Cognition and Heterogeneity in Supply Chain Planning:
A Study of Inventory Decision Making

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Brent B. Moritz

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Arthur V. Hill, Adviser

July, 2010

Acknowledgements

A PhD is the product of hard work, good guidance and many excellent teachers who have brought me to this point. It is impossible to thank all of the individuals who have been so important, but I will offer a special thanks to several individuals:

I would like to thank my advisor, Arthur V. Hill. Art has been a friend and mentor to me throughout my journey to complete the PhD. I admire his insight and his character, which flows out of his strong Christian faith. He also has the unique ability to have excellent research and outstanding teaching while maintaining relevance to practice. As I start my career as a professor, I hope to measure up to his standard.

I would also like to thank my committee members: Karen Donohue, who encouraged me to look into the area of behavioral operations and who has added significantly to this dissertation; Enno Siemsen, a friend, colleague and co-author and with whom I look forward to additional projects; Kevin Linderman, who has been a voice of sound wisdom and who has served as chair of the final exam committee and Charles R. Fletcher, who has provided his expertise in cognitive sciences and is the outside committee representative.

Thanks as well to my past teachers and professors, from grade school through now. As part of the PhD experience, I specifically thank the following professors from outside our department: Paul Johnson, from whom I took my first seminar in cognitive science and was exposed to a new world of behavioral theories of decision making; Andy Van de Ven, as his comments on early versions of my research proposal proved especially helpful at a formative time and Mirko Kremer, a friend, future colleague and co-author at Penn State. I would like to thank the faculty and staff of the Operations and Management Science department. Thanks as well to the Carlson School Dean's Research Grant program, which provided funding for the experiments. Lastly, I also have worked with many great PhD students at Minnesota who have influenced this research, including my friend Aravind Chandrasekaran.

I particularly acknowledge the important contributions of those who have directly help shape this dissertation, and I look forward to several quality journal publications that will come out of this research. In particular, Art Hill and Karen Donohue have contributed to chapters two and three, while Enno Siemsen and Mirko Kremer have contributed to chapter four. They each have helped implement this research design and improve this manuscript in many areas, while any deficiencies are mine.

To my wife Benita: We met, were engaged and married while I worked on this PhD. While the research work has helped develop me as a scholar, you have improved who I am in far more profound and important ways. I thank you for your patience, character and love and look forward to growing closer to you as we begin new adventures together. I love you.

Lastly, I thank God for the abilities and opportunities he has given me. We are made in his image to in charge of his creation (Psalm 8). Perhaps, in some small way, this research may provide more insight as to how we go about that task. *Soli Deo Gloria*.

Dedication

To my late grandmother, Elsie Fredeen, who always valued education

and

to my wife, Benita.

Abstract

This dissertation investigates supply chain inventory decision making using behavioral experiments. Within inventory management, selecting the order quantity in advance of unknown future demand is a central question. If a decision maker knows the cost of having too much inventory, the cost of lost sales due to a shortage and an estimate of future demand, the newsvendor model provides a normative order quantity. Prior research shows that individuals regularly order too much in low margin settings and too little in high margin settings relative to the optimal quantity. Prior research has suggested several heuristics and preferences which may correlate with observed behavior, but little research explains reasons why individuals order as they do. This research is an investigation of that behavior in three areas. Chapter two finds that cognitive reflection explains a significant amount of the variance in performance for industry professionals in a high margin repeated newsvendor problem setting. Chapter three considers inventory ordering behavior across a range of margin settings and specifically investigates cognitive dissonance relative to customer service expectations. While cognitive reflection predicts performance in medium and high margin settings, minimizing cognitive dissonance appears to guide behavior in low margin settings. Chapter four looks at forecasting behavior, which is an antecedent to inventory ordering decisions. Specifically, this research finds that while system neglect appears to drive forecasting behavior, individuals with high cognitive reflection exhibit less system neglect and also have decision times closer to the predicted optimal time.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Research Background

There are two fundamental questions of inventory management: When to order, and how much to order. A great deal of research in operations management has be been devoted to the question of how much to order. More recently, the question has been considered from a behavioral point of view. Inventory managers routinely place orders for inventory in advance of knowing the actual customer demand. If they order too much inventory, it must be stored, discounted or salvaged. If they order too little, customers will be dissatisfied, and profit lost for both current and future customers. This problem is called the newsvendor problem, because left over newspapers are worth little, while running out of newspapers results in lost revenue and dissatisfied customers. The analytic solution to this problem has been available in literature for more than a century[1].

The newsvendor problem appears in many contexts, both large and small. Retail stores such as a small coffee shop have to order pastries from the bakery before knowing how many customers will make a purchase that day. Other decision makers must decide the inventory of fashion goods, repair parts, cars and computers before knowing the demand. Sometimes this mandates placing an order weeks or months ahead of holiday selling season, with only limited information about the actual demand. In public health, decision makers must order essential products such as flu vaccine in advance of knowing how many doses will be required. Revenue management, a closely related problem, involves decision makers who have a set quantity of goods (like the number of aircraft seats or hotel rooms) and must make pricing decisions so as to maximize the expected revenue. Fortunately, if a decision maker can make a forecast of demand and can estimate the prices and costs associated with a particular decision, the newsvendor model (cf. Porteus 2002, Silver *et al.* 1998) can provide a solution.

[1] A variant of the newsvendor problem appeared in the literature in 1888 (Edgeworth). While that problem was about the quantity of cash banks should keep on hand to meet customer demand versus the amount to lend at a profit, the mathematics of the solution are strikingly similar in a supply chain context.

However, even among experienced inventory managers and MBA students who have been taught how to solve the newsvendor problem, a number of behavioral anomalies have been observed (cf. Schweitzer and Cachon 2000). These anomalies include sub-optimal behaviors such as anchoring on the mean, chasing prior demand, a preference to minimize *ex post* inventory error, risk aversion and Prospect Theory preferences. The main finding is that individual decision makers order too much in a low-margin setting, and too little in a high margin setting. However, the causes of this persistent pattern of behavior are not well-known or well-understood.

The motivation for this dissertation comes out of a desire to understand the individual behavioral factors that cause at least a portion of the sub-optimal behavior. There were two motivating questions as to why individuals behave in a particular way, and how the behaviors differ by individual. Much of the prior research focused on analyzing experimental data relative to specific heuristics or preferences. While this is not a bad approach, the answers seemed incomplete. The heuristics and preferences that had been studied might say *what* was happening or if a particular heuristic appeared to be in use, but did not offer a compelling explanation as to *why* that might occur. For example, anchoring on the mean demand might explain a portion of observed behavior, but why do some people anchor on the mean and not another value? That more foundational question did not have much, if any, theory available. Second, much of the prior research focused on average behavior in a particular supply chain setting. However, average behavior of a large number of individuals is not the same as individual behavior. While there are limitations when studying one individual, the inventory decision is typically made by an individual. Said another way, do all people place the same weight on a particular heuristic or preference? This motivates the approach of theory driven experimental research into individual differences in supply chain inventory decisions.

About the same time I was beginning to wrestle with these questions, I was taking a PhD seminar course in Cognitive Science from Dr. Paul Johnson. Here I was exposed to some theory from judgment and decision making and psychology that were new to me, and appeared to be relatively unknown within the operations management community. However, some of the theory appeared to be very applicable to the newsvendor inventory task. This also corresponded with the growth of Behavioral Operations as a sub-field within the discipline of Operations Management. Hence, I had an interesting problem context, behavioral theory that seemed applicable and some runway for investigating these questions. What follows in my dissertation is an introduction to the theories that may explain at least a portion of observed behavior, behavioral experiments designed to test aspects of these theories and conclusions that may help explain a portion of the observed behavioral anomalies.

1.2 Dissertation Structure

The structure of this dissertation is in a three-paper model. Each chapter of the dissertation is designed to eventually be published as a stand-alone journal article. The reader of this dissertation should be able to understand each chapter without having read all of the prior parts of the dissertation in detail. Hence, there is some modest repetition of the problem context and some crossover within the literature review in the introduction to each chapter. However, each chapter addresses a different aspect of the problem and uses different theoretical approaches to investigate behavior. An overview of the entire dissertation structure is in Table 1.1.

Table 1.1: Dissertation Structure and Research Activities

Dissertation Structure	Title	Research Activities
Chapter 1	Introduction	Research Overview
Chapter 2	Cognition and Individual Differences in the Newsvendor Problem: Behavior Under Dual Process Theory	Experiment comparing inventory performance relative to cognitive reflection in high margin setting.
Chapter 3	Asymmetric Ordering Behavior in Newsvendor Inventory Decisions: Customer Service and Cognitive Dissonance	Experiment comparing inventory performance relative to cognitive dissonance in high, mid and low margin settings.
Chapter 4	System Neglect and Individual Differences in Forecasting Behavior	Experiment comparing forecasting performance in different signal to noise ratios, and comparing time to reach a decision and performance.
Chapter 5	Conclusion	Overall results, limitations and future research.

1.3 Research Design

1.3.1 Chapter 2 – Essay 1

Chapter 2 begins with the newsvendor problem setting and an introduction to dual process theory as a potential explanation of individual performance. Individuals regularly order too much in a low margin setting and too little in a high margin setting, and prior observations offered limited explanation as to why this might occur. Based on literature from judgment and decision making, the dual process theory construct of cognitive reflection (Frederick 2005) is compared to performance in a high margin newsvendor problem setting. Participants in essay 1 were 313 experienced supply chain practitioners, who participated in a repeated newsvendor problem scenario. Their performance was measured in terms of expected profit, order quantity and variance in order quantity. In addition, performance was compared relative to three well-known heuristics/preferences, including anchoring on the mean, anchoring on the prior period demand and a preference to reduce *ex post* inventory error. Lastly, performance was compared to other possible explanations such as years experience and college major. The results show that cognitive reflection predicts performance in this setting: Individuals with higher cognitive reflection performed better and had a lower tendency to use non-optimal heuristics.

1.3.2 Chapter 3 – Essay 2

Essay 2 looks at ordering behavior and cognitive dissonance, specifically comparing behavior in high, medium and low-margin settings. Several prior studies have observed that ordering behavior tends to be asymmetric: Individual ordering behavior tends to be about twice as far above mean demand in a high margin setting relative to the amount below the mean in a low margin setting. Additionally, while cognitive reflection predicted performance in a high margin setting, this was not the case in a low margin setting where the ordering behavior was closer together.

The main difference between the settings is related to the change in customer service expectations. In a high margin setting, the optimal order quantity corresponds with a plan to satisfy more than the expected number of customers. In a low margin setting, the optimal order quantity corresponds with a plan to disappoint customers. Based on theory from judgment and decision making, cognitive dissonance (Festinger 1957) might explain this anomaly. One marker for dissonance is trivialization (Simon et al. 1995), and this research is specifically designed to look for trivialization by assessing importance of several decision factors. In particular, in the low margin context, there is statistical evidence supporting trivialization specific to customer service, while there was no evidence of increasing importance in the medium and high margin contexts.

1.3.3 Chapter 4 – Essay 3

While chapters 2 and 3 focus on the newsvendor problem itself, chapter 4 looks at individual differences in forecasting behavior. The newsvendor problem has a forecasting task followed by an inventory management ordering task. The inventory management task is based on an estimate the financial factors such as selling price, cost, salvage value and goodwill associated with a lost sale. However, a key antecedent to the inventory order decision is the forecast of upcoming demand. Chapter 4 specifically investigates the time series forecasting task. In this experiment, forecasting performance is evaluated relative to the mean absolute error across several conditions which vary in change and noise. This is compared to the system neglect hypothesis, cognitive reflection and the time to reach a decision. Results show that individuals with high cognitive reflection perform better when forecasting, as they tend to have lower system neglect under conditions of high noise. These individuals also have a decision time that is neither too quick or too long, while individuals with lower cognitive reflection scores tend to be further from the predicted optimal time for each condition.

1.3.4 Chapter 5 – Summary, Conclusions and Future Work

Chapter 5 offers concluding comments on the contribution of this work for academics and practitioners. In addition, several opportunities for future work are outlined.

Chapter 2

Essay 1: Cognition and Individual Differences in the Newsvendor Problem: Behavior Under Dual Process Theory

Chapter Summary:

Previous research has shown that individuals systematically and persistently deviate from the profit maximizing quantity when solving a newsvendor problem. This research posits that Dual Process Theory provides an underlying cognitive explanation for why individuals deviate from optimality. More specifically, this research explores the relationship between individual performance and a Dual Process Theory construct called cognitive reflection, which can be measured by the Cognitive Reflection Test (CRT). We experimentally test the relationship between cognitive reflection and newsvendor decision-making using 313 experienced supply chain professionals. We find statistically significant results showing that cognitive reflection is related to performance as measured by expected profit, order quantity, and order quantity variance. Cognitive reflection is also related to anchoring heuristics and preference to reduce *ex post* inventory error. Other potential explanations of individual heterogeneity, including college major, years of experience, and managerial position, are also evaluated and found to be less informative than CRT scores. These results suggest that Dual Process Theory contributes to a theoretical understanding of supply chain decision-making. These results can be used to inform managerial decisions regarding employee selection, training, task design, and decision support systems.

2.1 Introduction

Our understanding of how people make inventory decisions has advanced significantly in the past decade. The seminal work of Schweitzer and Cachon (2000) revealed that when facing newsvendor decisions, the average response across individuals is to select an order quantity between the profit-maximizing optimal quantity and mean demand. Subsequent work has tested different explanations for this average behavior (Kremer *et al.* 2010; Su 2008) and examined how the behavior changes with experience and training (Bolton and Katok 2008; Lurie and Swaminathan 2009; Bolton *et al.* 2008). More recently, the research scope has expanded to examine the impact of environmental factors, such as whether or not demand is censored (Rudi and Drake 2009) and whether decisions are performed individually or in groups (Gavirneni and Xia 2009). Some studies observed wide variation in ordering between individuals but have not identified causal factors to explain this observed variation.

Much of the prior newsvendor research has implicitly assumed that decision makers are homogeneous by reporting average results. However, research in many disciplines has pointed to the importance of measuring attributes of individual respondents and using this information to explain some of the variance in the results. For example, research in cognitive psychology (Stanovich and West, 2000) and consumer behavior (Hutchinson *et al.* 2000) has identified problems with unobserved heterogeneity, also known as individual differences. Within operations management, Doerr *et al.* (2004) have highlighted worker heterogeneity and its impact on the variability of performance in assembly lines.

In supply chain inventory research, several recent studies point to the importance of individual differences (heterogeneity) in decision makers. Croson and Donohue (2006) call for theoretical research that incorporates the biases of individuals. Su (2008) developed a model that applies bounded rationality to newsvendor decisions, while calling for additional research and theory to look at cognitive limitations of individuals. In addition to a general observation about individual heterogeneity in judgment, Bolton and Katok (2008) specifically call for theory to explain individual variance in newsvendor problems. Despite this recognition, to our knowledge, no research has been dedicated to identifying and measuring individual attributes that might explain this variation between individuals in the newsvendor task.

The goal of this paper is to develop theory to explain some of the variation in performance between individuals and then test this theory in a newsvendor experiment. Drawing from cognitive science, we develop a theoretical model based on Dual Process Theory to identify *ex ante* individual characteristics that can be used to predict decision outcomes. More specifically, this theoretical model focuses on the Dual Process Theory construct called cognitive reflection, which can be measured with the Cognitive Reflection Test (CRT) (Frederick, 2005). We hypothesize that cognitive reflection predicts newsvendor performance and is related to several behavioral heuristics/preferences. We test our hypotheses in a controlled human experiment conducted with 313 supply chain managers and analysts from three firms. Because we are interested in cognitive differences across individuals who face inventory decisions, it was important to draw our respondents from an experienced industry population. Using this subject pool also allows us to test the impact of other individual characteristics that are often considered in hiring decisions, such as college major, years of professional service, and managerial position. We find that cognitive reflection is a stronger predictor of performance than any of these individual characteristics. More importantly, we find that cognitive reflection predicts individual performance in the newsvendor task, and is related to several explanations of behavior, including anchoring on the mean, anchoring on the prior period (demand chasing), and a preference to reduce *ex post* inventory error. These findings have potential implications for managerial decisions regarding employee selection, training, task design, and decision support systems.

The paper continues in section 2 with an introduction to the cognitive theory used in our study. Section 3 develops the research hypotheses and outlines the experiment, which places decision makers in a simulated newsvendor environment. Section 4 reports the experimental results, and section 5 summarizes the contributions of the research and suggests opportunities for further research.

2.2 Theory Development

In the classic newsvendor model (cf. Silver *et al.* 1998; Porteus 2002), a decision maker is faced with selecting an order quantity Q to satisfy stochastic demand D during a single sales period. The decision maker incurs a cost c for each unit purchased, earns price p for each unit sold, loses customer goodwill g for each unit of unsatisfied demand, and receives a unit salvage value s for each unit of unsold inventory. The underage and overage costs are then $c_u = p - c + g$ and $c_o = c - s$. For a given order quantity Q and demand realization D, the realized mismatch cost for the period is $G(D,Q) = c_o(Q-D)^+ + c_u(D-Q)^+$ and the realized profit is

$\Pi(D,Q)=(p-c)D-G(D,Q)$. The normative approach to solving the newsvendor problem is to assume the decision maker wishes to maximize expected profit

$$\Pi(Q)=\int_{D=0}^{\infty}\Pi(D,Q)f(D)dD, \qquad (1)$$

where $f(D)$ is the demand density function. The optimal order quantity with respect to this objective is

$$Q^*=F^{-1}\left(\frac{c_u}{c_u+c_o}\right), \qquad (2)$$

where $F^{-1}(.)$ is the inverse of the cumulative demand distribution function.

