 13$. Aside from this bias, however, individual decision makers seem to qualitatively understand the relationship between pricing and ordering decisions, especially under the newsvendor setting. The observed median order/order-up-to levels given different prices are insignificant from the respective theoretical benchmarks (all p-values > 0.1).

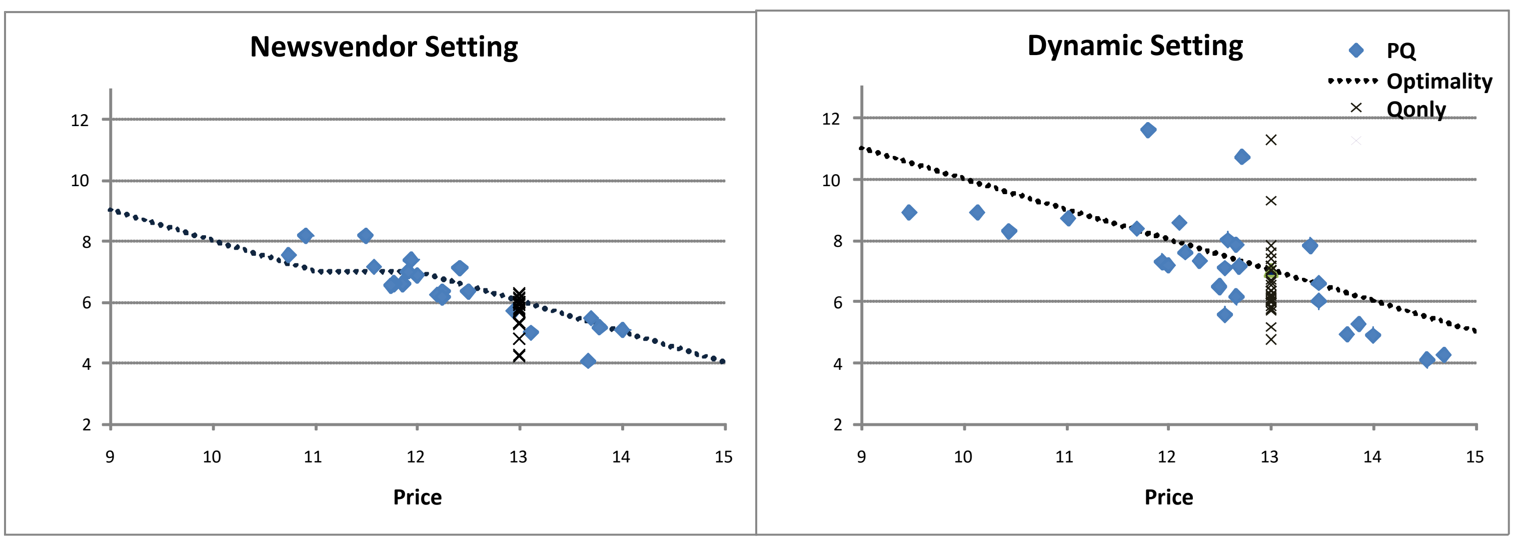

Figure 1: Order-Price Relationships.

Experimental Result 3: Subjects are not entirely myopic when dealing with inventory dynamics.

Under the single-period setting, subjects should make decisions based upon the expected payoffs in the current round. As for the dynamic case, we do not expect subjects to be perfectly forward-looking either to the end of the game (up to 50 periods as stated in the instruction), or treat it as an infinite horizon problem mathematically. At the same time, since decisions in the current period affect those later, we expect subjects to at least take some of the future payoffs into consideration. If a subject were entirely myopic when interacting with the multi-period inventory system, they would make decisions similar to those of the newsvendors. We thus compare treatments across the two inventory settings to test the conjecture.

First, we observe increased standard deviations of the ordering decision in the dynamic treatments in contrast with their newsvendor counterparts (2.72 vs. 0.94, p-value < 0.001 for Qonly; and 3.04 vs.1.44, p-value < 0.001 for PQ). This is consistent with the theory prediction that a base stock policy in the dynamic setting should generate larger order variance than a stationary optimal order quantity of the newsvendor model. Second, we find the standard deviation of the selling price in Ponly to be larger under the dynamic setting than the single-period case (1.98 vs. 0.81, p-value < 0.001). This again is in accordance with theory that subjects should price according to the current inventory levels instead of a list price. Both observations reveal that subjects respond to the dynamic environment by some decision rule other than the one used under a newsvendor setting. It provides indirect evidence that subjects are not entirely myopic.

Interestingly, if we assume that subjects adopt a heuristic that simply fixes the price at 13 under the dynamic Ponly treatment, an average profit of 17.42 per period would then be reached, greater than the current observed profit of 14.86. Thus, even though decision makers react to dynamics in a qualitatively correct way, their empirical performance is worse than the case where a suboptimal list price policy is applied. Thus, it is unlikely that subjects are perfect forward-looking.