Individual decision makers frequently deviate from the optimal order quantity defined by equation (2). Although the newsvendor problem has a long history of published research (cf. Edgeworth, 1888), deviations from the optimal quantity are frequently observed in actual and experimental contexts. Previous research has suggested a number of possible heuristics and preferences that may explain this behavior. For example, it has been confirmed that average ordering behavior is consistent with heuristics such as anchoring on the mean or a preference to reduce *ex post* inventory error, while evidence for demand chasing is weak or non-significant (Schweitzer and Cachon 2000). In a more complex multi-echelon setting, Bloomfield *et al.* (2007) found that on average, order quantities selected by individuals were not sensitive to relative costs. Bloomfield *et al.* (2007) also found that some of the same behavioral factors found in the newsvendor problem appear in situations where inventory is replenished over time, and that inventory errors are exacerbated with transit lags. Bolton and Katok (2008) found that performance improves when individuals are prevented from drawing conclusions from inappropriately small samples across multiple decision periods. However, they observe anecdotally that the tendency for "too-quick" conclusions based on small samples seemed to vary widely between individuals. Kremer *et al.* (2010) reported that for a subset of the sample, demand chasing is significant at the individual level. While these and other papers describe potential heuristics and preferences that individuals might use to solve a newsvendor problem, little theory has emerged to explain the observed heterogeneity between individuals or to predict performance based on individual attributes.

2.2.1 Dual Process Theory

To better understand the decision-making process of individuals in the newsvendor problem, we draw from the fields of cognitive science and judgment and decision-making for theory on individual choice. While a number of possible heuristics have been proposed to explain some of the observed performance, our research posits that Dual Process Theory (Stanovich and West 2000) provides a cognitive foundation for understanding and explaining individual heterogeneity observed in the newsvendor task. The two cognitive approaches of intuition and analysis are called Associative and Rule-Based processes by Sloman (1996), Correspondence and Coherence Theory by Hammond (1996) and System 1 and System 2 by Stanovich and West (2000). While not all scholars agree on the terminology and on certain aspects of the two approaches, the key finding is that these are different cognitive processes in problem solving. We follow Stanovich and West's (2000) suggestion and use the generic distinction System 1 and System 2.

System 1 processes include more automatic cognitive tasks such as facial recognition, solving trivial math problems, and driving to a familiar place. Most problems amenable to System 1 processing do not require special cognitive abilities or deliberate mental effort. Such problems are solved rapidly when set within the context of the task itself. For example, driving a familiar route is usually a System 1 process. In contrast, it may take more and different cognitive effort (System 2) to give someone else step-by-step driving directions on that same route. System 2 processes typically require specific cognitive effort, targeted analysis, or the deliberate use of designed "scaffolding" (Clark, 1998) suitable for the problem.[2] Table 2.1 is our synthesis of much of this cognitive science literature and compares these two cognitive approaches for decision-making.

[2] For example, calculating the volume of a sphere requires the equation $V = \frac{4}{3} \pi r^3$, knowledge of basic algebra, a measure of the radius, and often a computational tool (e.g., calculator or computer). If any one of these elements is absent, this relatively simple problem cannot be solved with precision. In other contexts, if a decision maker appropriately applies the "right" equation or objective function as well as the inputs, evaluating a decision against a normative criterion is relatively straightforward.

Table 2.1: Comparison of Dual Process Decision Making

Dimension	System 1 Processes	System 2 Processes
Type of Decision-Making	Descriptive	Normative
Type of Knowledge Used	Tacit	Explicit
Type of Learning	Experiential Learning: Learn by Doing	Analytical/Book Learning: Learn by Thinking
Task Construal	Context-Specific	Decontextualized
Archetype	Handyman	Engineer
Decision Basis	Intuition/Common Sense	Theory/Rules and Laws
Traditional Business Domain	Organizational Behavior	Operations Research

When studying a decision *ex post*, it is not clear which cognitive approach an individual might have used to arrive at a decision. For example, a chess grandmaster using an intuitive System 1 approach can glance at a chess board and immediately know the right move, in part based on a tacit understanding of the board situation and a rapid comparison to similar games in memory. In contrast, an amateur (or a sophisticated computer) could make *exactly the same move* after considering many possible moves using a deliberate, analytical System 2 approach. The point here is that *ex post*, it is nearly impossible to know which cognitive approach is being applied by the decision maker merely by looking at the results of the decision. Similarly, to explain individual performance in the newsvendor task, an observer can see the decision Q_t while using theory from cognitive science to explain the behavior.

While the processing of System 1 decision inputs may be an interesting question for future newsvendor research, these factors are often highly dependent on the problem context. In some newsvendor decisions, intuitive, descriptive, and experiential decision inputs (System 1 processes) may have a role in selecting the profit-maximizing order quantity. In some industry contexts, only limited relevant historical demand data is available to characterize future demand. However, some individuals may have a particularly keen intuitive sense for predicting future demand for fashionable and trendsetting items. In addition, in many contexts, lost goodwill (g) is hard to estimate. Bolton *et al.* (2008) suggest that individuals with relevant experience might develop an intuitive feel for the solution to the newsvendor problem. For example, an individual with no formal training in finding the optimal solution could infer that it would be better to have more units leftover in a high-profit condition and few or no units leftover in a low-profit condition. Such a context-specific, tacit approach might not exactly reach Q^* but is likely to be fast, flexible, and generally robust in a wide range of circumstances. Lastly, important decisions are rarely made only by one person. Individuals and teams at several levels of the firm have to make decisions regarding inventory based on available objective data plus experience, intuition, and a "gut feel" (Gigerenzer 2007) of the parameters.

While these forms of tacit knowledge may have a role in parameter selection, our research emphasizes individual differences in selecting the particular order quantity. The System 2–based logic of the newsvendor model is well established in equations (1) and (2). The decision maker must accurately apply the cost parameters (or at least the critical ratio) and understand the demand distribution. With this information and equation (2), the decision maker can solve for Q^*. Rather than test context-specific parameter estimation, this research focuses on individual differences in decision-making behavior assuming the parameters are available and stable. To our knowledge, no previous research has tested Dual Process Theory as an underlying explanation of observed individual behavior in the newsvendor problem. To deliver on this objective, we identified a measure of Dual Process individual differences that could be used to compare performance in an experimental setting.

2.2.2 Measuring Individual Dual Process Differences Using the CRT

Because of the hidden, complex, and interrelated nature of neurological functions, no specific test can directly and independently measure the System 1 and System 2 functions of an individual. However, surrogate measures of performance have been suggested in behavioral science. Kahneman and Tversky (1982) suggest studying systematic errors in reasoning because those errors expose cognitive limitations or reveal the processes and procedures governing statistical or logical intuition. Such observations might highlight the System 1 features that create an error, or highlight reasons why an error was not overridden by a System 2 process (Kahneman and Frederick 2002). Frederick (2005) observed that individual decision-makers differ in terms of cognitive reflection, or the tendency to allow their System 2 thinking to moderate their System 1 response. The

Cognitive Reflection Test (CRT) was developed to measure this trait. Frederick also found that individuals with high cognitive reflection are more patient and less impulsive when presented with several judgmental tasks.

Frederick (2005) noted that while the CRT score is correlated with IQ, it also measures important aspects of individual heterogeneity related to risk and discounting. He concluded that the CRT is a measure of the ability and disposition to avoid reporting the first response that comes to mind. High CRT individuals are also more temporally patient, preferring later and larger rewards. These differences in risk preference and temporal discounting are not correlated with IQ. Such behavior appears to be particularly important in the newsvendor problem, where sub-optimal preferences such as minimizing *ex post* inventory or heuristics such as anchoring on the prior period or anchoring on the mean may be common un-moderated System 1 responses.

The CRT consists of three quantitative items (Table 2.2) where the obvious, impulsive answer is incorrect. According to Frederick (2005), the obvious answer that immediately springs to mind on question 1 for nearly all respondents is $0.10, but upon reflection, the correct answer is $0.05. The obvious (but incorrect) answers for questions 2 and 3 are 100 minutes and 24 days, respectively. However, the correct answers are only found if each respondent moderates their System 1 intuition with a System 2 approach to arrive at the correct answer. Frederick (2005) defined the CRT score for an individual as the sum of the number of correct responses on the instrument.

Table 2.2: The CRT Instrument

Q1. A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost?
____ cents

Q2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____ minutes

Q3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half the lake? ____ days

The objective nature of the CRT makes it more appealing than other scales suggested for measuring individual cognition. For example, the Need For Cognition scale (Cacioppo and Petty 1982) asks respondents to rate themselves on questions such as, "I would prefer complex to simple problems," answered on an agree-disagree Likert scale. Other scales including the Cognitive Style Index (Tetlock 2005) and the Need for Closure scale (Webster and Kruglanski 1994) are also similar, self-reported measures of preference. In contrast, as a measure of the cognitive tendency of the individual, the CRT is short, easy to administer, unambiguous, and does not rely on self-reported data.

2.3 Research Hypotheses

Our first set of hypotheses examines the relationship between the cognitive reflection of individuals (as measured by the CRT) and performance as measured by expected profit, order quantity, and order quantity variance. Our second set of hypotheses examines possible behavioral biases, preferences and alternative explanations of behavior. Beginning with the performance measures, we expect that individuals with higher cognitive reflection will achieve higher expected profit because they are less impulsive and more likely to allow their System 2 thinking to mediate their System 1 initial response. This leads to our first hypothesis:

> *H1: Individuals with higher cognitive reflection will have higher expected profit.*

Note that expected profit is an absolute measure of performance independent of any particular demand realization.

Hypothesis 2 compares the order quantity of individual decision makers to the normative (optimal) order quantity from equation (2). We expect less-reflective decision makers to select order quantities closer to the mean demand, while more-reflective individuals will select order quantities closer to the optimal order quantity Q^*.

> *H2: Individuals with higher cognitive reflection will have order quantities closer to the optimal quantity.*

Hypothesis 3 focuses on the reliability of an individual's order quantity decisions over multiple periods. When asked to solve the same newsvendor problem in sequential periods where the parameters do not change, the decision maker's order quantity should not change. We can operationalize reliability as the variance of the individual's order quantities over time. Individuals with lower cognitive reflection are more impulsive (i.e., less reflective). These individuals are less likely to carefully consider the objective function and less likely to understand the underlying system. This leads to larger and more frequent changes in the order quantity, which will increase order quantity variance across multiple periods. This leads to Hypothesis 3, which states:

> *H3: Individuals with higher cognitive reflection will have lower order quantity variance.*

Though related, Hypotheses H1, H2, and H3 measure three different aspects of performance. A decision maker who orders Q^* each period will have a higher expected profit (H1) than a second decision maker who orders exactly the mean demand each period. However, both individuals will have the same (zero) variance. H2 looks at the actual order quantity while H3 provides insight into the reliability of the individual decision maker and provides additional evidence of individual heterogeneity.

Prior behavioral research points to a number of possible explanations for why individuals deviate from Q^*. These include using heuristics, such as anchoring on mean demand and chasing prior demand realizations, and adopting a different objective function, such as minimizing *ex post* inventory error. Schweitzer and Cachon (2000) were the first to consider these decision models in a newsvendor context. Bostian *et al.* (2008) expanded these decision models to include a learning component, while Su (2008) proposed a complementary explanation based on a random error framework. Most recently, Kremer *et al.* (2010) tested the robustness of the models to changes in problem complexity and context and found further evidence for "guided" decision strategies. Our second set of hypotheses focus on the relationship between cognitive reflection and the use of these decision rules.

Our first hypothesis in this set focuses on the mean anchor heuristic, which suggests that decision makers use the mean demand as the starting point for order quantity decisions. Given that individuals with higher cognitive reflection tend to avoid going with the first response that comes to mind, we hypothesize that they will be less prone to anchor on the mean demand, which is a natural response. This leads to the following hypothesis:

> *H4a: Individuals with higher cognitive reflection will exhibit less anchoring on the mean.*

Similarly, the second anchoring heuristic includes a tendency to order inventory based on the prior period demand (i.e., to "chase demand"). Given that individuals with higher cognitive reflection are less likely to go with the first response that comes to mind, we hypothesize that they will be less likely to adjust their order based on the prior period demand (i.e., to chase demand) than individuals with lower cognitive reflection.

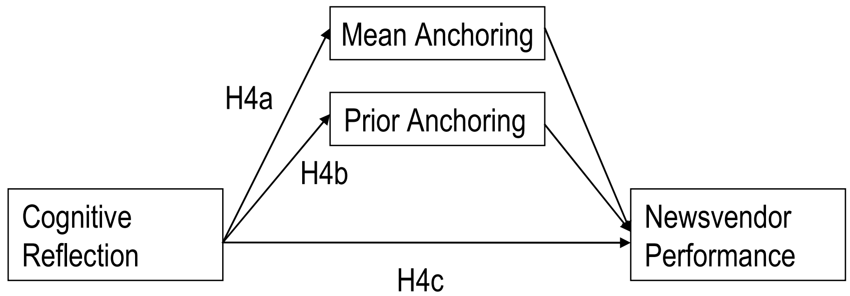
> *H4b: Individuals with higher cognitive reflection will exhibit less anchoring on the prior period demand.*

Hypotheses H4a and H4b suggest that individual cognitive reflection tendency predicts anchoring behavior, which can be observed based on the *ex post* order quantities. By definition, a decision maker who anchors on a value other than Q^* would have lower performance in terms of expected profits. In contrast, Hypothesis H4c suggests that cognitive reflection can predict performance *ex ante* even if the *ex post* anchoring assessments are included.

> *H4c: Even when individual anchoring tendencies are included, individuals with higher cognitive reflection will have higher expected profit.*

Figure 2.1 illustrates the proposed relationship between Hypotheses H4a, H4b, and H4c and performance.

Figure 2.1 Mediation Model of Anchoring and Cognitive Reflection



In particular, the model shows the potential direct impact of cognitive reflection on performance. If only H4a and H4b are supported, then cognitive reflection is only related to the use of anchoring heuristics. However, if H4c is supported, then cognitive reflection is also directly related to performance beyond *ex post* assessments of anchoring.

Finally, we consider the preference for minimizing *ex post* inventory error, which suggests that individuals derive personal utility from choosing the *ex post* realized demand, even though this order quantity is unlikely to maximize expected profit. Our hypothesis is that individuals with lower cognitive reflection will place more weight on this preference because System 1 decisions are likely to include other types of personal utility. In contrast, increased System 2 reflection leads to a profit-maximizing approach and away from other personal preference(s).

> *H5: Individuals with higher cognitive reflection will exhibit lower preference for minimizing ex post inventory error.*

Our main hypotheses all focus on testing the relationship between cognitive reflection and individual behavior either directly through performance measures (H1-H3) or indirectly by fitting outcomes to proposed decision rules (H4-H5). An obvious follow-up question is whether or not other individual characteristics, especially those that are often considered as part of a task selection process, are linked to cognitive reflection or help explain performance outside of this measure. Our last hypothesis focuses on three such characteristics (college major, years of business experience, and managerial position) and tests whether cognitive reflection is a better predictor of performance once these other factors are included. It is conceivable that certain college majors (e.g., engineering, supply chain, and finance) may be more inclined to use System 2 processing when compared with other majors. Similarly, Bolton *et al.* (2008) found that years of experience for practitioners significantly reduced performance in the newsvendor problem, while a higher managerial position significantly improved performance. Our finding should augment the prior results and provide additional insights into how these individual characteristics influence performance.

> *H6: When compared with college major, total business experience in years, and managerial position, cognitive reflection is a stronger predictor of expected profit performance in the newsvendor problem.*

2.4 Experimental Design

We used an online behavioral experiment to test these hypotheses. The online experiment was programmed in Java and web-enabled. The experimental design was similar to prior newsvendor experiments where respondents made repeated order quantity decisions over many periods. However, our experiment was different in four ways. First, the simulated demand was normally distributed rather than uniformly distributed as in most prior studies. This was important for external validity since demand is rarely uniformly distributed in practice, and our subject pool of professionals was more accustomed to this type of demand. Additionally, our choice followed Su's (2008) recommendation of studying decision biases under non-uniform demand. The second difference is that rather than running the experiment in a campus laboratory environment, the experiment was conducted online. This medium was necessary to accommodate our unique subject pool of supply chain professionals and to allow respondents to complete the exercise at their own workstations which may improve ecological validity (cf. Berkowitz and Donnerstein, 1982). Third, none of the industry respondents was directly compensated because of the difficulty of compensating hundreds of practitioners in three different firms and across multiple locations. In return for their participation, each firm obtained a confidential detailed benchmarking report comparing their aggregate performance to that of the other participating firms. Lastly, because we are particularly interested in individual differences, it was critical to focus on a single treatment and draw from a large subject pool to ensure that a sufficient number of subjects with a range of cognitive tendencies were available for analysis. After consultation with our industry partners, we chose to focus on a high margin condition since it was most similar to the actual margin conditions experienced by the three firms in our study. Specifically, we selected $p = 4$, $c = 2$, $g = 8$, $s = 0$, which implies $c_u = 10$ and $c_o = 2$. A single demand stream (based on a random draw from the normal distribution with $\mu_D = 100$ and $\sigma_D = 20$) was used to ensure that we could compare results across companies with sufficient statistical power, which results in $Q^* = 119.4$ units.

The experiment was performed with supply chain managers and analysts at three large, well-known, Fortune 500, supply-chain-intensive firms. All of the respondents were supply chain managers or analysts who typically made inventory or inventory-related supply chain planning decision such as allocating warehouse space or providing logistics/transportation support. The process was initiated in each firm by a senior executive who sent an e-mail to potential respondents on an existing distribution list inviting them to participate and promising confidentiality. Respondents were prevented from responding more than once. Once launched, the experiments took place over a two-week window at each firm. Before the experiment closed, at least one reminder notice was sent out. The respondents appeared to take the task seriously based on their participation time, their comments provided in an optional open-ended feedback section, and follow-up interviews. Response rates by firm were 53.8%, 50.4% and 25.1%, respectively, for an aggregate response rate of 36.6%. Given the anonymous nature of the instrument, we specifically looked for systematic bias in our samples. We provided company-specific detail about the number of responses by department to each firm. Based on this data, senior managers at each firm indicated that the responses were a representative sample of employees within their organization. These managers also indicated that the respondents were actively engaged in the exercise.

The cost parameters and demand distribution parameters were clearly stated to minimize estimation errors. This approach allowed our research to focus on individual differences in decision-making in a consistent problem context. In other words, to the extent possible, the differences observed can be attributed to individual differences in ability to estimate the critical ratio and appropriately apply it to the demand distribution, rather than differences in estimating the cost and demand parameters themselves.