To summarize, we identify systematic behavioral deviations from predictions by normative theories, especially that more inventory control options does not necessarily enhance profit performance. One possible explanation is that, for a boundedly rational agent, more freedom in decision making means more room to err. The experimental results provide some insights of where the theory breaks down. In the next section, we propose a behavioral model to explain these behavioral departures.

4. BEHAVIORAL MODEL

Our model is motivated by two general principles. First, people make mistakes. We incorporated this idea using the probabilistic choice framework (McFadden 1976, McKelvey and Palfrey 1995). Second, people do not look forward into the future perfectly when dealing with dynamics.

4.1 Probabilistic Choice with Partial Forward-Looking

The core idea behind the probabilistic choice framework is that individuals are subject to random errors while making decisions. We consider a representative agent who has the following utility function when evaluating an action *i* among possible alternatives:

$$U_i = \sum_{t=0}^{T} \beta_t (E\pi_t) + \varepsilon$$

where $E\pi_t$ is the expected profit at period *t*. $t = 0$ is set as the reference time period where an initial decision is made. ε is a random variable with zero mean, which can be interpreted as the error made in calculating the sum of the expected profit. β_t is a weight the agent assigned to future profits in period *t*. While it has a similar structure to the standard discounting factor, there is one key difference. β_t can decrease at a rate not constrained by the exponential decay. Thus, we do not restrict it to represent discounting based on the future value of money. In a laboratory experiment where subjects are expected to be paid within a short period of time, there is no true discounting. However, we still do not expect β_t=1. It is possible in this framework for the agent to place less weight on profits further away in the future because they may have less confidence of the validity of their estimation. Ultimately, the exact reason of this weighing is less important than the validation of its existence, as shown next in Section 5.

To make the model tractable, we make two further simplifying assumptions: 1) β_0 = 1 (without loss of generality), $\beta_1 = \beta$ and β_t = 0 for all $t > 2$. Thus, we reduce the model to two terms, one for the current period and the other for the next period. This is a strong assumption assuming that an agent does not consider what will happen beyond the immediate next period. The full dynamic optimization process is computationally intensive because of the combinatorial explosion of the possibilities. In addition, previous experimental studies suggest that human decision-makers would consider only few periods ahead. As we can also see in the later section, estimates of β from the two-period model using experimental data are significantly below 1, which helps justify this assumption. 2) ε is distributed with the extreme value distribution, which is a standard assumption used in choice models (McFadden 1976).

Under these assumptions, the choice probabilities are modeled by a multinomial logit distribution. Given specific parameter settings of our study, the action space is two-dimensional and consists of possible combinations of price and order quantity. Subjects were restricted to choose integer decisions within a certain reasonable range. Thus, modeling the action space as a finite set is adequate, however, the model can be generalized to continuous decision spaces (see Morgan 1992). Let i = 1... n be a finite set of possible actions. The probability that the decision maker chooses alternative i is given by:

$$p_i = \frac{e^{\gamma U_i}}{\sum_{i=1}^{n} e^{\gamma U_i}},$$

where Ui is the utility defined above and γ is a parameter related to the distribution characteristics of ε. At γ = 0, the agent randomly selects his/her decision with equal probability for all possible choices. As $\gamma \to \infty$, the agent picks the choice with the highest utility with probability 1. Hence, the traditional utility maximization model is a special case of the above probabilistic choice model. γ can be interpreted as the degree of rationality. An agent is completely irrational (random choices) at γ =0 and fully rational when $\gamma \to \infty$. In Figure 2, we plot three choice distributions given different levels of gamma for illustration. The probability distribution of the low gamma case is quite flat as the agent makes almost random decisions. As gamma increases, distributions become more concentrated on the utility-maximizing decision. Although the best alternative is no longer chosen with probability one, it is still the mode of the choice distribution.

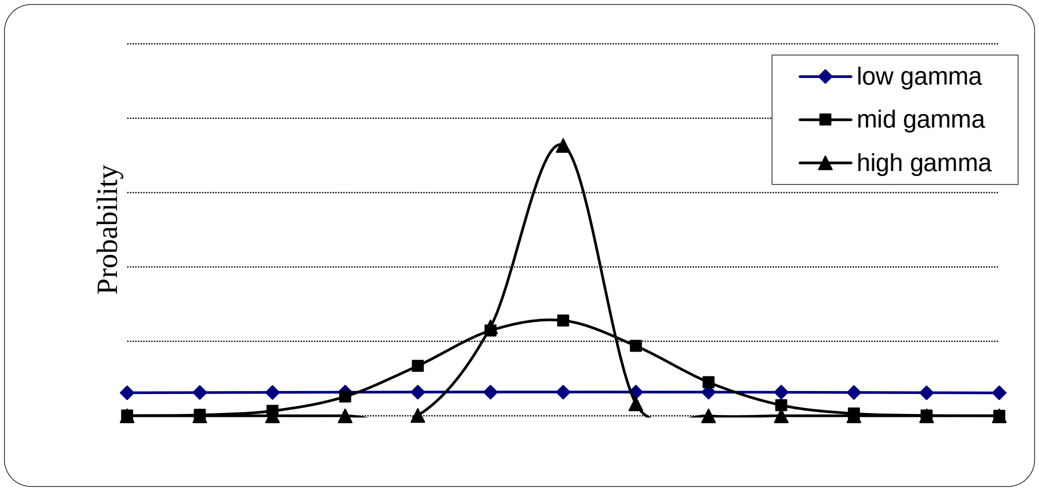

Figure 2: Probability Distributions of Decisions under Probabilistic Choice Model.