The experiment proceeded as follows: First, each respondent provided basic demographic information about their work experience and education. Next, each respondent made order quantity decisions over twelve periods at a simulated store selling milk. At the end of each period, respondents received an updated report with the actual customer demand, the number of units discarded or short, and a calculation of their weekly financial performance. As with previous studies, the instructions noted that the demand distribution remained stationary (normally distributed with $\mu_D = 100$ and $\sigma_D = 20$).[3] After the newsvendor simulation was completed, each respondent was given the CRT.

2.5 Analysis and Results

2.5.1 Summary Observations

A total of 319 professionals participated in this study. Before doing any analysis, we removed six subjects, either because the time stamp showed they completed the entire exercise in an inappropriately short period of time or because they appeared to have outside knowledge of the demand stream.[4] For our final sample, n = 313. Following Frederick (2005), each individual was classified into one of four CRT groups based on their answers to the instrument. Table 2.3 shows the sample demographics and breakdown by CRT score and by firm.

[3] More detail regarding the parameters of the problem and one of the instruction screens is in the appendix.
[4] We recorded the total elapsed time from start to finish for the entire instrument. As an additional check, we regressed time against performance but found no statistically significant relationship.

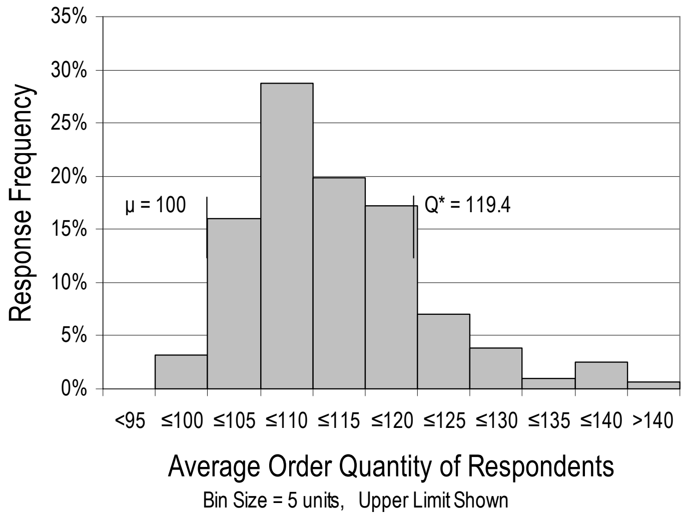
Table 2.3: Respondent Demographics and Split by CRT Score

Firm	Number of respondents	Average Years of Professional Experience	Standard Deviation of Years of Experience	Frequency by CRT Score			
				0	1	2	3
Firm A	67	16.1	10.6	23	18	19	7
Firm B	124	8.3	7.1	39	28	32	25
Firm C	122	18.3	11.5	18	21	43	40
Total	313	13.9	10.8	80	67	94	72

The respondents were heterogeneous in their responses. As in previous studies, the average order quantity across respondents deviated toward the mean, $\bar{Q}$ = 112.4 units versus Q^* = 119.4 units. The average expected profit across all respondents was \$916.80 compared with an optimal expected profit of $\Pi(Q^*)$ = \$940.00/week.[5] Figure 2.2 shows the distribution of average order quantities by respondent. The range of average orders varied from 96 to 160, the middle 50% range was 106 to 117, and the median was 110. While 83% of respondents exhibited an average order between Q^* and μ_D, the magnitude of this tendency to bias toward the mean varied across individuals. In the next subsections, we test whether this variation can be explained by an individual's level of cognitive reflection.

[5] Appendix 2 presents a graph of the expected profit function.

Figure 2.2 Average Order Quantity by Respondent



2.5.2 Decision Performance by CRT Score

We test the impact of cognitive reflection on our three performance measures using a GLM ANOVA procedure. Figure 2.3 summarizes the results. Consistent with their higher proportion of CRT = 2 and CRT = 3 respondents, Firm C performed slightly better than the other two companies and throughout the subsequent analysis we control for firm effects where appropriate. We find that performance differs significantly by CRT group for each performance measure (expected profit: $F = 14.275$ (3, 301), $p \leq 0.001$, order quantity: $F = 8.537$ (3, 301), $p \leq 0.001$, and order quantity variance: ($F = 4.156$ (3,301) $p = 0.007$.) Figure 3 also shows that expected profit, order quantity, and order quantity variance all move closer to their optimal values as the level of cognitive reflection increases. To test the significance of this trend, we used a Tukey HSD procedure to perform all pairwise comparisons by CRT group for each performance measure. The results show that the difference between the low CRT group (CRT = 0) and those exhibiting higher reflective tendencies (CRT = 2 and CRT = 3) is particularly strong ($p \leq 0.001$ for all performance measures). We report all pairwise comparisons in this paper for completeness, though Frederick (2005) largely focuses on the difference between CRT = 0 and CRT = 3. The results show that performance generally increases with CRT score, though there is no statistical difference between the highest-performing groups

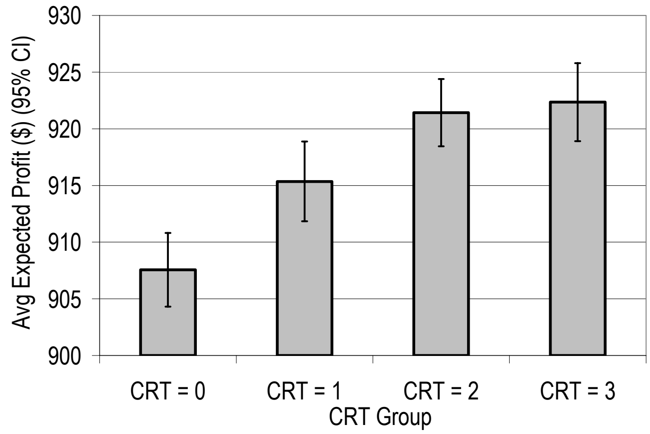
Figure 2.3 Comparison of Expected Profit, Order Quantity and Order Quantity Variance by CRT Group

2.3a Expected Profit[a]

CRT Group	*Average Exp. Profit*	*SE*
0	$907.64	1.701
1	$915.25	1.795
2	$921.42	1.577
3	$921.69	2.196

CRT Group Significance
$F = 14.275\ (3, 301)$, $p \leq 0.001$

Pairwise Multiple Comparisons (Tukey HSD):
$\mu_0 < \mu_1, \mu_2, \mu_3$ $p < 0.006$
$\mu_1 < \mu_2, \mu_3$ $p < 0.030$
$\mu_2 < \mu_3$ p = n/s

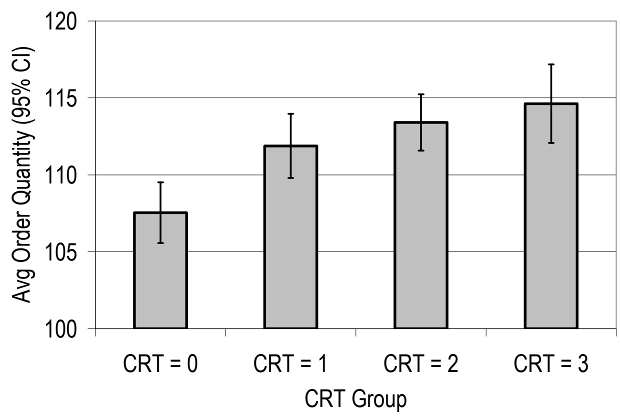


2.3b Order Quantity[a]

CRT Group	*Average Order Qty.*	*SE*
0	107.543	1.004
1	111.878	1.060
2	113.406	0.931
3	114.625	1.296

CRT Group Significance
$F = 8.537\ (3, 301)$, $p \leq 0.001$

Pairwise Multiple Comparisons (Tukey HSD):
$\mu_0 < \mu_1$ $p = 0.011$
$\mu_0 < \mu_2, \mu_3$ $p \leq 0.001$
$\mu_1 < \mu_2$ p = n/s
$\mu_1 < \mu_3$ $p = 0.070$
$\mu_2 < \mu_3$ p = n/s

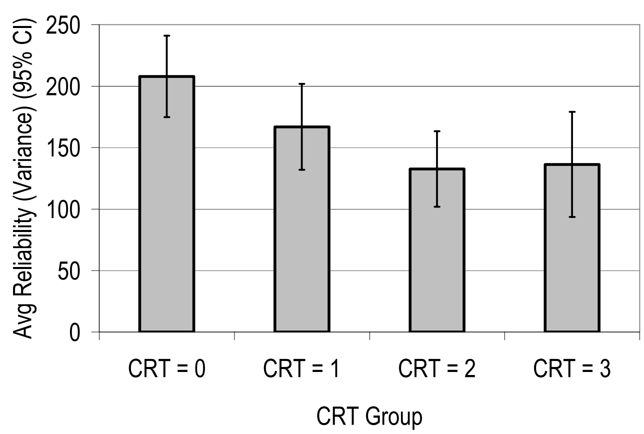


2.3c Order Quantity Variance[a]

CRT Group	*Average Variance*	*SE*
0	207.9	16.82
1	166.9	17.75
2	132.7	15.59
3	136.3	21.72

CRT Group Significance
$F = 4.156\ (3, 301)$, $p \leq 0.007$

Pairwise Multiple Comparisons (Tukey HSD):
$\mu_0 > \mu_1$ $p = 0.091$
$\mu_0 > \mu_2, \mu_3$ $p \leq 0.001$
$\mu_1 > \mu_2, \mu_3$ p = n/s
$\mu_2 > \mu_3$ p = n/s



[a] Controlling for firm effect
[b] We report $(1- \alpha_i)$ to make interpretation easier and to be directionally consistent with β_i and δ_i.

(CRT = 2 versus CRT = 3). Further analysis (not shown) indicated that these performance results were similar within each of the three firms. In summary, individuals from the lowest CRT groups have a lower expected profit, order further from the optimal quantity, and exhibit higher order quantity variance than those in the highest CRT groups. H1, H2, and H3 are supported.

2.5.3 Differences in Use of Anchoring Heuristics by CRT Group

To test the presence of anchoring behavior, we fit the data against a series of candidate decision models. First, we follow Bostian *et al.* (2008) by operationalizing the mean anchoring heuristic as a linear partial adjustment model:

$$Q_{it} = \mu_D + \alpha_i (Q^* - \mu_D) + \varepsilon_{it} \tag{3}$$

We use OLS regression to calculate the average score α_i for each respondent i that best reflects the deviation in their order quantity in a given period (Q_{it}) from the average demand (μ_D). As shown in Figure 2.4, we find that low CRT groups anchored on the mean more often than high CRT groups (F = 8.537 (3,301) $p \leq 0.001$). H4a is supported.

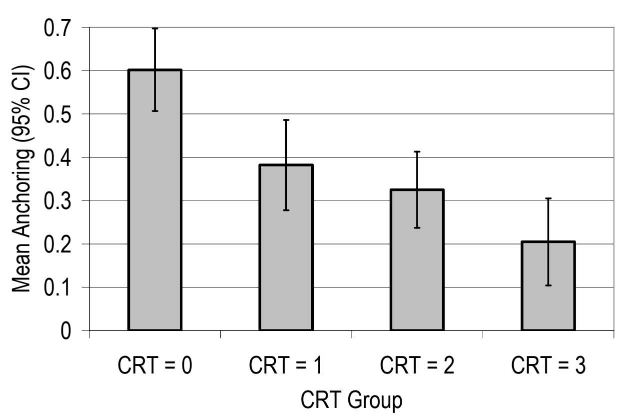
Figure 2.4 Comparison of Mean Anchoring Tendency by CRT Group

Mean Anchoring Tendency $(1-\alpha)$[a,b]

CRT Group	***Average Anchoring* $(1-\alpha)$**	***SE***
0	0.602	0.048
1	0.382	0.053
2	0.325	0.045
3	0.205	0.051

CRT Group Significance
F = 8.537 (3, 301), $p \leq 0.001$

Pairwise Multiple Comparisons (Tukey HSD):
$\mu_0 < \mu_1, \mu_2, \mu_3$ $p < 0.011$
$\mu_1 < \mu_2$ p = n/s
$\mu_1 < \mu_3$ $p = 0.070$
$\mu_2 < \mu_3$ p = n/s



[a] Controlling for firm effect.
[b] We report (1- α_i) to make interpretation easier and to be directionally consistent with β_i and δ_i.

We next operationalize the second anchoring heuristic (demand chasing / Prior Q Anchoring) as a linear partial adjustment model (again following Bostian *et al.* 2008):

$$Q_{it} = Q_{it-1} + \beta_i (D_{t-1} - Q_{it-1}) + \varepsilon_{it} \tag{4}$$

where subject i considers the previous order (Q_{it-1}) and adjusts the order quantity based on the realized demand in the previous period (D_{t-1}). Again using OLS regression, we calculate β_i for each respondent i that measures the propensity to chase demand. As shown in Figure 2.5, the tendency to chase differed by group (F = 4.206 (3,301) $p \leq$ 0.006), with low CRT groups exhibiting more chasing behavior. H4b is supported.

Figure 2.5 Comparison of Anchoring and Adjustment Tendency (Chasing) by CRT Group

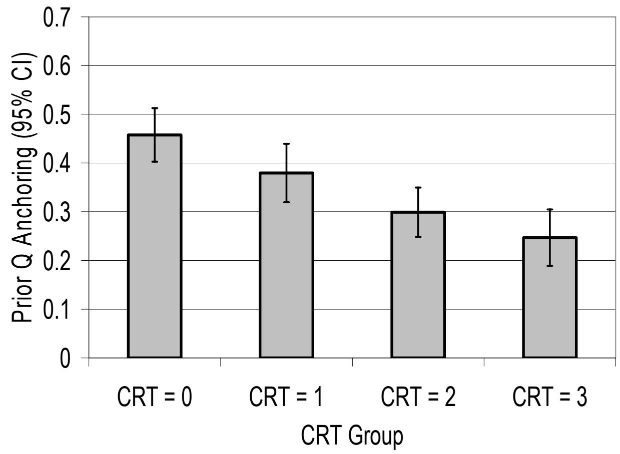
Prior Q Anchoring and Adjustment Tendency (Chasing) (β)[a]

CRT Group	*Average Chasing* (**β**)	*SE*
0	0.458	0.028
1	0.379	0.030
2	0.299	0.026
3	0.247	0.029

CRT Group Significance
F = 4.206 (3, 301), $p \leq 0.006$

Pairwise Multiple Comparisons (Tukey HSD):

$\mu_0 < \mu_1$	p = n/s
$\mu_0 < \mu_2, \mu_3$	$p \leq 0.001$
$\mu_1 < \mu_2$	p = n/s
$\mu_1 < \mu_3$	p = 0.005
$\mu_2 < \mu_3$	p = n/s

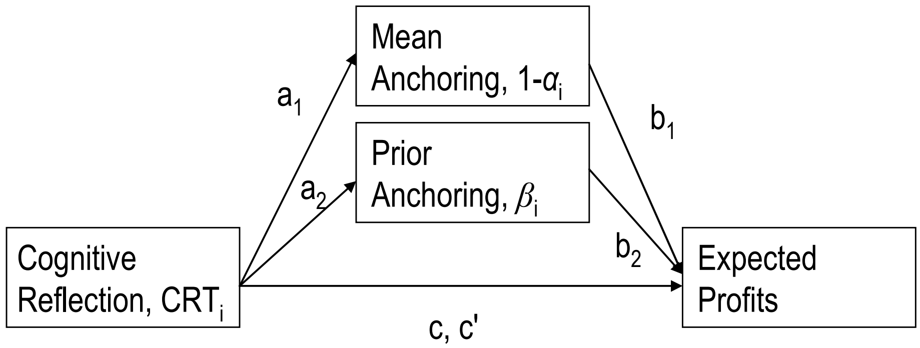


[a] Controlling for firm effect.

Specific to chasing, this result is in contrast to previous research that shows only limited support for the chasing heuristic. Many respondents in our experiment exhibited chasing behavior and moved their responses in the direction of the most recent demand, a rational strategy if demand is changing. However, the key finding is that the *amount* of chasing varied based on cognitive reflection. The low CRT group exhibited nearly twice the amount of chasing as the high CRT group. This offers a potential explanation of the weak or non-significant chasing results observed in prior studies such as in Schweitzer and Cachon (2000).

Next, we look at H4c, where the anchoring heuristics are included in the mediation model. We apply the Sobel test of mediation (Baron and Kenney 1986) as well as notation for mediation analysis to test for the impact of cognitive reflection on performance with the anchoring heuristics in the model, as shown in Figure 2.6.

Figure 2.6 Mediation Model and Results



Path	Coefficient	Standard Error	*t*	*p*-value
a_1	-0.072	0.014	-5.646	0.000
a_2	-0.126	0.022	-5.695	0.000
b_1	-28.667	2.543	-11.272	0.000
b_2	-14.204	1.462	-9.715	0.000
c	5.533	0.753	7.359	0.000
c'	1.682	0.624	2.694	0.008
$R^2 = 0.513$, $R_{adj}^2 = 0.508$, $F = 108.52$ (3,309) $p \leq 0.001$				

Not surprisingly, the mediation assessment shows a relatively strong impact of anchoring on expected profits (paths b_1 and b_2) in a partial mediation model. However, the significance of the c' path is of particular interest. The c path is the direct impact of cognitive reflection on results without the anchoring heuristics, and is the baseline comparison of the direct effect. The fact that the c' path is significant (t = 2.694, p = 0.008) shows that CRT measures both the *ex ante* tendency of an individual to anchor (paths a_1 and a_2) and is related to *ex post* performance in its own right. The significance of the c' path points to inclusion of cognitive reflection in the model even when an *ex post* assessment of anchoring might be used to explain performance. H4c is supported.