4.2 Modeling Implications

In this section, we summarize major implications from the behavioral model introduced above when applying to our game settings. Monte Carlo simulations are used to study the model under the newsvendor and the dynamic inventory settings. A "software" agent makes decisions in settings identical to our treatments (i.e., Qonly, Ponly and PQ). In every period, the agent calculates the expected profit for all potential decisions. All decisions were restricted to integers. Since prices have a finite range of [0, 15], an expected profit can be calculated for each possible price. Ordering decisions, on the other hand, are not bounded. Theoretically, as long as the decision space is not infinite, there is positive probability to choose even very large quantities. However, in practice, the probability of doing so diminishes quickly. We therefore placed an upper bound on the quantity decision in the simulation. The bound was chosen so that it would not impact the final results. Each simulation run lasts for 36 periods, and 500 simulations are used to calculate the average behavioral performance given different β and γ. For illustration purpose, we report simulation results from the behavioral model in which β is set to be zero. In this case, the agent is myopic who computes the expected profit of the current period only. While actual decision distributions will change when β is positive, the comparative static results shown below are robust.

We first examine the profit performance. Figure 3 plots the average period profit per round under each treatment for $\gamma \in [0, 1]$ (the range was chosen in accordance with gamma estimations from the experiments reported in Section 5). We can see that profits increase with γ in all treatments. The behavioral model suggests that when γ increases, the deviations from the optimal policy reduce and thus profits increase. Note that differences in profit performance decrease as γ increases among all treatments, which is consistent with the idea that when agents become more rational, behaviors would converge to the optimal solutions by the normative theory.

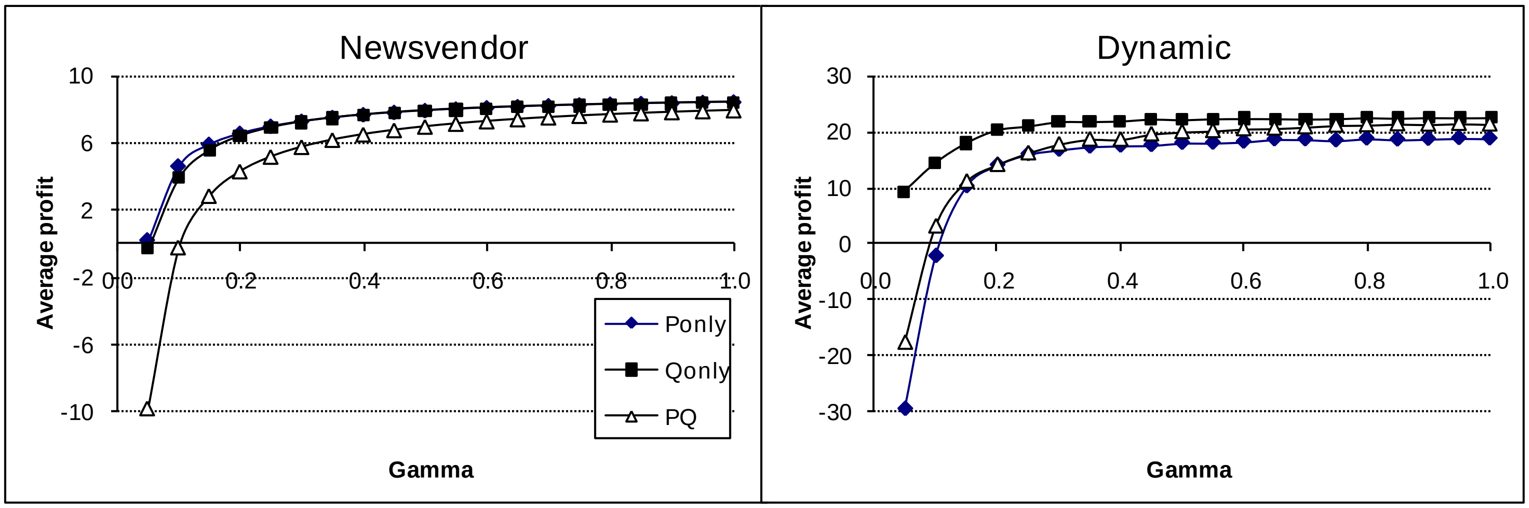

Figure 3: Profit Performance as a Function of Gamma Predicted by the Behavioral Model.