2.5.4 Difference in Utility Function by CRT Group

As mentioned in section 3, previous research has shown that average behavior is consistent with the preference to minimize regret associated with *ex post* inventory error. Following Bostian *et al.* (2008), we operationalize the expected utility for respondent *i* as the expected profit less delta times the expected regret for not selecting the realized demand.

$$U_i(Q_{it}) = \Pi(Q_{it}) - \delta_i R(Q_{it}) + \varepsilon_{it} \qquad (5)$$

where expected regret in period *t* is defined as $R(Q_t, D_t) = |Q_t - D_t|^2$. The δ_i parameter is the implied degree of induced regret for each individual found through a double exponential maximum likelihood logit model. As shown in Figure 2.7, analyzing *ex post* regret δ_i, the difference by CRT group was again significant (F = 5.569 (3,301) $p \leq$ 0.007). While it is not possible to distinguish if a respondent acts on a preference to minimize *ex post* regret or anchors on the mean (which also minimizes expected regret for symmetric demand), the results hold by CRT group, and H5 is supported.

Figure 2.7 Comparison of Minimizing *Ex Post* Inventory Error by CRT Group

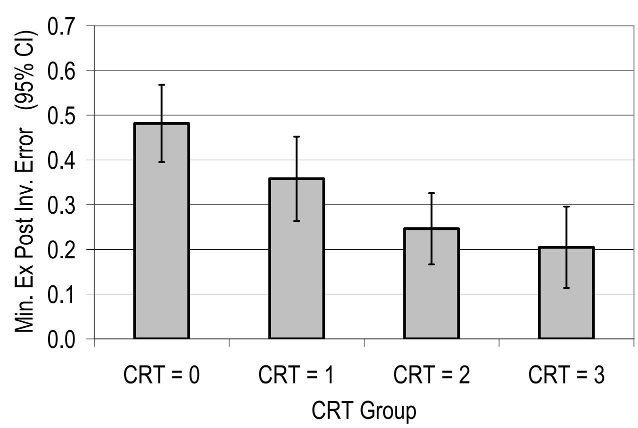
Minimizing *Ex Post* Inventory Error Preference (δ)[a]

CRT Group	*Average Error Preference (δ_i)*	*SE*
0	0.482	0.044
1	0.358	0.048
2	0.246	0.040
3	0.205	0.046

CRT Group Significance
F = 5.569 (3, 301), $p \leq 0.007$

Pairwise Multiple Comparisons (Tukey HSD):

$\mu_0 < \mu_1$	p = n/s
$\mu_0 < \mu_2,\mu_3$	$p \leq 0.001$
$\mu_1 < \mu_2$	p = n/s
$\mu_1 < \mu_3$	p = 0.100
$\mu_2 < \mu_3$	p = n/s



[a] Controlling for firm effect.

2.5.5 Differences in Performance: Alternative Explanations

We investigated alternative explanations of performance including college major[6], years of experience, and managerial position. For this portion of the analysis, we apply the Bonferroni multiple comparisons adjustment because we are testing a smaller number of comparisons, limited by the data available for each combination of interest. In looking at

[6] The discussion of performance by college major was not the primary focus of our study. This discussion is limited by sample size, particularly for certain majors that are infrequent in our sample of 313 supply chain professionals. Additionally, for this portion of the analysis, we do not control for firm effect, as many of the engineering/physical science majors were found in one firm.

Table 2.4: Comparison of Expected Profit by Major and CRT Group

College Major	Complete Sample n	$E[\pi]$[d]	*SE*	CRT Group and College Major[a] $E[\pi]$ *SE* 0	1	2	3	F (df, df_{error})	Pairwise Multiple Comparisons[e]
Liberal Arts	15	\$913.61	*5.287*	\$890.52 *7.722*	-	-	-		
Business: Accounting/Finance	20	\$921.11	*3.569*	\$913.36 *5.474*	-	\$925.85 *5.474*	-		
Business: Marketing/ Management	142	\$913.78	*1.295*	\$909.56 *2.265*	\$911.38 *2.517*	\$920.94 *2.213*	\$915.23 *3.129*	4.936 (3, 138) p=0.003	$\mu_0<\mu_1$ p=*n/s*, $\mu_0<\mu_2$ p=0.003, $\mu_0<\mu_3$ p=*n/s* $\mu_1<\mu_2$ p=0.026, $\mu_1<\mu_3$ p=*n/s*, $\mu_2<\mu_3$ p=*n/s*
Business: Supply Chain / Operations	57	\$919.84	*2.065*	\$906.33 *4.397*	\$917.71 *4.395*	\$923.00 3.351	\$926.87 *3.652*	5.406 (3, 53) p=0.003	$\mu_0<\mu_1$ p=*n/s*, $\mu_0<\mu_2,\mu_3$ p=0.010 $\mu_1<\mu_2,\mu_3$ p=*n/s*, $\mu_2<\mu_3$ p=*n/s*
Education / Social Sciences	12	\$919.63	*4.500*	-	-	-	-		
Engineering / Physical Sciences	36	\$925.13	*2.674*	-	\$915.52 *5.548*	\$923.10 *3.741*	\$929.05 *2.924*	2.484 (2, 34) p=0.100	$\mu_1<\mu_2$ p=*n/s*, $\mu1<\mu3$ p=0.100 $\mu_2<\mu_3$ p=*n/s*
Other	24	n/a	n/a	n/a	n/a	n/a	n/a		

F (df, df_{error})	3.605 (6, 299), p=0.002

F (df, df_{error})[b]	1.426 (6, 278), p=n/s	F (df, df_{error})[c]	9.31 (3, 278), $p\leq$0.001

[a] Only cells with $n \geq 5$ are reported.
[b] F Values for college major with CRT Included
[c] F Values for CRT with college major Included
[d] F Maximum $E[\pi]$ = \$940.00, $E[\pi]_{Q=mean}$=\$904.25
[e] Pairwise multiple comparison of group means (Bonferroni)

Table 2.5: Comparison of Expected Profit by Years of Experience and CRT Group

Years Exp.	Complete Sample			CRT Group and Years of Experience[a] $E[\pi]$ *SE*									
	n	$E[\pi]$[d]	*SE*	0		1		2		3		F (df, df_{error})	Pairwise Multiple Comparisons[e]
≤ 1 Year	21	\$918.27	*4.421*	-	-	\$913.49	*5.335*	\$918.61	*4.939*	\$928.78	*4.939*	1.693 (2,16) p=n/s	
1-5 Years	65	\$916.45	*1.858*	\$910.71	*4.118*	\$912.93	*3.684*	\$924.54	*3.113*	\$917.62	*3.460*	3.120 (3, 61) p=0.031	$\mu_0<\mu_1,\mu_3$ p=n/s, $\mu_0<\mu_2$ $p\leq0.046$ $\mu_1<\mu_2$ p=0.087, $\mu_1<\mu_3$ p=n/s, $\mu_2<\mu_3$ p=n/s
5 - 15 Years	107	\$918.55	*1.468*	\$908.18	*2.459*	\$919.34	*3.077*	\$922.39	*2.694*	\$924.29	*3.478*	7.263 (3,103) $p\leq0.001$	$\mu_0<\mu_1$ p=0.028, $\mu_0<\mu_2,\mu_3$ $p\leq0.001$ $\mu_1<\mu_2,\mu_3$ p=n/s, $\mu_2<\mu_3$ p=n/s
>15 Years	120	\$915.10	*1.357*	\$904.19	*2.700*	\$912.64	*3.134*	\$919.81	*2.505*	\$923.79	*2.744*	10.156 (3,116) $p\leq0.001$	$\mu_0<\mu_1$ p=n/s, $\mu_0<\mu_2,\mu_3$ $p\leq0.001$ $\mu_1<\mu_2$ p=n/s, $\mu_1<\mu_3$ p=0.051, $\mu_2<\mu_3$ p=n/s

F (df, df_{error})	0.710(3, 309), p=n/s

F(df, df_{error})[b]	1.038 (3, 297), p=n/s	F (df, df_{error})[c]	6.080 (3, 297), $p\leq0.001$

[a] Only cells with $n \geq 5$ are reported.

[b] F Values for years of experience with CRT included

[c] F Values for CRT with years of experience included

[d] Maximum $E[\pi]$ = \$940.00, $E[\pi]_{Q=mean}$ = \$904.25

[e] Pairwise multiple comparison of group means (Bonferroni)

Table 2.6: Comparison of Expected Profit by Managerial Position and CRT Group

Managerial Position	Complete Sample			CRT Group and Managerial Position E[π] *SE*									
	n	E[π][c]	*SE*	0		1		2		3		F (df, df_{error})	**Pairwise Multiple Comparisons**[d]
Non-Managers	245	\$915.90	*0.949*	\$907.66	*1.682*	\$913.56	*1.955*	\$922.26	*1.717*	\$921.01	*2.074*	15.076 (3,241) $p \leq 0.001$	$\mu_0<\mu_1$ $p=n/s$, $\mu_0<\mu_2,\mu_3$ $p \leq 0.001$ $\mu_1<\mu_2$ $p=0.006$, $\mu_1<\mu_3$ $p=0.057$, $\mu_2<\mu_3$ $p=n/s$
Managers	68	\$917.19	*2.021*	\$900.89	*5.765*	\$921.40	*4.231*	\$919.66	*3.114*	\$926.81	*3.114*	5.264 (3, 64) $p=0.003$	$\mu_0<\mu_1,\mu_2,\mu_3$ $p<0.028$ $\mu_1<\mu_2,\mu_3$ $p=n/s$, $\mu_2<\mu_3$ $p=n/s$

F (df, df_{error})	4.949 (1, 311) $p=0.026$

F (df,df_{error})[a]	0.233 (1, 305) p=n/s	F (df, df_{error})[b]	11.726 (3, 305) $p \leq 0.001$

[a] F Values for managerial position with CRT included

[b] F Values for CRT with managerial position included

[c] Maximum E[π] = \$940.00, E[π]$_{Q=mean}$=\$904.25

[d] Pairwise multiple comparison of group means (Bonferroni)

expected profits between groups (Table 2.4), major is a significant predictor of performance (F = 3.605 (6,299) p = 0.002), with Business (Supply Chain/Operations) and Engineering/Physical Science majors performing especially well. However, if cognitive reflection is included in the analysis, the effect of major is no longer significant (F = 1.426 (6,278) p = n/s) while cognitive reflection is significant (F = 9.310 (3,278) $p \leq$ 0.001). The implications are that if one knows nothing about an individual except their college major, it is possible to estimate expected profit performance. However, cognitive reflection is a better predictor than major. In addition, mean expected profits increase with CRT score within each of the majors (e.g., Marketing/Management majors (F = 4.939 (3,138) p=0.003), Supply Chain/Operations majors (F = 5.406 (3,53) p=0.003), and Engineering/Physical Science majors (F = 2.848 (2,34) p=0.100)). The highest performing group of individuals were Engineering/Physical Science majors with CRT = 3.

Turning to the self-reported years of business experience (Table 2.5), previous research has found that more years of experience decreases expected profit in repeated newsvendor problems (Bolton *et al.* 2008). Our research found no significant relationship between years of experience and average expected profit. Years of experience was not a significant predictor of average expected profit for the categories shown (Table 7), using either a GLM ANOVA procedure (F = 0.710 (3, 309) p = n/s) or using a regression of actual (non-categorized) years of experience (available on request). Additionally, as before, CRT is a robust predictor of performance across the range of experience (F = 6.080 (3, 297) $p \leq$ 0.001).

Lastly, we compared the performance of managers versus individual contributors (Table 2.6). We identified those individuals who are likely to be managers (n = 68) based on their title as managers, directors or vice-presidents, or as non-managers (n = 245) typically identified as analysts. Taken separately, managers had higher expected profits than non-managers (F = 4.949, (1, 311) p = 0.026). However, when cognitive reflection is included, the expected profit of managers and non-managers were not statistically different (F = 0.233 (1, 305) p = n/s), although CRT is again significant (F = 11.726 (3, 305) $p \leq$ 0.001). We conclude that managers perform better than non-managers in our study. However, cognitive reflection is a better predictor of performance than college major, years of experience, or managerial position. Therefore, we conclude that H5 is supported.

2.6 Conclusions

Taken as a whole, our results confirm that individual cognitive heterogeneity based in Dual Process Theory predicts performance in the newsvendor problem. Individual differences in cognitive reflection (as measured by the CRT) predict performance as measured by expected profit, order quantity, and order quantity variance. Hypotheses H1, H2, and H3 are supported, which means that individuals with higher cognitive reflection are statistically more likely to perform better in the newsvendor problem environment. Cognitive reflection is also directly related to use of other anchoring heuristics and preferences, which shows that the strength of those errors and biases are not homogeneous. H4a, H4b and H5 are supported. In addition, even when the two anchoring heuristics are included in a mediation model, cognitive reflection also directly impacts performance (H4c). Lastly, this research investigated other possible individual covariates such as college major, years of experience, and managerial position. While some of these factors were modestly significant predictors of performance in their own right, cognitive reflection was clearly the best predictor of performance (H6). Across the sample of experienced practitioners, individuals who regularly use a System 2 process to moderate a System 1 response (as measured by the CRT) performed better in this experiment. This makes a strong case for considering individual heterogeneity in cognitive reflection in supply chain decision-making contexts such as the newsvendor problem.

These results provide a first step toward answering Bolton and Katok’s (2008) call for robust behavioral theory with respect to multiple order quantity misjudgments, with several implications for future research. Most prior studies have reported the average results of decision makers in newsvendor studies. Important individual differences in cognition should be considered in newsvendor experiments since individual heterogeneity can have significant effects on the results. Practically speaking, when conducting future behavioral supply chain experiments, researchers should not expect a sample population of management students at one university to perform as well as a group of engineering-oriented students at another university due to differences in cognitive reflection. As a research implication, when simulating expected supply chain performance, we believe it is appropriate to consider individual heterogeneity relative to some of the known heuristics and preferences outlined above. Overall, we find that applying aspects of Dual Process Theory can explain a significant amount of the variance in observed behavior. Cognitive reflection is supported as a behavioral indicator of performance in the newsvendor setting, and this is robust relative to several other alternative explanations.

This research also has direct and implied implications for practitioners. The results suggest that using the CRT as a screening tool may help managers select individuals who perform better in a newsvendor task environment. In particular, individuals with low cognitive reflection performed worse than medium and high reflective individuals on all three metrics. These results may also be useful in developing training and education programs by helping individuals recognize their own decision-making tendencies. For example, it may be possible to develop training exercises that encourage or develop the use of System 2 processes to moderate System 1 responses in this supply chain context. In addition, the results may help in the design of decision support systems that interact with human decision makers. Such design aspects could include providing a moving average of demand rather than emphasizing the actual prior period demand, which would dampen one source of variation and encourage more systematic responses. Anecdotally, one of our participating firms was considering new software promoted as being more reactive to customer demand. After seeing the results of this study at their firm, senior managers were alerted to the implications of a more reactive system. Appropriate design mechanisms could be especially important where individuals are likely to have an inappropriately high order-quantity variance. Lastly, managers are cautioned that when faced with stochastic demand, individual employees do not react the same across several known decision heuristics. Managers may be able to attenuate aspects of sub-optimal behavior based on how they ask questions, perhaps by not only asking for sales during the last period but also asking how recent sales compare to the trend.

This study has several limitations. First, several individual respondents with low CRT scores still performed well in the newsvendor task, which suggests that the CRT is not a perfect predictor of success. This should be of comfort to practitioners, who may be able to perform at a high level regardless of their CRT score, particularly when they possess relevant, task-specific experience. Second, this research is based on an experimental task environment, using a design that is similar to other newsvendor studies. Based on feedback from the participating firms, the experimental task was generally relevant to the inventory task facing practitioners. The participating firms indicated that high-performing individuals in the experiment are likely to also have high performance in actual supply chain settings. Although the experiment is empirically grounded, practicing managers and analysts may be able to perform better with significant training, experience, or use of task-specific software that we are unable to replicate in an experimental setting. Third, other behavioral or cognitive factors or measurement instruments may be better predictors of success, which could be the subject of future research.

These findings suggest a number of possible follow-up studies. This research is focused on a high-margin newsvendor context. Low profit conditions may be more likely to activate cognitive dissonance (Festinger, 1957) because the goal of profit maximization may be in direct conflict with the goal of satisfying customers and minimizing lost goodwill. Such potential cognitive dissonance may be stronger than Dual Process-based cognitive reflection. This research could also be applied to other, more complex supply chain contexts such as multi-stage inventory problems. We expect that the individual heterogeneity present in this simple newsvendor setting will also impact performance in other inventory settings where replenishment decisions are made across multiple periods, such as those subject to the bullwhip effect (Lee *et al*. 1997). Similarly, it would be interesting to extend this research to other supply chain decision contexts such as forecasting and inventory pooling where Dual Process Theory might provide insights into decision-making behavior. Consistent with previous studies of learning in the supply chain (cf. Bolton and Katok, 2008), individual cognition may impact how decision makers approach problems, how they search, and how they learn when faced with supply chain tasks. Lastly, this research could be extended to consider the impact of team-based decision-making, including differences among leaders and overall team composition.

Chapter 3

Essay 2: Asymmetric Ordering Behavior in Newsvendor Inventory Decisions: Customer Service and Cognitive Dissonance

Chapter Summary

When solving a newsvendor problem, individuals systematically and persistently deviate from the profit maximizing order quantity. Previous research has shown that individuals tend to order too much in a low margin setting and too little in a high margin setting. This "pull to center" effect is asymmetric relative to the mean of the demand distribution and is typically about twice as strong in low margin contexts than in high margin contexts. This research posits that cognitive dissonance surrounding customer service expectations causes this behavioral anomaly. Specifically, planning to *a priori* disappoint customers causes cognitive dissonance even though doing so maximizes expected profit in a low-margin setting. When individuals experience cognitive dissonance, they attempt to mitigate it. We test this proposition in a behavioral experiment with 127 participants, showing that the amount of dissonance mitigation differs based on margin condition. We also compare these results to prior research using cognitive reflection. These results suggest that cognitive dissonance based on customer service expectations explains a portion of the observed asymmetry in ordering behavior in low margin conditions. This research also confirms that cognitive reflection explains a portion of the behavior in the high and medium margin conditions. This insight into how cognitive dissonance impacts supply chain decision making contributes to behavioral operations management theory. These results inform managerial decisions regarding employee selection, training, task design and reward structures.

3.1 Introduction

In many contexts, a manager must order inventory in advance of knowing customer demand. With estimates of the unit cost, price, customer goodwill, salvage value, and parameters of the demand distribution, the newsvendor model (cf. Porteus 2002) can be used to determine the order quantity that maximizes expected profit in a single selling season (or period). If the inventory manager orders more than the realized demand, excess inventory must be salvaged or scrapped; if the manager orders less than the realized demand, revenue, profit and customer goodwill are lost. A number of behavioral studies have shown that individual decision-makers perform sub-optimally, even when they are familiar with the newsvendor model and have the required parameters available. If individuals apply the normative solution, they will maximize expected profit in the absence of other individual preferences that may influence ordering behavior.

Several studies have shown that observed behavior is not consistent with the goal of maximizing expected profit. Schweitzer and Cachon's (2000) well-known study found that individuals regularly order too little in high margin settings and too much in low margin settings.[1] A number of heuristics and preferences have been proposed that partially explain this behavior. In particular, individuals may anchor on the mean and insufficiently adjust toward the optimal order quantity, may have a preference to reduce *ex post* inventory error, or may tend to chase prior period demand. Other explanations such as bounded rationality (Su 2008), experience with a particular setting (Bolton and Katok 2008), and managerial experience and training have also been investigated (Bolton *et al.* 2008). Moritz *et al.* (2009) found that the dual process construct of individual cognitive reflection explained a significant portion of the variance in performance in high margin settings, but cognitive reflection did not predict performance in a low margin setting. Similarly, a number of recent papers have observed a behavioral asymmetry based on the margin condition. Schweitzer and Cachon (2000) note that, "This too low/too high pattern of choices was not symmetric across low and high profit conditions. Orders were closer to expected demand for low profit products than for high profit products." (p. 418). However, the specific causes of this observed asymmetry between margin conditions has remained an open question. This paper proposes that cognitive dissonance (Festinger, 1957) associated with customer service in the low margin context is a primary cause for the asymmetry, and tests this hypothesis in a behavioral experiment.

---

[1] The high margin condition is defined as a critical ratio greater than 0.5 and a low margin condition as a critical ratio less than 0.5, where the critical ratio is defined as $c_u/(c_u + c_o)$, with underage cost c_u (price – cost + customer goodwill) and overage cost c_o (cost – salvage value) (cf. Silver *et al.* 1998).

The remainder of this paper is organized as follows. The next section develops the theory for the paper starting with newsvendor research on the asymmetry between margin conditions followed by an overview of cognitive dissonance theory. Section 3 develops the research hypotheses and outlines the experiment, which places individuals in low, medium and high margin settings. Section 4 reports the experimental results and section 5 concludes that paper with a discussion of implications and future research.

3.2 Theory Development

3.2.1 Behavioral Asymmetry in Margin Conditions

Regarding asymmetry in order quantity based on margin condition, Bostian *et al.* (2008, p. 608) conclude, "We view exploration of the asymmetry as the most interesting line of inquiry for future newsvendor experiments and modeling of experimental data." Average order quantities are about twice as far above the mean in high margin conditions when compared how far the average order quantity is below the mean in corresponding low margin conditions. This observed asymmetry has not been explained by prior research, and we follow this motivation in the behavioral experiment outlined later in this paper.

Many studies have observed this asymmetry: In study 1 of Schweitzer and Cachon (2000), average order quantities were 18 percent above the mean demand in the high margin condition and 10.6 percent below the mean in the low margin condition. In study 2a, average orders quantities were 25 percent above and 5 percent below the mean demand, while in study 2b order quantities were 4 percent above and 2.7 percent below the mean in the high and low margin conditions respectively (Schweitzer and Cachon, 2000). These results are similar to the baseline condition findings in Bostian *et al.* (2008), where average orders quantities were 32 percent above the mean and 22 percent below the mean in the high and low margin conditions. In the doubled payoff condition, orders were 22 percent above the mean and 12 percent below the mean in high and low margin conditions (Bostian *et al.* 2008). In the base condition of Bolton and Katok (2008), orders were 22 percent above the mean and 12 percent below the mean in the high and low margin contexts. In the neutral frame condition of Kremer *et al.* 2010, orders were 11 percent above the mean and 7 percent below the mean, while in the operations frame condition, orders were 18 percent above the mean and 10 percent below the mean in the high and low margin conditions (Kremer *et al.* 2010).

In studies designed specifically to investigate the effects of order frequency and feedback frequency, the observed asymmetry was not clear. Bostian *et al.* (2008) found that the average order quantity declines with decreasing order frequency. In Lurie and Swaminathan (2009), the asymmetry was not consistent with other research. While specifically investigating the impact of feedback frequency (study 2 of Lurie and Swaminathan 2009), order quantities were 6% above the mean in the high margin condition, but 37% below the mean in the low margin condition. Still, overall results in nearly all experimental investigations indicate that observed asymmetry appears to be nearly twice as strong in the high margin versus low margin conditions (Bostian *et al.* 2008). The research presented in this paper seeks to explain this observed asymmetry in ordering behavior based on margin context.

3.2.2 Cognitive Reflection and Performance

Prior research has shown that the dual process construct of cognitive reflection (Frederick, 2005) predicts performance in a high margin newsvendor environment but not in low margin settings (Moritz *et al.* 2009). However, prior research has not investigated cognitive reflection and order asymmetry across margin conditions, as well as potential interactions between cognitive reflection and cognitive dissonance. As a brief description, cognitive reflection has its roots in dual process theory, a broad framework describing the two types of decision making as proposed by many cognitive psychologists (cf. Stanovich and West 2000). System 1 processes are rapid, context-specific intuitive responses to a problem while System 2 processes are logical, structured, analytical approaches to solving the same problem. Typically, both System 1 and System 2 processes are active in a decision making task. Frederick (2005) proposed the construct of cognitive reflection and developed a measurement instrument called the Cognitive Reflection Test (CRT). The CRT measures the tendency of individual to allow System 2 processes to override or endorse a System 1 answer that may be immediately available (Kahneman and Frederick, 2002).

In a study of experienced practitioners (n = 313), Moritz *et al.* (2009) found that individual cognitive reflection (as measured by the CRT) predicted a significant portion of the variance in performance in a newsvendor experiment in a high margin environment (Moritz *et al.* 2009). Individuals with higher cognitive reflection tended to have higher expected profits, lower order variance, anchored less on the mean, had a lower tendency to chase most recent demand and had a lower apparent tendency toward a minimizing *ex post* inventory error preference. However, an experiment in a similar low margin context (n = 47, again experienced practitioners) found no difference in performance based on cognitive reflection. In a follow-on study to Moritz *et al.* (2009), 88 participants were placed in both high and low margin conditions, with approximately equal numbers seeing each margin condition first. Cognitive reflection was again a good predictor of performance in a high margin condition ($F = 4.149$ (2,85) $p = 0.019$), but did not predict performance in a low margin condition ($F = 0.559$ (2,85) $p = 0.579$). The main difference between the low and high margin conditions in this experimental setting was treatment of and value attributed to the cost of lost customers. Based on these results, we looked for additional theory that would explain the observed behavior, and turned to cognitive dissonance theory as a potential alternative explanation.

3.2.3 Cognitive Dissonance Theory

While the literature on cognitive dissonance is large (see Cooper and Fazio 1984 or Cooper 2007 for excellent reviews), Festinger's original theory (1957) was relatively simple. Cognitive dissonance occurs when an individual must make a decision between two or more alternatives, and more dissonance is created when the choice is difficult. Simply holding two or more cognitions does not automatically imply dissonance if those cognitions are unrelated, i.e., it is possible to know the weather forecast for tomorrow and the cost of a new computer. However, if a person holds two alternative and related cognitions, those alternatives are dissonant if one of them implies a decision opposite of the other (i.e., I can purchase either the sports car or the SUV, but not both). Individuals experiencing cognitive dissonance seek to reduce it. This need to reduce dissonance is a human drive that has been likened to hunger or thirst (cf. Aronson, 1997). The amount of dissonance and the pressure to reduce it are larger if both alternatives are similar in their desirability (Brehm, 1956). The total amount of dissonance can be considered the sum of all discrepant factors times their importance divided by the sum of all consonant conditions times their importance (Cooper, 2007).

In Brehm's (1956)[2] classic study of cognitive dissonance in consumer behavior, 225 female college students were asked to rate eight different household appliances. After reviewing their ratings, the experimenter gave the subjects their choice of two appliances. In the high dissonance condition, the choice was between two similarly rated appliances (i.e., second and third), while in the low dissonance condition, the choice was between appliances rated differently (i.e., second and eighth). After including some other manipulations, the subjects were asked to rate the items for a second time. Relative to their initial values, the ratings of the chosen alternatives were higher, while the ratings of the alternative not selected were lower. Taken on its own, this observation might appear to show evidence of the endowment effect (cf. Thaler, 1980) relative to possession of the appliance. However, subjects had possession of the same appliance in both dissonance conditions, so the endowment effect does not explain the observed change in ratings for the item that was not chosen. More importantly, the change of the mean ratings was greater for the high dissonance conditions, and numerous other studies have confirmed these general results.

Several important conditions associated with dissonance theory have been added as more experimental work has been completed (Cooper, 2007). We summarize the four key conditions outlined by Cooper (2007) and suggest an inventory management context for each one. First, individuals must have freedom in the decision, so a choice made under duress or mandated by managerial policy is not subject to the same amount of dissonance. For example, if company policy is to always order the mean demand, the decision maker has no real choice and may psychologically revoke the decision by

[2] As a historical note, Festinger was Brehm's dissertation advisor at the University of Minnesota, where the initial theory of cognitive dissonance was developed and tested, through this experiment and others.

relying on pre-set policy. Such psychological revocation ("I had no choice") tends to greatly reduce dissonance. Second, individuals must be committed to their behavior. For example, if an inventory manager has the opportunity to change the order after seeing customer demand, dissonance associated with the initial inventory quantity decision will be greatly reduced. Third, the choice must result in aversive or undesirable consequences. If an inventory decision has no potential aversive consequences (such as customer satisfaction or profitability), then no dissonance is created. Fourth, the consequences of the decision must be foreseeable. For example, if an inventory manager orders an appropriate amount but the supplier ships the wrong product, no dissonance is created specific to the initial quantity decision. While the supply failure may have other adverse consequences, such consequences were not foreseeable or caused by the selected quantity. To the extent possible, the newsvendor experiment developed in this paper meets the conditions of free choice, commitment, undesirable consequences and foreseeable consequences.

Relative to cognitive dissonance, individuals assess their performance relative to what behavior would be expected in a given situation. In effect, individuals evaluate their decision behavior relative to questions like, "Did I make the right decision? Was I competent?" The self standards model (Stone and Cooper, 2001) notes that individuals compare their performance with two referents: First, the *self concept*, which in a supply chain context refers to an individual's outlook on competence or perceived role in customer satisfaction, profitability or other factors. Second, there is a *normative social standards* approach, as to what performance or expectations are made especially salient by the environment. In an inventory management context, such normative references may include managerial statements, i.e., "We are a customer focused company" or "The customer is always right." The main difference between these two approaches is that if a specific normative standard is easily accessible, individual self-concept referents are less important relative to dissonance (Stone and Cooper, 2001).

While cognitive dissonance is a major theory of human behavior and is a fundamental theory to psychology and judgment and decision making, it has several limitations. First, it is not specific in the sense that there are limited measures of dissonance and those measures tend to be self reported. Second, in some cases it is difficult to determine if cognitive dissonance is the only factor at work in a particular decision or if other factors such as social desirability or high anxiety (cf. Oshikawa, 1972) are also at work. Third, the precise psychological and physiological mechanisms of dissonance creation remain elusive.

Because prior research suggests that because individuals have a need to reduce dissonance, investigating dissonance reduction behavior gives insight into the behavioral factors that initially created dissonance. Cognitive dissonance can be reduced in one of three modes (Festinger 1957, Simon *et al.* 1995). With the first mode, individuals change their behavior or attitude or even revoke a previous choice. In the context of consumer behavior, this might mean returning the sports car to the dealer and buying the SUV instead. While changing or revoking behavior may happen in some instances, there is limited research about revocation and repeated decision contexts that are frequently observed in supply chain environments. Most existing literature surrounds irrevocable commitments or one-time choices, rather than the context of repeated decisions. The implications of revocation on repeated choices may be an area for future research, but is beyond the scope of this paper. The second mode of reducing dissonance is by adding consonant cognitions that would reduce the inconsistency between alternatives, such as seeking out new information that would support the choice. For example, reading more sales literature about a recent purchase or seeking confirmation from friends who have made a similar choice would add consonant information supporting the choice behavior, though the extent of such behavior is unclear in supply chain context. The third mode is "decreasing the importance of elements involved in the dissonant relations" (Festinger, 1957 p. 264). This third mode is referred to as trivialization (Simon *et al.* 1995) where an individual reduces the importance of specific cognitions to reduce dissonance.

3.2.4 Research Hypotheses

Our research is designed to investigate cognitive dissonance in inventory decisions. In particular, we are specifically interested in cognitive dissonance caused by customer service expectations and how dissonance differs by margin condition. In a high margin condition, selecting an order quantity that optimizes costs and profits also results in planning to satisfy more than the expected (mean) number of customers. However, in a low margin condition, these two goals are in direct conflict: In order to optimize costs and expected profit, inventory managers should plan to disappoint customers on average. We hypothesize that planning to disappoint customers creates cognitive dissonance.

> *H1: Cognitive dissonance related to customer satisfaction will be higher in a low margin condition than in a high margin condition.*

Measuring cognitive dissonance is difficult because of the three modes of dissonance reduction discussed above (Festinger 1957, Simon *et al.* 1995). Based on the first mode of dissonance reduction, we expect that individuals may reduce dissonance by increasing their order quantity toward the mean in a low margin condition, in effect changing their behavior. While this mode is intuitively appealing and is consistent with cognitive dissonance as an underlying explanation of the ordering behavior described above, the precise mechanisms behind this decision are difficult to detect and even more difficult to measure. We do not expect the second mode of dissonance reduction (adding consonant conditions) will be directly active because this experiment had no differences in the amount, kind or presentation of information available in the different margin conditions. At least in this controlled laboratory environment, there were no obvious opportunities to seek out new confirmatory or supportive information in any of the margin conditions. However, individuals may consider spurious information (such as illusory trends) or struggle more with order quantity decisions in low margin cases where customer goodwill (inclusion of the cost of lost or disappointed customers) is negligible.

We also want to validate prior research on cognitive reflection and compare those results in margin conditions where cognitive dissonance is strongest. If individuals first seek to reduce higher levels of cognitive dissonance, we expect that cognitive reflection (as measured by the CRT) is less likely to indicate performance. Such behavior is more likely to happen in low margin conditions.

> *H2: Cognitive dissonance is stronger than cognitive reflection in low margin conditions.*

3.3 Research Design

3.3.1 Experimental Design

Our research design was specifically geared toward investigating the third mode of dissonance reduction, trivialization. To test for trivialization in inventory decision making, we developed a two-part research design, separated by several days. In the first part of the study, a five-minute pre-test was performed. We collected basic demographics, and subjects were given basic information about the upcoming study. Subjects were asked to both rank and rate (1 = unimportant, 5 = extremely important) the factors they considered important in an inventory order quantity decision. A subset of the instructions for the pretest is listed in the appendix. The pre-test instructions were written to be easily understandable by the subjects, give them enough information to understand the kind of inventory management task they would be facing, and be as short as possible. Subjects were told that they would have to order inventory in a retail environment in advance of knowing actual customer demand and that their compensation would be based on minimizing the total costs[3] throughout the simulation. However, no detailed instructions were given on how to minimize those costs. The limited information was designed to reduce the chances of enterprising subjects studying the newsvendor problem prior to the experiment and thus improving their performance when participating in the inventory simulation.

The first part of the study was web-enabled and designed to be completed five days in advance of the laboratory study. We chose to build in this time delay in order to reduce a possible carryover effect between the pre-test measures of the importance factors, decision making on the simulation, and post-test measures. While it may be possible to complete the pre-test, inventory simulation and post-test in a short time frame, we wanted the newsvendor inventory simulation to be as similar as possible to prior research. In addition, when answering the post-test questions of importance, we wanted the recent decision making experience in the simulation to be more salient than their prior answers on the pre-test.

---

[3] In the newsvendor problem, maximizing expected profit is equivalent to minimizing the expected opportunity cost plus the acquisition cost.

The second part of the study was a controlled laboratory experiment which placed the subjects in a simulated repeated newsvendor environment. Subjects were given information on how to place orders, the distribution of past demand, and assurance that future demand would be drawn from the same distribution. In addition, subjects were given the cost, profit and customer goodwill parameters for their margin condition. Each subject participated in only one margin condition. Subjects were given a benchmark mismatch cost (per period and cumulative) which would maximize their payoff for the experiment and were compensated based on minimizing the total mismatch cost. Payoffs ranged from \$10 to \$15, rounded up to the nearest whole dollar amount and were equalized across margin treatments. Cash was the only incentive offered to the participants. For each period, subjects saw a history of recent demand, their order quantity from the previous period, the realized customer demand in that period, the mismatch cost associated with their order, and a running total of the cumulative mismatch costs through the current period. After the inventory simulation was complete (but before the final payoffs were announced), subjects were again asked to rate the importance of possible decision factors, using the same format as the pre-test. We also administered the CRT instrument (Frederick, 2005). While results were anonymous, we used a unique subject identifier in the pre-test and laboratory experiment to track performance and change in rating by subject.

3.3.2 Details About the Experiment

Subjects were recruited via a posting to the student subject pool affiliated with the business school at a large university. Following Su's (2008) recommendation, simulated demand was sampled from a normally distributed data set, with mean demand $\mu_D = 100$ and standard deviation of demand $\sigma_D = 30$.[4] For the actual realized sample, the average demand was $\bar{D} = 97.0$ and the standard deviation was $s_D = 27.1$. Subjects were given 52 periods (simulated weeks) of demand history in a time series graph and as a histogram. The experimental parameters for the three margin conditions are given in Table 3.1.

[4] We did not provide subjects with either a calculating tool or a normal table. We believe that including such decision aides would have increased complexity without providing any additional insight into the behavioral factors of interest.