Modeling Result 1: The "Optimizer's Paradox": restrictions on inventory control options can lead to higher profits.

Consistent with Experimental Result 1, the behavioral model predicts that reduced decision freedom can improve profits. This is counter to the optimization theory that states more decision freedom will never decrease performance. More specifically, in the newsvendor setting, the behavioral model predicts similar profit performance in Qonly and Ponly, and both of which strictly dominate the performance of PQ for all values of gamma. In the dynamic setting, profits in Qonly are always higher than that in PQ. The Ponly case, however, is predicted to perform worse than PQ. This is driven by the fact that the ordering policy is set to be suboptimal with a constant order (of 6). The ordering of profit comparisons amongst various treatments is identical to what we observed in the experiments.

We next look at the pricing decisions implied by the behavioral model with a focus on the dynamic setting. Figure 4 shows the simulation results on the average and standard deviation of the selling price as a function of γ for the dynamic treatments. Qonly, where the pricing policy is set exogenously at optimality, is also included in the graph for reference.

Modeling Result 2: Price and order comparisons amongst treatments are consistent with experimental observations.

Consistent with Experimental Result 2, the behavioral model implies that average prices in Ponly and PQ to be lower than the optimal price of 13. The "underpricing" is in line with the "midpoint bias" suggested by Su 2008 under the newsvendor setting. The basic intuition is that the probabilistic choice model, with limited support in the decision dimension, predicts a shift to the midpoint (a price of 11.50 in our case) of the range of the support. This reasoning applies to the pricing decisions in our setting as well. The shift is more prominent when gamma is at its low end as illustrated by Figure 4. Furthermore, the behavioral model predicts the average price to be higher in Ponly than in PQ. Recall that orders are fixed at 6 in Ponly. Yet in PQ, orders can be in the range between 0 and 15 with a mid-point of 7.5 (assuming the subjects will not order beyond the highest possible demand under the lowest price of 8). If we assume the demand can go up to 23, the highest possible demand given a zero price, the order will be even higher which further strengthens the intuition we are trying to illustrate. Applying the same idea of the mid-point bias, the model predicts the order quantity to be higher than 6 in PQ with a further reduced selling price. Although the pricing differences observed in the experiments are not significant across the treatments but they are along the direction suggested by the behavioral model.

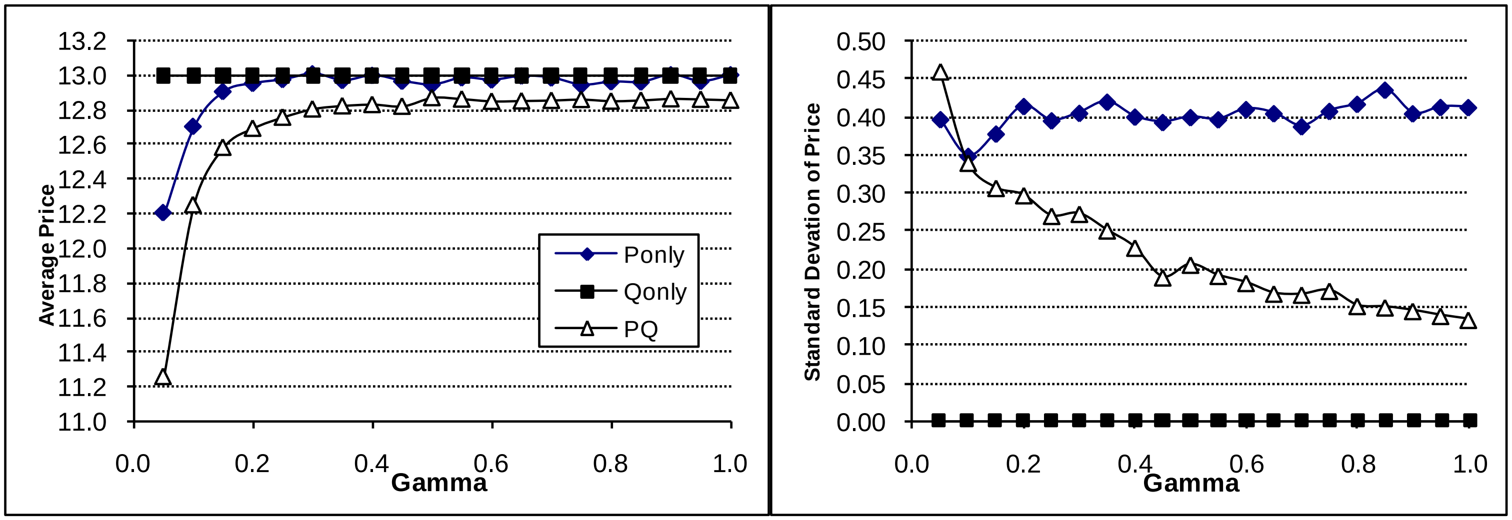

Figure 4: Pricing Decisions under Dynamic Setting as a Function of Gamma Predicted by the Behavioral Model.