Table 3.1: Experimental Parameters

Margin Condition	Experimental Parameters Price (p)	Cost (c)	Goodwill[a] (g)	c_u	c_o	Critical Ratio	Optimal Order (Q^*)
High	\$4.00	\$2.00	\$4.00	\$6.00	\$2.00	0.75	120.2
Medium	\$3.00	\$2.00	\$1.00	\$2.00	\$2.00	0.50	100.0
Low	\$3.00	\$2.25	\$0.00	\$0.75	\$2.25	0.25	79.8

[a] In the low margin condition, participants were told that the customers were very loyal and would continue to shop from them even if they occasionally ran out of inventory late in the period (g = 0).

3.4 Analysis and Results

3.4.1 Overall Results

Table 3.2 presents the overall results for the repeated single-period newsvendor experiment. A total of 127 individuals participated in the study.[5] The asymmetry in average order quantities relative to the mean is clearly visible based on the margin condition (105.92 and 97.48). Consistent with prior research, the deviation between the average order quantity and the mean demand in the high margin condition was about twice as large as the deviation from the mean demand for the low margin condition. The average cash compensation was $12.69 per subject and was similar across all margin conditions.

[5] Prior to any analysis, one subject was excluded due to order quantities that we far from the average (i.e., orders of 4567 and 6789 when average demand was 100).

Table 3.2: Overall Results

	Margin Condition	*n*	*mean*	*std. dev*	*SE*	Comparison of Group Means by Margin Condition[d]
Average Order Quantity[a]	High	47	105.92	7.143	1.042	F = 16.851(2,127) p < 0.001
	Medium	28	99.36	7.282	1.376	
	Low	52	97.48	7.853	1.089	
Average Order Quantity Variance by Individual[b]	High	47	208.96	195.184	28.470	F = 0.240 (2,127) p = 0.787
	Medium	28	181.21	129.886	24.546	
	Low	52	193.44	172.665	23.944	
Expected Profit[c]	High	47	502.55	14.134	2.062	F = 300095 (2,127) p < 0.001
	Medium	28	146.57	3.888	0.735	
	Low	52	36.50	6.336	0.879	

[a] Difference between High and (Medium, Low) significant at < 0.001;
Difference Low to Medium not significant.
[b] No group comparisons significant
[c] All group comparisons significant at < 0.001
[d] ANOVA results, all group comparisons via Tukey HSD

3.4.2 Cognitive Dissonance and Performance

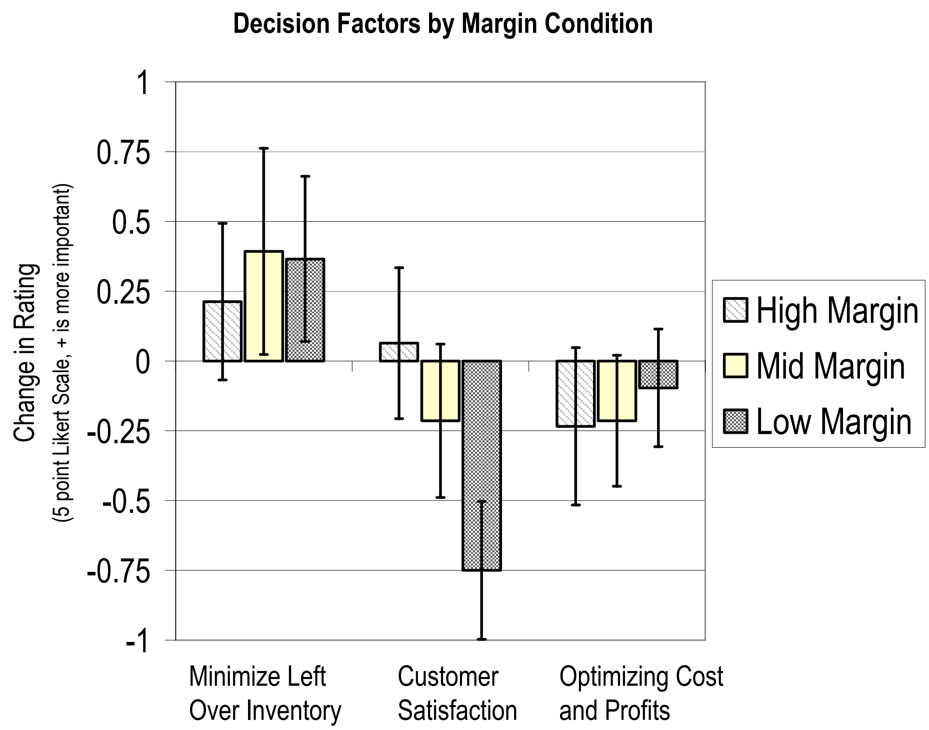
Next, we compared the pretest and post-simulation ratings of importance by subject.[6] A comparison of the change in importance by subject and by margin condition is shown in Table 3.3 and Figure 3.1.

[6] We collected both rating data (1-5 Likert scale) and ranking data. The ranking data and rating data results were not statistically different (Goodman-Kruskal gamma, $p < 0.01$). Therefore, we present only the rating data here.

Table 3.3: Dissonance Reduction by Trivialization

	Margin Condition	*n*	Mean change in rating	std. dev	*SE*	*t* test, change from 0 *t*	*p-value*
Minimize Left Over Inventory	High	47	0.213	0.977	0.142	1.494	0.142
	Medium	28	0.393	0.994	0.188	2.091	0.046
	Low	52	0.365	1.085	0.150	2.428	0.019
Customer Satisfaction	High	47	0.064	0.942	0.137	0.465	0.644
	Medium	28	-0.214	0.738	0.140	-1.536	0.136
	Low	52	-0.750	0.905	0.125	-5.978	<0.001
Optimizing Costs and Profits	High	47	-0.234	0.983	0.143	-1.633	0.109
	Medium	28	-0.214	0.630	0.119	-1.800	0.083
	Low	52	-0.077	0.788	0.109	-0.704	0.485

Figure 3.1 Dissonance Reduction by Trivialization



The results specific to customer satisfaction are particularly important in this research. While no statistically significant changes in importance were found in the high and medium margin conditions, the mean importance rating declined significantly in the low margin condition ($t = -5.978$ $p < 0.001$). This is consistent with trivialization as a mode of dissonance reduction (Simon *et al.* 1995). An alternative explanation might suggest that customer satisfaction should be lower in a low-margin setting due to the low cost of a lost customer and the relatively high cost of the product. However, a consistent line exposition based on this alternate explanation would also argue for a corresponding increase in customer satisfaction in the high margin setting. Given that there was no statistical increase in customer satisfaction in the high margin setting ($t = 0.465$, $p = 0.644$), this alternative explanation is not supported. Therefore, we conclude that trivialization of the dissonant customer service cognition is supported in the low margin context and H1 is supported.

The other two decision factors showed modest though directionally similar adjustments with experience in the simulation: Minimizing left over inventory appeared to be slightly more important and optimizing costs and profits appeared slightly less important, though these results were either marginally significant or not significant. These apparent shifts may be due to the participant experience in the simulation, such as exactly how the participants would be compensated, saliency of left over inventory and specific detail on the costs of left over inventory. In contrast to the results of customer satisfaction, the adjustments are both modest and directionally similar across margin conditions, suggesting that minimizing dissonance specific to customer satisfaction (H1) has the greater impact on performance.

3.4.3 Cognitive Reflection and Performance

Next, we evaluated performance by CRT group. Following Frederick (2005), the total number of correct responses on the CRT instrument was calculated for each participant, with those scoring 2 or 3 combined into one larger group. These were also broken out by margin condition, as shown in Table 3.4.

Table 3.4: Cognitive Reflection and Performance by Margin Condition

Margin Condition	Performance Measure	Optimal Value[a]	*CRT = 0*	*CRT = 1*	*CRT = 2,3*	Comparison of Group Means by Margin Condition[b]
High	*n*		*8*	*10*	*30*	
	Order Quantity	120	102.1	102.5	108.1	$F = 4.087\ (2,45)\ p = 0.023$
	Order Quantity Variance	-	329.3	272.8	155.8	$F = 3.519\ (2,45)\ p = 0.038$
	Expected Profits	$523.73	$491.50	$495.80	$507.90	$F = 7.559\ (2,45)\ p < 0.001$
Medium	*n*		*11*	*5*	*12*	
	Order Quantity	100	95.2	100.0	102.9	$F = 4.034\ (2,25)\ p = 0.030$
	Order Quantity Variance	-	183.9	291.6	132.7	$F = 3.046\ (2,25)\ p = 0.065$
	Expected Profits	$152.13	$145.34	$145.00	$148.30	$F = 2.431\ (2,25)\ p = 0.108$
Low	*n*		*16*	*14*	*22*	
	Order Quantity	80	96.8	99.4	96.7	$F = 0.589\ (2,49)\ p = 0.559$
	Order Quantity Variance	-	231.0	194.2	165.6	$F = 0.657\ (2,49)\ p = 0.523$
	Expected Profits	$46.40	$36.20	$35.50	$37.40	$F = 0.394\ (2,49)\ p = 0.677$

[a] Optimal values based on the demand distribution and given margin parameters

[b] ANOVA results for each performance measure in a margin condtion.

Based on these results, it is clear that cognitive reflection is able to predict performance in the high and medium margin conditions. However, this result does not hold in the low margin condition where the group means are consistently close together and statistically indistinguishable. These results are consistent with Moritz *et al.* (2009). Given that we observe cognitive dissonance specific to customer satisfaction in the low margin condition, this suggests that the decisions are first based on drive to minimize cognitive dissonance. Second, contingent on the margin condition where dissonance relative to customer service expectation is not present, individuals with higher cognitive reflection perform better.[7] This supports H2.

3.5 Conclusions

The normative solution to the newsvendor problem is well known and the newsvendor model is taught in many (if not most) supply chain management courses. A common expectation is that individuals should be rational, seek to maximize expected profits and

[7] We also checked for interaction effects between cognitive dissonance, cognitive reflection and performance, but none of these results were statistically significant. Given the relatively small differences in effect sizes in the low margin condition, it might be that a very large sample size is required.

generally consistent in their decisions as the economic parameters change. However, much prior research has indicated that individuals systematically and persistently deviate from profit maximizing choices for the newsvendor problem, and these deviations occur with both highly functional MBA students (cf. Schweitzer and Cachon 2000) and experienced practitioners (cf. Bolton *et al.* 2008, Moritz *et al.* 2009). In addition, several studies have observed the asymmetric order behavior in high and low margin conditions and have suggested it as a particularly promising line of inquiry for future experimental and modeling work (Bostian *et al.* 2008). We have observed a similar asymmetry in ordering behavior where demand is sampled from a normally distributed data set.

This research has proposed and tested the hypothesis that cognitive dissonance specific to customer service expectations is one explanation of the observed asymmetric ordering behavior. In particular, individual decision makers employ trivialization as a primary mode of dissonance reduction when faced with a low margin newsvendor setting. In such a setting, the optimal newsvendor solution is to plan to dissatisfy some customers in each period, and this appears to create dissonance. Cognitive dissonance also appears to be stronger than cognitive reflection, suggesting that individual decision makers first act so as to reduce their dissonance and then are impacted by other factors such as cognitive reflection. While cognitive reflection predicts performance in high and mid-margin settings, the presence of cognitive dissonance (as in a low-margin setting) appears to have a stronger impact on newsvendor inventory decisions.

These findings have several implications for practitioners and academics. First, these experimental results confirm the notion that individuals are not motivated solely by expected profit and appear to change the weights they put on decision factors *depending on the margin context.* Our experimental results show that decision makers decrease the importance that they place on satisfying customers in a low margin setting, but do not increase the importance that they place on satisfying customers in a high margin setting. Such behavioral anomalies are obviously not included in the normative theory, but are active in a descriptive observation of actual behavior. Second, mitigating dissonance suggests that firms and managers should consider self-expectations of performance specific to cognitive dissonance (see section 3.2.3). Where individuals have a self-concept as a referent, they may think about their own performance and make a judgment about their competency. However, if there is a strong normative social standard at work in an organization, this provides an easily accessible, salient reference point for individuals to evaluate their behavior. Companies, departments and individual managers have a substantial role in developing these normative social standards for performance. This is especially true as many employees have had prior experience in customer-facing service positions. If managers repeatedly emphasize customer service rather than optimal low-margin decision making, this provides a powerful behavioral incentive and normative standard for employees in evaluating their decisions. Third, these self-expectation models have implications on the design of reward systems: Many annual reviews emphasize external or internal customer service, and this could motivate or emphasize different inventory decision-making behavior.

This research has some limitations and opportunities for future research. This research is limited in that it uses a student subject pool rather than experienced practitioners who may differ in their ability to understand and manage tradeoffs associated with inventory decisions. Additionally, there may be context-specific factors that we are unable to replicate in an experimental setting, or other explanations of observed behavior. This may include other measures of the individual tendency to want to please others, or other specific individual factor(s) that may better explain observed behavior. With respect to opportunities for future research, there appears to be limited research on cognitive dissonance and repeated decisions in general, so additional specific research in this area appears warranted. More specific to the newsvendor problem context, a follow on study that tests other modes of dissonance reduction may provide interesting results, perhaps by varying the degree of emphasis on customer service. In some environments, shortages (lost sales) are observed while they may be hidden in other contexts. The visibility and saliency of shortages may have an impact on the amount of dissonance created by a decision. In contexts where managers are responsible for multiple products, dissonance based on customer service expectations may cut across product or margin contexts. There also may be an interaction between experience in a low-margin context and the steps taken to mitigate dissonance. Additionally, this research is focused on individual decision makers. We believe that the group effects of cognitive dissonance may impact performance and this could be the subject of future research. Lastly, while this research shows an application of cognitive dissonance theory to newsvendor inventory decision making in a relatively simple context, we believe other supply chain management topics can also be explored through this theoretical lens. For example, many firms have tension between their conflicting goals of motivating sales with an aspirational forecast while holding manufacturing responsible to build to an operational plan that minimizes excess inventory. A large gap between an aspirational forecast and planned production may create friction in the organization, especially in conditions where maximizing expected profit results in a plan to disappoint a large number of customers. Recognizing, evaluating and mitigating these dissonant cognitions may open up further research opportunities.

Chapter 4

Essay 3: System Neglect and Individual Differences in Forecasting Behavior

Prediction is very difficult, especially about the future. – Niels Bohr

Chapter Summary:
This chapter analyzes how individuals make forecasts based on time series data. Based on a behavioral model of forecasting, we compare individual performance using a behavioral experiment. We find evidence that individuals with high cognitive reflection demonstrated higher performance and are more consistent in their forecasts, demonstrating lower system neglect. We also compared the amount of time that it takes for individuals to reach a decision. Either responding too quickly or taking too much time is detrimental to performance, and individuals with high cognitive reflection had times that were closer to the estimated optimal time.

4.1 Introduction

Demand forecasting is central to successful management of inventory, capacity and customer satisfaction. In many demand environments, decision-makers must make and update forecasts based on historical time series data. Forecasting is also a central antecedent to placing orders for inventory in advance of knowing customer demand. In fact, Schweitzer and Cachon (2000) specifically identify forecasting as a task requiring managerial judgment and suggest that the order quantity decision (as in a newsvendor problem) can be automated once the forecast is complete. In the context of forecasting, while quantitative methods such as moving averages, exponential smoothing, regression and time-series decomposition are available, firms frequently use judgmental forecasts or a combination of methods incorporating human judgment (Sanders and Manrodt 2003). Even of firms that primarily employ quantitative methods, 77.5% report moderate or high degrees of managerial adjustment to software-generated forecasts (Sanders and Manrodt 2003). While the quantitative methods are well known and widely available, far less is known about how individuals generate forecasts or make adjustments to statistically-generated forecasts. Individual judgment may be subject to many systematic biases and errors (Hogarth and Makridakis 1981), anchoring and adjustment (Lawrence and O'Connor, 1992) and a range of other suboptimal behaviors (Lawrence *et al.*, 2006). Some of these behaviors may be tied to the system neglect hypothesis, the idea that individuals react to the observed signals and tend to ignore the underlying system that generated the signals (Massey and Wu, 2005). Recent research has found support for the system neglect hypothesis in time series forecasting (Kremer *et al.*, 2010). Cognitive reflection (the tendency of an individual to let his/her analytical/systematic cognition to override or endorse the intuitive answer that may be immediately available) has been shown to be related to individual level risk preferences and time preferences in preferred outcomes (Frederick, 2005). Similarly, one would expect that taking more time to reach a decision might improve forecasting performance, but there is relatively little research in a time series context. Hence, this paper investigates aspects of individual-level forecasting behavior, including the relationship between cognitive reflection and system neglect, as well as the time to reach a decision.

We investigate individual-level time series forecasting behavior in a controlled laboratory setting. This forecasting task had participants observe a time series of demand and make a forecast over sequential periods. The time series was generated from a random walk with noise, such that the stability of the time series in the amount of change and noise was precisely controlled. This artificial data generating function allowed us to compare performance to normative predictions based on perfect knowledge of the generating function. Individuals with high cognitive reflection performed better (lower absolute error) than those with low cognitive reflection. Consistent with Kremer *et al.* 2010, we observe that individuals exhibit system neglect by over-reacting to demand signals in stable environments and under-reacting in less stable environments. However, individuals with high cognitive reflection react closer to the optimal value, particularly in conditions that have high noise. Looking at the time to reach a decision, individuals who took longer to generate their forecast performed poorer than those who took a shorter time, though answering too quickly was also detrimental to performance. Individuals with high cognitive reflection typically had decision times closer to the predicted optimal time for each condition and had lower forecast error.

We proceed in this paper as follows: The next section provides an overview of the literature, while § 3 covers the theoretical developments. Section 4 discusses the results, while § 5 offers a discussion and conclusion.