As for the decision variance, the model predicts it to be lower in PQ than in Ponly. When gamma increases, the pricing policy under PQ converges to the list price with smaller variations, whereas the price in Ponly requires dynamic adjustments. When β is positive, standard deviations of the pricing decision enlarge in general under these two treatments, but the relative comparisons remain the same. In the newsvendor setting, the behavioral model suggests the pricing standard deviation to be larger in PQ for all gamma values, and the difference is more seemingly when gamma is less than 1. The standard theory implies the exact same pricing strategy to be used in Ponly and PQ under the newsvendor model. The larger decision variance predicted in PQ by the model therefore can be attributed to the fact of increased decision freedom.

Lastly, we examine the stocking decisions implied by the behavioral model. Figure 5 displays simulation results on averages and standard deviations of the order quantity for the three treatments under the dynamic setting. In Qonly, our model predicts the average order quantity to be close to 6, which indeed was observed in the experiment. The average order in PQ, however, is predicted to be above 6; and the treatment difference is more significant at lower levels of gamma. The intuitive explanation for the above results can again trace back to the mid-point bias. In Qonly, when the price is set to be 13, the resulting critical fractile is 50%. We do not see much variation from the mean demand of 6. In PQ, since the mid-point bias can cause the price to be below 13, the corresponding stocking decision is then higher than 6 due to the demand function. The larger decision variance in PQ can again be explained by the increased decision complexity. When β is positive, decision variances go up under both Qonly and PQ.

The ordering strategy converges to the base stock policy as decision makers become less myopic. Yet at any given value of beta, the standard deviation of orders in PQ is no less than that in Ponly.

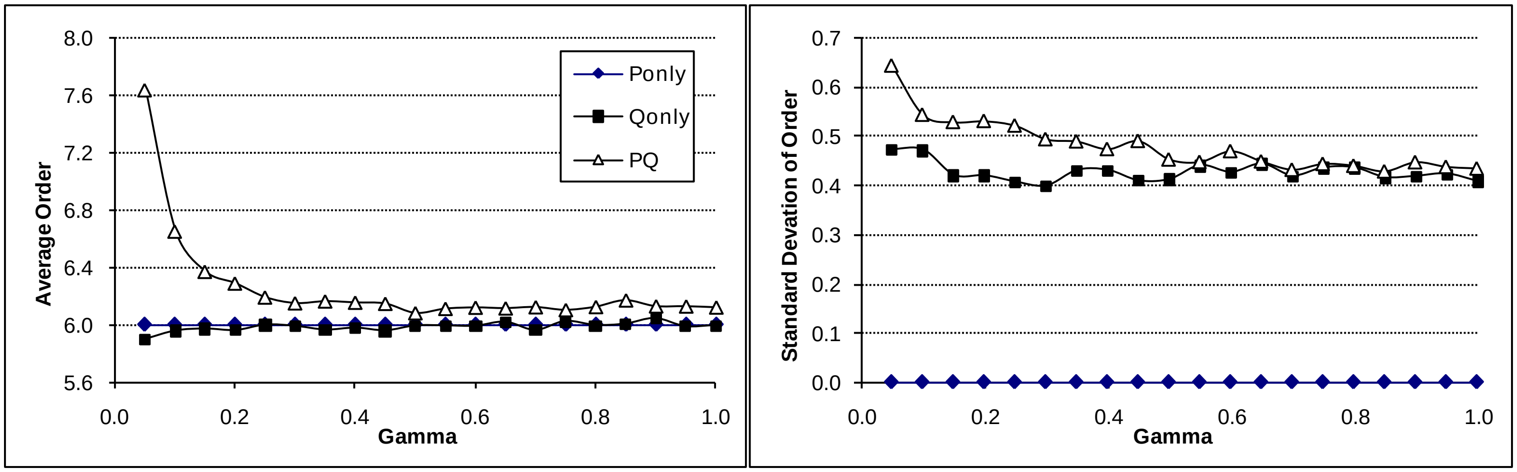

Figure 5: Ordering Decisions as a Function of Gamma by the Behavioral Model.

It is worth noting that results in this section are obtained under the assumption that γ remains the same across various treatments. From model estimations results reported in Section 5, we find that γ actually decreases as the "complexity" of the decision making increases (for example, from Qonly to PQ). The smaller γ estimated from PQ does not change the comparative results discussed above. In fact, it makes the gap even larger. The basic comparative statics of the model predictions still hold even if γ changes across treatments, as long as γ decreases with increased decision complexity. To conclude, directional predictions from the proposed behavioral model are consistent with lab observations in general. In the next section, we apply the model to estimate behavioral parameters of the subjects for their tendency to make random errors and their ability to look forward into the future.

5. MODEL ESTIMATION RESULTS

The method of maximum likelihood estimation is used to estimate γ and β, with individual differences being independent across subjects. Estimations of both γ and β are performed at an aggregate level, using observed decisions from all subjects under a treatment. For treatments under the newsvendor setting, since the initial stocking level is always 0, decisions are simulated to be independent across periods. For treatment under the dynamic setting, we assume that, conditional on the starting inventory in a period, decisions are independent. The total number of data points used = (the number of subjects) x (the number of periods) x (number of decisions in a period). Table 3 reports the corresponding behavioral estimates.

Table 3: Estimation Results from the Behavioral Model.

Treatment		Gamma	Beta	Log Likelihood
Newsvendor	Ponly	0.29	n/a	528.51
	Qonly	0.27	n/a	553.24
	PQ	0.13	n/a	1463.56
Dynamic	Ponly	0.11	0.14	1644.43
	Qonly	0.14	0.33	1938.82
	PQ	0.10	0.35	2528.67

In all treatments, log-likelihood ratio tests reject the null hypotheses that $\gamma=0$, or $\gamma=\infty$ (p-values < 0.001). We can conclude that subjects are neither perfectly rational ($\gamma=\infty$) nor did they make totally random decisions ($\gamma=0$). In other words, while subjects respond to profit incentives in making decisions, they are not capable of full optimization. The behavioral model proposed can capture the level of decision variations due to bounded rationalities. Furthermore, we note that estimates in PQ are significantly *lower* than those in Ponly and Qonly under each of the two inventory settings (all-values < 0.05). This again demonstrates that subjects have a higher propensity to make mistakes (lower gamma) when they can manage inventory through both pricing and ordering decisions (i.e., more complexity in decision making). Note that it is not possible to compare gamma estimates between the newsvendor and dynamic settings due to different profit scale under the respective models.

In treatments under the dynamic setting, log likelihood ratio tests reject the hypothesis that $\beta = 0$ with p-values < 0.001. This is strong evidence that subjects are not entirely myopic and do consider the future payoff while deciding for the current period. Moreover, we can reject the hypothesis that $\beta = 1$ in all three treatments with p-values < 0.001. This result cannot be explained by the discounting behavior. In our experiments, subjects are paid in cash right after they finish the game. Hence, we expect subjects to treat a dollar made in the last period the same as a dollar earned in the first period. An alternative explanation is that, due to limited cognitive ability, decision makers have to reduce the weight placed on periods further into the future while calculating the expected payoffs. This behavior is consistent with what was found in the beer game: subjects underweight the supply line, which are shipments to be received in the future (Sterman 1989).

Figure 6 provides an illustration for the goodness-of-fit of our behavioral model. The graph is generated using data and estimates from the dynamic Qonly treatment (it is picked for convenience. Similar graphs can be obtained from other treatments). In particular, we plot the observed frequencies of the order-up-to level"freq"), versus predictions from a myopic model with $\beta = 0$ ("myopic"), and the limited-forward looking model given the estimated β ("beta"). As we can see, the limited-forward looking model is an improvement over the myopic model, which is statistically confirmed by log likelihood ratio tests as well.

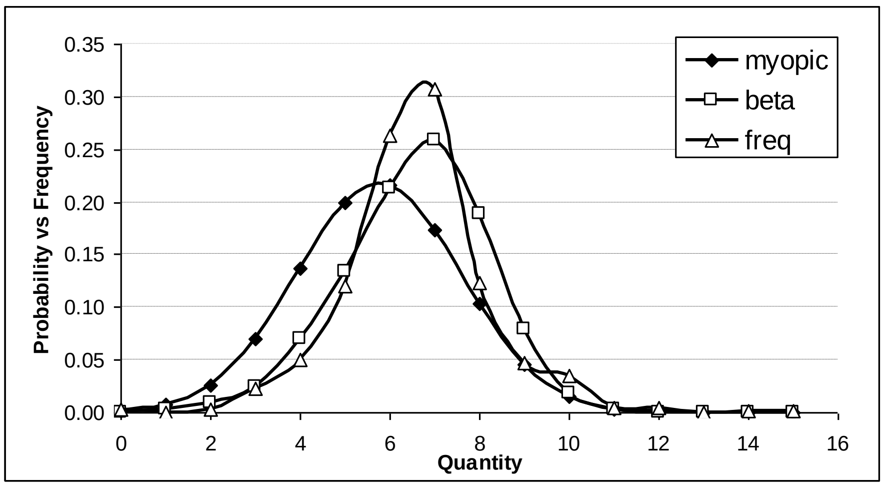

Figure 6: Goodness-of-fit of the Behavioral Model under the Dynamic Qonly Treatment.