4.2 Background and Related Literature

4.2.1 Judgmental Forecasting

A number of previous studies have compared judgmental forecasts with quantitative techniques. Though the results are decidedly mixed, Lawrence *et al.* (1985) show that judgmental forecasts perform as well as or better than quantitative methods, while Makridakis *et al.* (1993) find no improvement from inclusion of judgmental factors. Other studies find that including judgment results in poorer performance (Carbone and Gorr 1985, Sanders 1992). A recent study shows that large judgmental adjustments improve forecasting performance, while smaller adjustments led to worse forecasts (Fildes *et al.* 2009). These mixed results may be partially attributed to the two main types of adjustments to statistically-generated forecasts. First, there are identifiable causal factors or tacit knowledge that individual managers may use to adjust a statistically-generated forecast. For example, a hardware store may have past data on weekly rock salt sales. Using time series decomposition, a forecast of demand can be statistically generated and used to place an order for salt. However, a manager might recognize that a major winter storm is likely and adjust the demand forecast in advance of this specific causal factor. The software that generated the initial forecast does not have access to such information, so the managerial adjustment is causally identifiable and may be an example of a large adjustment leading to improved performance (Fildes *et al.* 2009). Unfortunately, such adjustments are idiosyncratic to a specific managerial context and difficult to generalize.

The second main type of adjustment is due to individual level factors unrelated to the external context. These individual-level factors may be considered managerial biases and because of these biases, warnings against including any human judgment in forecasting have been well documented (cf. Hogarth and Makridakis 1981). These biases may include anchoring and insufficient adjustment, chasing, illusory correlation, optimism, hindsight bias and a range of other heuristics and preferences. Despite warnings against these behaviors, many firms still incorporate human judgment in forecasting decisions because the adjustments are believed to have a positive impact on performance. The perceived benefits of including causal factors may improve performance more than the possibility of reduced performance due to other individual-level biases. For this reason, comparisons of judgmental point forecasting often exclude domain-specific factors where the forecaster has no outside knowledge not contained in the time series (Lawrence *et al.*, 2006). However, we still know very little about individual-level factors which influence the final forecast. The goal of this research is to better understand these factors and to provide guidance to decision makers who work in this area.

4.2.2 Behavioral Operations

This research falls within the sub-field of behavioral operations management (Bendoly, Donohue and Schultz 2006). This field seeks to study human behavior and cognition in ways that impact the design, management and improvement of operating systems (Gino and Pisano, 2006). These questions are typically addressed with experimental studies, where participants are asked to make operations decisions in several contexts. In an inventory management context, Schweitzer and Cachon (2000) place individuals in a newsvendor environment and observe that individuals order too much in a low-margin condition and too little in a high margin condition. This bias toward the mean demand is quite robust and has been replicated in many other independent studies (cf. Bolton and Katok 2008; Kremer, Minner and Van Wassenhove, 2010). A number of potential heuristics and preferences have been proposed to describe these errors including minimizing ex-post inventory error; anchoring and insufficient adjustment, anchoring on the mean or random choice errors (Su 2008). All of these are examples of behavioral deviations from the normative newsvendor solution, a variant of which has been in the literature for more than a century (Edgeworth 1888). Additional research in more complex, multi-echelon scenarios has confirmed the presence of the bullwhip effect, which may be caused by rational operational decisions such as responses to price fluctuations and order batching (Lee *et al.* 1997) in addition to behavioral factors such as inappropriate demand forecasting (Sterman 1989). These forecasting factors have been observed when participants are given a known and stationary demand distribution (Croson and Donohue 2003) as well as with constant and deterministic demand (Croson *et al.* 2009).

There are at least two key challenges relating to behavioral research: Group versus individual factors and a relative lack of explanatory theory as to why individuals may respond as they do. First, one of the specific challenges is that sub-optimal behavior occurs at an individual level, while research methods tend to be focused on population-specific behaviors. For example, in Schweitzer and Cachon's seminal paper (2000), average performance of all subjects was shown to exhibit a preference to minimize ex-post inventory error, as well as a tendency to anchor on the mean, where subjects frequently order too much in a low margin setting and too little in a high margin setting. Such findings have been widely replicated in many studies (cf. Bostian *et al.* 2008; Kremer, Minner and Van Wassenhove 2010), but average performance of all individuals is different than individual performance: One individual alternating between orders of 100 and 200 may have the same average performance as an individual ordering 150 every period, but the observed behavior is quite different. Some recent research has begun to address this issue, though more work remains. Bolton and Katok (2008) note that individuals appear to draw conclusions from inappropriately small samples, but observe that the nature of the conclusions vary widely across individuals. Su (2008) applies bounded rationality to inventory decisions at the individual level. Hence, it is important to focus on individual decisions to the extent possible.

The second main challenge is a relative lack of theory that explains why individuals make sub-optimal decisions. Among other heuristics and preferences, anchoring on the mean, demand chasing and a preference to minimize ex post inventory error may explain aspects of observed behavior, but they do not explain the reasons why the behavior occurs or what individual factors may influence a subject to apply a particular heuristic.

Bolton and Katok (2008) call for better theory to explain multiple kinds of inventory misjudgments. However, an inventory choice is preceded by (a) an estimate of the probability distribution for future demand (a forecasting task) and (b) a service-level decision based on economic and customer-service goals. These antecedents may influence the inventory choice task and be more closely identified with behavioral heuristics such as anchoring and chasing. Hence, we choose to investigate the forecasting task, and our goal is to look more deeply at individual level performance.

More specifically, we look to investigate three individual factors that may explain performance: system neglect, cognitive reflection and the time to reach a decision while including other control factors such as age, education, and college major. While other factors may be of interest, the next section outlines the theoretical foundation for a focus on these areas and the subsequent experimental results reported in investigated in § 4.

4.3 Theory

4.3.1 Judgment in Forecasting

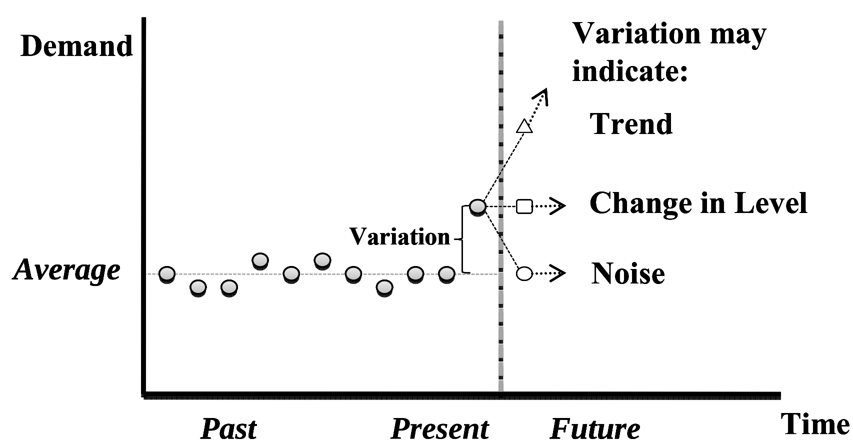
Several studies have investigated the processing steps made by individuals in judgmental forecasting. These tend to focus on inclusion of systematic change such as trend or seasonality in a time series forecasting task (cf. Andreasson and Kraus 1990; Lawrence and O'Connor 1992; Bolger and Harvey 1993). While the tasks and definitions differ, many studies use some version of the anchor and insufficient adjustment heuristic (Harvey, 2007). Often, recent demand is the proposed anchor (Epley and Gilovich 2001), while adjustments are made toward the level. However, over-adjustment tends to be more common than under-adjustment (Lawrence and O'Connor 1995). In addition, feedback might be included (Sanders 1997) while the nature of the task environment may impact performance (Lawrence *et al.* 2006).

Part of the challenge in forecasting is to separate random noise from structural change. While not specific to time series forecasting, this task has been the subject of literature on regime change detection (Barry and Pitz 1979). The main finding is that individuals overreact in environments characterized by stability and high noise, while under-react in environments that lack stability but with low noise (Griffin and Tversky, 1992). These findings have been extended by Massey and Wu (2005), who note that under the system neglect hypothesis, individuals place too much weight on the recent signal while ignoring the underlying system which generates the signal. Kremer *et al.* (2010) find support for the system neglect hypothesis in time series forecasting. In the time series context we study, we also focus on time series which have no underlying trends or seasonal factors, and where there are no outside causal factors which would support managerial intervention.

4.3.2 Behavioral Forecasting

In generating a forecast, both quantitative and judgmental methods include the most recent data point. While the weight on any particular point is well-specified under quantitative methods, the behavior of an individual is less certain. Using judgmental forecasting, when the recent data point differs from a long-run average, that individual must decide if the variation implies a change in the level, a change in the trend, or is just random noise and should be ignored. This is shown graphically in Figure 4.1:

Figure 4.1 Variation and Forecasting



We can characterize the forecast F_{t+1} made in time period t (for demand in period t+1) by the following equation, where the forecast is a function of the level L_t, a possible trend component T_t and "trembling hands" noise ε_t.

$$F_{t+1} = L_t + T_t + \varepsilon_t \tag{1}$$

However, unless individuals have perfect knowledge about future demand, time series forecasts will incorrect at least some of the time. Hence, individuals must have some opportunity to adjust a forecast in response to the difference between the actual and the forecast, $E_t = A_t - F_t$. In looking at an estimate of level, the size and direction of the adjustment is typically considered in the next forecast, so a model includes a smoothing factor α, the weight placed on the most recent forecast θ_L, an initial anchor 1-θ_L, and an initial constant C.

$$\begin{aligned} L_t &= \theta_L F_t + \alpha\left(A_t - F_t\right) + \left(1-\theta_L\right)C \\ &= \left(\theta_L - \alpha\right)F_t + \alpha A_t + \left(1-\theta_L\right)C \end{aligned} \tag{2}$$

Note that (2) is a generalization of the well-known single exponential smoothing method, which allows for individuals to consider the weight they place on recent forecasts and initial anchors.

Previous research has shown that in addition to making errors in the level, individuals are quick to see trends where they do not exist (DeBondt 1993), where a trend is defined as a systematic change in the level. Hence, a model of judgment should include an estimate of the trend and a weighted estimate of the impact of the trend β used by an individual. In addition, this estimate of a trend may vary with time and the recall factor θ_T should moderate the impact of the trend.

$$\begin{aligned} T_t &= \theta_T T_{t-1} + \beta\left(\left[L_t - L_{t-1}\right] - T_{t-1}\right) \\ &= \left(\theta_T - \beta\right)T_{t-1} + \beta\left(L_t - L_{t-1}\right) \end{aligned}. \tag{3}$$

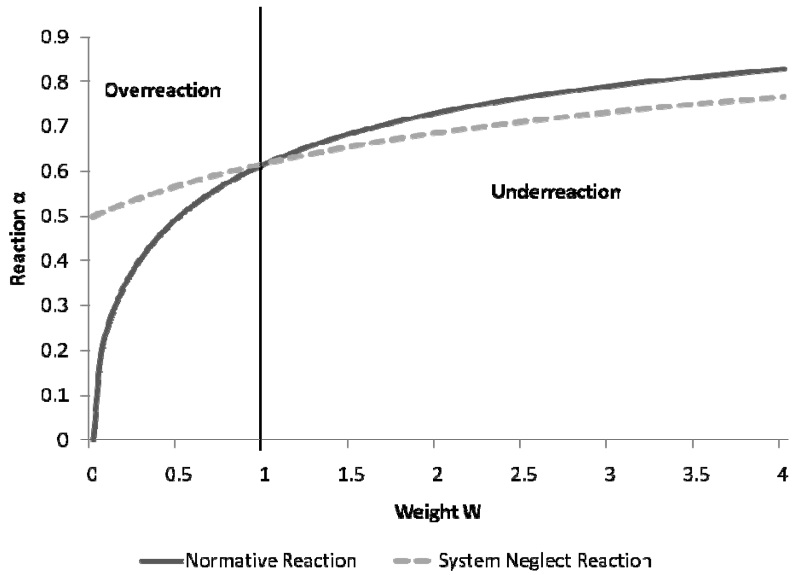
Note that (3) is a generalization of a double exponential smoothing model; where $\theta_T = 1$ it is double exponential smoothing, while if $\theta_T = \beta$, it is only a one-period change in level with no trend memory.

The task of estimating level, trend and error is also related to the system neglect hypothesis (Massey and Wu 2005). System neglect suggests that a decision maker will place more emphasis on the strength of a signal and less weight on the system that generated the signal. The immediate saliency of a signal (a demand point in this case) can be greater than recognition of the underlying properties describing demand over time. This is aligned with the concept of weight of the evidence W. This can be interpreted as the change (c) to noise (n) ratio, where $W = c^2/n^2$, or the amount of emphasis that should be placed on a particular forecast error $A_t - F_t$. If W is high, this represents large change, while if W is low, this represents noisy data. In the context of a level estimate, α should increase as W increases. If the demand can be characterized as Brownian motion (random walk with no trend; $\beta^* = 0$), single exponential smoothing is the optimal heuristic, with α^* depending only on the change to noise ratio, as follows in (4) (McNamara and Houston 1987, Harrison 1967).

$$\alpha^*(W) = \frac{2}{1 + \sqrt{1 + 4\frac{1}{W}}} \qquad (4)$$

Equation (4) allows us to calculate α^* as a normative benchmark, which depends on the weight of the evidence W and not the absolute magnitude of either change or noise. The change in alpha relative to the weight of evidence can be expressed graphically as in Figure 4.2.

Figure 4.2 System Neglect Behavior



This figure shows that where the weight of the evidence is low, individuals tend to over-react relative to the normative α. Where the weight of the evidence is high, individuals tend to under-react relative to the normative α.

To test the system neglect hypothesis, we estimate the smoothing factor (α) used by decision maker in a behavioral experiment. For completeness, we will also estimate the values of trend (β), the weight of the most recent forecast (θ_L) and trend memory (θ_T) in section 4.4.

4.3.3 Dual Process Theory: Cognitive Reflection and Individual Differences

While system neglect has been suggested as a theoretical explanation for sub-optimal performance (Massey and Wu 2005, Kremer *et al.* 2010), prior results have been focused on the average performance of the sample, looking at forecasting heuristics and the average behavior of the respondents. System neglect appears to explain a portion of observed behavioral deviations from a normative response. However, this suggests a further fundamental question: Do all individuals perform equally as well with respect to these heuristics and can the strength of these responses at the individual level be predicted? We believe that dual process theory may provide an underlying cognitive explanation of forecasting performance. Based on theory from judgment and decision making, we provide an overview of dual process theory (Stanovich and West, 2000). In addition, we highlight the cognitive reflection test (Frederick, 2005) as a measure of the reflective tendency of the individual. We expect that the strength of these forecasting heuristics will vary based on the cognitive tendency of the individual decision maker.

Many researchers in the cognitive sciences have proposed dual process theories of decision making under the broad categories of intuition and analysis. While the literature is extensive and there are differences in some of the aspects of each process, the key finding is that intuition and analysis are two different cognitive processes. These are generically titled "System 1" and "System 2" by Stanovich and West (2000). System 1 processes are typically automatic, leading to a quick answer based on tacit inputs. These reactions are typically descriptive and based on experience, activating intuitive and common sense approaches to a problem. System 2 processes are focused on normative solutions to explicit problems, often analytical and theoretical in nature, and generally take more time and targeted effort to arrive at an answer. Asking someone the sum of eight and ten is a relatively simple System 1 process, while asking that same individual to calculate the cube root of eighty to three decimals would require more or different cognitive processes (or access to a calculator and use of the appropriate problem solving steps to arrive at the answer). System 1 processes quickly provide answers to a problem, while System 2 processes use a structured approach to arrive at an answer. Cognitive reflection is the tendency of an individual to let a System 2 process endorse, monitor or correct a System 1 initial response (Kahneman and Frederick, 2002). Dual process theory also appears to be directly related to forecasting, as quantitative forecasts follow a specific, coherent procedure. Much of the warnings against including qualitative judgment in forecasting are based on a lack of structure surrounding the interaction of judgment and a statistical model (Bunn and Wright, 1991).

Frederick (2005) has suggested that the cognitive reflection tendency of the decision maker has an impact on the decisions made by that individual. Based in dual process theory, Frederick proposed the Cognitive Reflection Test (CRT) as one measure of the cognitive tendency of the individual. In this instrument, each individual is given three questions where the obvious, intuitive answer is incorrect. Consistent with cognitive reflection, the correct answer is found only if the individual uses a System 2 process to override the System 1 answer that springs to mind for most individuals. The CRT is given in appendix 4.1.

We would expect that individuals who can better utilize System 2 processes for decision making are better forecasters. Therefore, we expect that individuals with high cognitive reflection perform better in forecasting and have lower forecast error. This may be caused by two factors: First, decision makers with high cognitive reflection suffer less from system neglect and may exhibit behavior where their implied α is closer to the normative benchmark. The system neglect hypothesis appeals to intuitive decision making, reasoning that individuals have a tendency to neglect the weight of a signal and emphasize the strength of the signal, since the signal itself is emphasized and immediately visible. Arguably, individuals with a stronger 'cognitive brake system' (higher cognitive reflection) are less influenced by an intuition based on a perception of the most recent data point, and therefore better able to properly evaluate the weight they should attribute to a signal. Second, individuals with high cognitive reflection are likely to be more consistent in their forecasts, reducing the noise inherent in their decisions. Further, we expect that as individuals move away from intuition to more reasoning in their decision making, they are better able to focus their mental efforts into analyzing the available data (Kahneman 2003). These mental efforts will lead to a reduction of errors (Dickhaut, Rustichini and Smith 2009), which in our context implies a lower trembling hands variance $\sigma(\varepsilon)$ in the decision making process. We hypothesize:

> HYPOTHESIS 1 (COGNITIVE REFLECTION AND PERFORMANCE): *Subjects with high cognitive reflection have higher performance (lower forecast error) than those with low cognitive reflection.*

HYPOTHESIS 2 (COGNITIVE REFLECTION AND SYSTEM NEGLECT): *Subjects with high cognitive reflection have an α closer to the normative value than subjects with low cognitive reflection.*

4.3.4 Decision Speed and Performance

An additional question of interest is how much time individuals take to reach a decision and how that relates to performance. The time to make a response is usually seen as dependent on the difficulty of the task, with more difficult tasks taking more time (Rustichini 2008). This would imply that the variation in forecast time is mostly due to the different conditions used in the experiment. Taken at face value, one would expect that a longer, more deliberative process would be better than a fast decision. However, in a complex task where all the attributes were known with certainty, individuals using conscious thought and spent more time focused on the attributes performed poorer than those who were forced to make a rapid decision (Dijksterhuis *et al.* 2006). In the time series forecasting task, longer deliberation may allow individual decision makers to emphasize or include perceived changes in level and trend even if those changes are illusory and detrimental to performance. One can make the argument that intuitive responses are generally faster (Kahnemann 2003), since they do not require additional deliberations and reasoning. However, this is confounded by a system neglect argument: If individuals understand the system, they place less weight on the most recent data point and more weight on the underlying system which generated the data. If an individual is making a rapid response based on a sound understanding of the system, this should improve performance but if the individual is responding based on an inappropriately high weight on the most recent data point, it should be detrimental to performance.