Lastly, we check if there is any behavioral adjustment over time with the help from the behavioral model. Behavioral parameters are estimated for the first half and the last half of the game (each with 18 periods). Table 4 summarizes the results. Under the dynamic setting, we note that gamma increases while beta decreases from the first to the second half of the game significantly in *all* three treatments as shown by the likelihood ratio tests. The increase in gamma can be interpreted as subjects "learned" to reduce errors with more experience. On the other hand, the decrease in beta seems to suggest that subjects tend to be more myopic over time. We speculate that, since the calculation of future payoffs becomes more challenging due to the convolution of probabilities, subjects may choose to focus more on improving the evaluation of the current period, and thus pay less attention to the future periods. Under the newsvendor setting, we observe similar increasing trend in the gamma estimates. However, none of them is statistically significant. This observation is line with Bolton and Katok 2008, which showed learning from experience is difficult and slow under the newsvendor problem.

Table 4: Time Trend in Behavioral Estimates.

Treatment			Gamma	Beta	LRT p-value
Dynamic	Ponly	1-st Half	0.09	0.15	0.0004
		2-nd Half	0.14	0.13	
	Qonly	1-st Half	0.13	0.38	0.0150
		2-nd Half	0.16	0.28	
	PQ	1-st Half	0.07	0.43	0.0000
		2-nd Half	0.14	0.28	
Newsvendor	Ponly	1-st Half	0.26	n/a	0.3527
		2-nd Half	0.33	n/a	
	Qonly	1-st Half	0.25	n/a	0.8229
		2-nd Half	0.28	n/a	
	PQ	1-st Half	0.11	n/a	0.0810
		2-nd Half	0.14	n/a	

6. CONCLUSIONS AND FURTURE WORK

We present a behavioral study that systematically compares empirical decision making under two classical inventory management models (single-period versus multi-period) in the operations management literature. Our research differs from previous studies in that agents are provided with different options to control inventory, which affects the complexity of their decision making. Using controlled experiments, we identify biases of underpricing and understocking in human subjects. Their profit performance does not necessarily improve with more inventory control options. This contradicts optimization theories and we coin this phenomenon the “Optimizer’s Paradox”. A behavioral intuition behind this result is that more decision freedom or options may allow more room for human errors. We propose a behavioral model to capture behaviors observed in the laboratory. The model is developed based on two principles: agents are prone to make mistakes in utility evaluations, and they have limited capability to look forward under a dynamic setting. The behavioral model is successful in prediction decision biases and profit comparisons from the experiments. Estimations from the model further help understand adjustment in the empirical behaviors. The principles used for modeling are not limited to the specific experimental settings of this study.

This study offers several managerial implications. If more control options can lead to worse-off behavioral performance, restricting decision freedom may become a necessary strategy in practice. Feng and Gallego (1995) provide field evidence that decision restrictions are an important management strategy in the real world. It also questions the popular notion of “employee empowerment” in the organizational behavior literature. It is important to point out that decision restrictions do not guarantee better performance (as evidenced by the dynamic Ponly treatment). Therefore, it is necessary to tailor restrictions based upon specific scenarios. The model we propose can provide such a framework to understand how restrictions should be optimized behaviorally.

There are several limitations of the study, which future research can build upon. First, we only consider a dynamic setting without any delays. In practice, the replenishment decisions are often subject to lead times whereas pricing adjustment is relatively immediate. A natural research extension is to study the effect of inventory control options that differ in feedback delays. It is also interesting to apply our model to such dynamic scenarios to see if its predictions are robust. The current behavioral model is static and does not include behavioral preferences such as risk or loss aversion. Future research may consider the modeling framework of Experience Weighted Attraction (EWA) by Camerer and Ho (1998) to incorporate more behavioral effects into the discussion. Another direction for future work is to incorporate strategic interaction into the theory. For example, consider the inclusion of an upstream player who can modify the transfer price dynamically while determining his own production level. This is requires a series of new experiments using a modified version of the beer game. It also demands to extend the current choice model to an equilibrium one to capture behavioral interactions. We consider this research offers an important first step to bridge the gaps.

REFERENCES

Anderson, S.P., De Palma, A., and Thisse, J.F. (1992), *Discrete Choice Theory of Product Differentiation*, MIT Press, Cambridge, MA.

Bearden, J.N., Murphy, R. O. and Rapoport, A. (2008), Decision biases in revenue management: some behavioral evidence. *Manufacturing & Service Operations Management*, Vol. 10, Issue 4, pp. 625 - 636.

Bendoly, E., Donohue, K. and Schultz, K.L. (2006), Behavior in operations management: assessing recent findings and revisiting old assumptions. *Journal of Operations Management*, Vol. 24, Issue 6, pp. 747-752.

Bertsekas, D.P. (1976), *Dynamic programming and stochastic control*, New York: Academic Press, pp. 395.

Block, H.D. and Marschak J. (1960), Random ordering and stochastic theories of response. In I. Olkin, ed. *Contributions to Probability and Statistics*, Stanford University Press, Stanford, CA, pp. 97–132.

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

Busemeyer, J.R. (1999), International Encyclopedia of the Social and Behavioral Sciences: Methodology. *Mathematics and Computer Science*, Amsterdam: Pergamon.

Camerer, C. and Ho, T. (1999), Experience-weighted attraction learning in normal form games. *Econometrica*, Vol. 67, No. 4, pp. 827-874.

Chen, K.Y. and Wu, Y. (2020). Learning under the Economic Order Quantity problem. SJSU Working paper.

Diehl, E. and Sterman, J.D. (1995), Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Process*, Vol. 62, Issue 2, pp. 198-215.

Edwards, W. (1962), Dynamic decision theory and probabilistic information processing. *Human Factors.* Vol. 4, pp. 59-73.

Federgruen, A. (1993), Cebtralized planning models for multi-echelon inventory systems under uncertainty. *Logistics of Production and Inventory*, pp.133-173.

Federgruen, A. and Heching, A. (1999), Combined price and inventory control under uncertainty. *Operations Research*, Vol. 43, Issue 3, pp. 454-475.

Feng, Y. and Gallego, G. (1995), Optimal starting times for end-of-season sales and optimal stopping times for promotional fares. *Management Science*, Vol. 41, No. 8, pp. 1371-1391

Ho, T.H. and Weigelt, K. (1996), Task complexity, equilibrium selection, and learning: an experimental study. *Management Science*, Vol. 42, No. 5, pp. 659-679.

Johnson, E. and Payne, J. (1985), Effort and accuracy in choice. *Management Science*, Vol. 31, No. 4, pp. 395-414.

Katok, E., Donohue, K. and Leider, S. (2018), The Handbook of Behavioral Operations. Wiley.

Lim, N. and Ho, T. (2007), Designing price contracts for boundedly rational customers: does the number of blocks matter? *Marketing Science*, Vol.2 6, No. 3, pp. 312-326.

Luce, R. D. (1959), *Individual Choice Behavior: A Theoretical Analysis*, Wiley, New York.

McFadden, D. (1974), Conditional logit analysis of qualitative choice behavior. In P. Zarembka, ed. *Frontiers in Econometrics*, Academic Press, New York.

McFadden, D. (1976), Quantal choice analysis: a survey. *Annals of Economic and Social Measurement*, Vol. 5, No. 4, pp. 363-390.

Morgan, B.J.T. (1992), Analysis of Quantal Response Data. *Monographs on Statistics and Applied Probability*, Chapman & Hall: London.

Palfrey, T.R. and McKelvey, R.D (1995), Quantal response equilibria for normal form games. *Games and Economic Behavior*, Vol. 10, Issue 1, pp. 6-38.

Perera, H.N., Fahimnia, B. and Tokar, T. (2020), Inventory and ordering decisions: a systematic review on research driven through behavioral experiments, *International Journal of Operations & Production Management*, Vol. ahead-of-print.

Petruzzi, N.C. and Dada, M. (1999), Pricing and the newsvendor problem: a review with extensions. *Operations Research*, Vol. 47, Issue 2, pp. 183-194.

Porteus, E. L. (2002), *Foundations of Stochastic Inventory Theory*, Stanford University Press, Stanford, CA.

Rapoport, A. (1966), A study of human control in a stochastic multistage decision task. *Behavioral Science,* Vol. 11, pp. 18-32.

Rapoport, A. (1967), Variables affecting decisions in a multistage inventory task. *Behavioral Science*, Vol. 12, Issue 3, pp. 194-204.

Rapoport, A. (1975), Research paradigms for studying dynamic decision behavior. In D. Wendt & C. Vlek, eds. *Utility, Probability, and Human Decision Making,* Dordrecht-Holland: Reidel, pp. 347-369.

Rubinstein, A. (1998), *Modeling Bounded Rationality*, MIT Press, Cambridge, MA.

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

Simon, H. A. (1955), A behavioral model of rational choice. *Quarterly Journal of Economics,* Vol. 69, Issue 1, pp. 99–118.

Simon, H. A. (1982), *Models of Bounded Rationality*, MIT Press, Cambridge, MA.

Sterman, J. (1989), Modeling managerial behavior: misperceptions of feedback in a dynamic decision making experiment. *Management Science*, Vol. 35, No. 3, pp. 321-339.

Su, X. (2008), Bounded Rationality in Newsvendor Models. *Manufacturing & Service Operations Management*, Vol. 10, No. 4, pp. 566-589.

Thowsen, G.T. (1975), A dynamic, nonstationary inventory problem for a price/quantity setting firm. *Naval Research*, pp. 305-321.

Tversky, A. and Kahneman, D. (1971), Belief in the law of small numbers. *Psychological Bulletin*, Vol. 76, pp.105-110.

Veinott, A.F. Jr. (1965), Optional policy for a multi-product, dynamic, nonstationary inventory problem. *Management Science*, Vol. 12, Issue 3, pp. 206-222.

Wu,Y. and E. Katok. (2006), Learning, communication, and the bullwhip effect, *Journal of Operations Management*, Vol. 24, Issue 6, pp. 839-850