This discussion highlights that the relationship between decision time and forecasting performance may be non-linear. While fast decisions may be driven by intuition and contain too little deliberation, slow decisions may contain too much deliberation or inclusion of spurious factors. Final choices under either extreme may be sub-optimal, suggesting that a potential "sweet spot" exists for the time one should spend on a decision. Hence, we investigate two hypotheses surrounding decision speed:

HYPOTHESIS 3 (DECISION SPEED AND PERFORMANCE): *For too low or too high decision times, forecasting performance is lower than an optimal time given a particular context.*

HYPOTHESIS 4 (COGNITIVE REFLECTION AND DECISION SPEED): *Individuals with high cognitive reflection will have a decision speed closer to the estimated optimal time.*

4.3.5 Other Individual Differences

In addition measuring the above data, we selected a number of other individual attributes to include as control variables that might explain at least a portion of forecasting performance. We wanted to control for education and experience, where some individuals may have relevant skills or backgrounds. We selected area of study (college major) as an important individual attribute, as we wanted to determine if this predicted performance. As our subject pool was affiliated with the business school, we wanted to determine if having completed a course in operations or supply chain improved performance, as these courses typically contain at least an overview of forecasting methods. We also controlled for highest degree attained, in part to look for individuals who may be completing an advanced degree. We elected to control for age and years of work experience, as it is possible that individuals with more experience might perform better. We also elected to include gender. This follows Frederick (2005) who found that a group difference by cognitive reflection and gender. In particular, we were interested to see if $\sigma(\varepsilon)$ varied by gender.

With any list of individual attributes, there clearly are other individual factors that we did not choose. We did not have an independent measure of statistical ability, in part because this would have added significant length to the study and seemed to correlate with other attributes such as major. We also did not choose IQ. While Frederick (2005) found that cognitive reflection was correlated with several measures of academic achievement (such as the SAT and ACT), these measures were not correlated with individual attributes such as risk preferences and time preferences that are more strongly correlated with cognitive reflection.

4.4 Experiment

4.4.1 Experimental Design

This experimental research has subjects make repeated forecasts based on a time series of demand data. The subjects observe the most recent period of demand, and have access to all past history of demand in both graphic and tabular forms. They also have access to their previous forecasts, absolute and relative forecast errors. The demand stream is generated from a random walk with noise. The experimental conditions are based on the degree of change *c* (none, moderate, high) and the degree of noise *n* (low, high), resulting in six demand environments. Each environment has 30 historic data points and 50 new data points drawn from four datasets within each environment. This results in a 6*4=24 between subject design. An example time series from each demand environment is shown in appendix 4.2. The forecasting task was implemented in zTree (Fischbacher 2007). Subjects were compensated with a participation fee of $5, plus compensation directly tied to their individual accuracy (1-MAPE) across the 50 periods of forecasting decisions. Most subjects earned $14-15 in total and no other compensation was offered.

4.4.2 Data

The study was conducted in a controlled laboratory environment at a large, public university in the USA. The 252 participants were part of the subject pool affiliated with the business school, and had the option to sign up in response to an online posting. The population was about half undergraduate students, with the remainder being graduate students or university support staff. Each of the 24 conditions had approximately the same number of participants. Before completing the analysis, all forecasts were examined looking for large outliers based on absolute forecast error. In a handful of cases, obvious typographical errors were corrected, while if the intent could not be determined, the forecast was deleted. Such instances represented less than 0.1% of all forecasts.

4.4.3 Analysis

Equations (1)-(3) provide a basis for a model of decision making in time series forecasting. Combining the equations, we can get an estimate of the forecast as follows, where Δ symbolizes first differences, i.e. $\Delta D_t = D_t - D_{t-1}$:

$$\begin{aligned} F_{t+1} &= \alpha E_t + \theta_L F_t + \alpha\beta\Delta D_t + \alpha\beta\left(\theta_T - \beta\right)\Delta D_{t-1} \\ &\quad + \beta\left(\theta_L - \alpha\right)\Delta F_t + \beta\left(\theta_L - \alpha\right)\left(\theta_T - \beta\right)\Delta F_{t-1} \\ &\quad + \left(\theta_T - \beta\right)^2 T_{t-2} + \left(1 - \theta_L\right) C + \varepsilon_t \end{aligned} \tag{5}$$

This can be estimated by from the data by fitting

$$F_{t+1} = a_1 E_t + a_2 F_t + a_3 \Delta D_t + a_4 \Delta D_{t-1} + a_5 \Delta F_t + a_6 \Delta F_{t-1} + \text{constant} \tag{6}$$

This specification is not exact, because we do not have independent prior estimates of trend T_{t-2}, and it is possible that prior trend estimates change across the 50 periods of data from each participant. However, as it is likely that $(\theta_T - \beta)^2$ is a small number, this potential error is unlikely to have a large influence on the variables of interest. In addition, adding ΔD_{t-2} and ΔF_{t-2} into the model moves the problem back one period: The omitted variable becomes T_{t-3} with a coefficient of $(1-\beta)\ \theta_T\ (\theta_T - \beta)^2$. Since β and θ_T should both be ≤ 1, this coefficient is lower in absolute value than $(\theta_T - \beta)^2$. We tested these terms using pre-test data and found them to be non-significant. Hence, for parsimony, we focus on the parameters of interest in (6), where $\alpha = a_1$, $\beta = a_3/a_1$, $\theta_L = a_2$, $\theta_T = a_3/a_1 + a_4/a_3$ and $C = \text{constant}/(1 - a_2)$.

This is a multilevel model, where each observation in time t is nested within in subject i, who is in turn nested in demand dataset s and again nested in experimental condition (i.e., demand environment) c. Thus, the error terms are also nested, and we use random coefficients to estimate a_1 and a_3 while using fixed coefficients for the remaining terms as there was very little variance in those terms. Note that we do estimate trembling hands error variance ε_t separately for each condition.

4.4.4 Hypotheses Tests

As a first test to determine what causes the variance in our forecasting error, we estimate a nested random effects model (with subjects nested in datasets nested in conditions) with heteroskedasticity (where error variances are allowed to change according to condition) to predict the absolute forecast error. Results show significant variation at the condition level ($\sigma = 18.22$, $p \leq 0.01$), the dataset level ($\sigma = 5.68$, $p \leq 0.01$) and the individual level ($\sigma = 0.99$, $p \leq 0.01$). While it is expected that forecasting performance is to a large degree determined by the actual condition the subject is in (i.e. the difficulty of the forecast), as well as by the actual dataset s/he is seeing, the significant variance at the individual level provides evidence that individual characteristics can have a large impact on forecasting performance.

The first individual characteristic of interest is cognitive reflection, as measured by the CRT score (Frederick, 2005). Following Frederick (2005), we coded each correct answer as '1' and then formed a sum over all three items to calculate the CRT score. To test whether the items are reliable, we estimated a CFA model on these three dichotomous items (using polychoric correlations) in MPlus 5. The results from the analysis show that all three items load significantly on one construct. While reliability of the second item lower, the other two reliabilities are fairly high, leaving the average variance extracted of the scale at around 55%. Cronbach's $\alpha = 0.61$; while this is fairly low it is above the minimum proposed for new scales (Nunally 1978).

To further test whether individual CRT scores can predict forecasting performance, we include CRT scores and additional individual level demographic variables in the analysis. Area of study was coded into 7 different categories, with liberal arts being the omitted category. 'No Operations Course Taken' is a dummy variable that captures whether subjects have ever taken a class in operations or supply chain management. Age is the subjects' reported age, work experience is a dummy variable capturing whether the subject has had some prior work experience (1 = prior work experience). Education measured the subject's education using 5 categories (No High School, High School, Some College/2 yr. degree, 4 yr. degree, Masters/PhD degree) and was included as a continuous variable. Gender is a dummy variable that captures the subjects' gender (1 =

female). The model allowed heteroskedasticity and estimated a different error variance for each condition. Table 4.1 summarizes our results.

Table 4.1: Individual Differences in Absolute Forecast Errors

	Estimate	Standard Error
CRT Score	-0.58**	(0.16)
Area of Study		
Business: Finance/Accounting	-0.18	(0.56)
Business: Other	-0.37	(0.59)
Education, Psychology or Social Sciences	0.39	(0.52)
Engineering or Physical Sciences	0.07	(0.61)
Law, Dentistry or Medicine	1.24*	(0.62)
Other	0.53	(0.51)
No Operations Course Taken	-0.50	(0.47)
Age	-0.02	(0.02)
Work Experience	0.37	(0.37)
Education	-0.29	(0.21)
Gender	-1.19**	(0.32)
Constant	40.17**	(7.62)
σ_c (random intercept, condition)	18.18**	(5.39)
σ_s (random intercept, dataset)	5.75**	(1.05)
σ_i (random intercept, individual)	0.00	(0.00)
N	12,548	

Notes: ** $p \leq 0.01$, * $p \leq 0.05$. Heteroskedasticity is significant ($p \leq 0.01$) but error variance difference are not reported.

Individuals with a higher CRT score have a lower absolute forecast error on average, suggesting that these individuals are better forecasters. Two additional control variables are significant: Law, dentistry or medicine students are worse at forecasting than liberal arts students, while women are better forecasters than men. Together, these variables seem to explain existing variance in forecasting performance at the individual level, since when these variables are included in the estimation, the variance of the random intercept at the individual level reduces to 0. Hypothesis 1 is supported.

Next, we examine whether we can track these performance improvements to changes in our behavioral model for differences in CRT scores. To do so, we re-estimate Eq. (6) in each condition separately, allowing the effects of E_t, $F_{t\text{-}1}$, ΔD_t, $\Delta D_{t\text{-}1}$, ΔF_t, $\Delta F_{t\text{-}1}$ as well as the intercept to change in the CRT score. In particular, we compare the results of low CRT (0 correct on the instrument) with high CRT (2 or 3 correct on the instrument), allowing the variance of the error term to change for high CRT scores (≥2). In Table 4.2, we show the parameters for low and high CRT scores. Note that different parameters for low and high CRT were only entered if an underlying structural estimate showed significant ($p \leq 0.05$) changes in the CRT score. In addition, it is appropriate to focus primarily on the observed α values and $\sigma(\varepsilon)$, though we provide all parameters for completeness.

Moving specifically to the system neglect hypothesis, we compare the observed α compared to α^* in Table 4.2: In low W (c^2/n^2, i.e. conditions 1 & 2), subjects over-react, while in high W (condition 5), subjects under-react. In general, this shows that individuals react more strongly to increases in change than in noise, but this effect tends to level off as the amount of change increases. However, what is particularly interesting is that under high noise (conditions 2, 4, and 6) the observed α is lower for the high CRT group. This provides support for H2, and suggests that individuals with high cognitive reflection appear to be less responsive to noise when making forecasts. Additionally, for individuals high in cognitive reflection, the α tends to level off more slowly. In conditions of high noise, high CRT individuals also exhibit less reaction to change. In the language of exponential smoothing, for any change to noise ratio, individuals with high cognitive reflection place the same or lower weight on the most recent data while increasing the weight on earlier points.

Table 4.2: Individual Differences in Forecasting Behavior

	Condition 1		Condition 2		Condition 3		Condition 4		Condition 5		Condition 6	
(c, n)	0,10		0,40		10,10		10,40		40,10		40,40	
CRT	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**
α*	0.00	0.00	0.00	0.00	0.62	0.62	0.22	0.22	0.94	0.94	0.62	0.62
α	0.38	0.38	0.53	0.35	0.77	0.59	0.67	0.54	0.72	0.72	0.71	0.46
β	0.44	0.14	-0.18	0.44	0.18	0.23	0.01	0.01	0.49	0.49	0.25	0.68
θ_L	0.67	0.67	0.68	0.87	1.00	1.00	0.90	0.90	1.00	1.00	0.94	0.94
θ_T	0.43	0.11	*n/a*	0.65	0.10	0.16	*n/a*	*n/a*	0.88	0.65	0.46	0.80
$\sigma(\varepsilon)$	5.95	5.95	19.32	21.24	11.19	6.97	29.29	21.97	20.76	17.36	34.93	25.95

Notes. Different values within a condition for the low CRT (0) and high CRT (2 or 3) columns were only entered if the parameters were different at $p \leq 0.05$. For $\sigma(\varepsilon)$, LR tests for homoskedasticity were used.

Third, we compare the trembling hands noise $\sigma(\varepsilon)$ and find that where any change c is present (conditions 3-6), the noise is much lower for subjects with a high CRT scores in conditions 3-6. This also supports H1. In summary, this analysis points to the conclusion that cognitive reflection (as measured by the CRT) may lower biases by lowering overreaction due to system neglect. The most persistent performance improvements of individuals with high CRT scores appears to result from less emphasis on the most recent data and a more consistent way of forecasting, (i.e. lower trembling hands error), for subjects with high CRT scores.

4.4.5 The Time to Make a Decision

This subsection discusses some exploratory insights related to the time it takes subjects to make a decision. On average, subjects took 11.6 seconds to make a forecast, but there was considerable variance in this variable ($\sigma = 8.02$). To test these ideas, we estimated a random effects model predicting response time using CRT scores, the prior demographic variables, as well as nested random effects at the condition, dataset and individual level. The results from this analysis are reported in Table 4.3.

Table 4.3: Individual Differences in the Time to Respond

	Estimate	Standard Error
CRT Score	-0.82**	(0.25)
Area of Study		
Business: Finance/Accounting	0.77	(0.43)
Business: Other	-0.89	(0.37)
Education, Psychology or Social Sciences	0.37	(0.69)
Engineering or Physical Sciences	0.34	(0.71)
Law, Dentistry or Medicine	0.42	(0.66)
Other	0.49	(0.59)
No Operations Course Taken	1.28	(0.80)
Age	0.09**	(0.04)
Work Experience	0.37	(0.58)
Education	-0.32	(0.34)
Gender	-0.77	(0.54)
Constant	9.18**	(2.13)
σ_c (random intercept, condition)	0.78*	(0.36)
σ_s (random intercept, dataset)	0.00	(0.00)
σ_i (random intercept, individual)	3.86**	(0.19)
N	12,550	

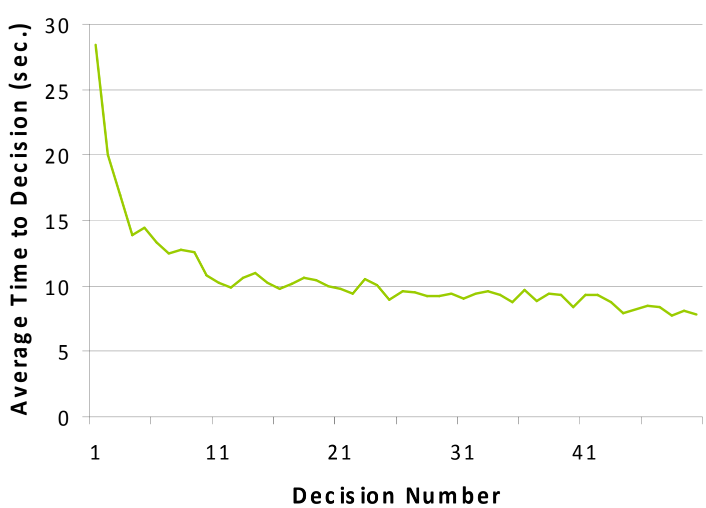
Notes. ** $p \leq 0.01$, * $p \leq 0.05$.

There are random effects at the condition level (σ = .78) which not surprisingly indicates that more difficult tasks take longer. Similarly, with the significance effects of the CRT scores on time to reach a decision (-0.82), there appears to be an effect on response time. However, much more variance is at the individual level (σ = 3.86), even after controlling for CRT scores and other demographics. In other words, while the difficulty of the forecasting task explains a small amount of the variance in the time it takes to make a forecast, far more variance exists at the individual level, indicating that some people are simply much faster at forecasting than others.

4.4.6 Decision Speed and Performance

Next, we looked to evaluate the impact of the time to make a forecast on performance, measured as absolute forecast error (AFE). Several factors may influence performance beyond condition-specific effects. First, individuals appear to learn with experience in the forecasting task and take less time as they make additional forecasts (Figure 4.3). The degree of learning can be estimated by a traditional learning curve. Second, as described above, there is variance in performance by individual. Third, if individuals experience a larger forecast error (as in more difficult forecasting tasks), this likely to increase the decision time.

Figure 4.3 Learning and Experience in Forecasting Task



To bring these factors together, control for learning effects and allow for comparisons across different seeds and conditions, will compare performance relative to a "predicted optimal time" for a particular decision after allowing for random intercepts. The simplest model is a learning curve model as shown in (7):

$$\ln(\text{decision time}) = a_0 + a_1^i \ln(\text{decision no.}) + a_2 CRT \ln(\text{decision no.}) \quad (7)$$
$$+ a_3 \ln|E_t| + \text{random effects} + \varepsilon_t$$

This estimates a learning curve, where the time to reach a decision in period is a function of a constant, experience with the forecasting task (decision number), individual cognitive reflection (≥ 2) and experience, absolute forecast error in a period and random effects by condition, demand seed and individual. Again, we estimate these with a random effects model, as shown in Table 4.4.

Table 4.4: Learning Curve and Parameter Estimation

Variable	Estimate	Standard Error
ln(decision no.)	-0.23**	0.01
CRT ln(decision no.)	-0.03†	0.02
$\ln\|E_t\|$	0.02**	<0.01
constant	2.91**	0.03
σ_i (ln(decision no.))	0.11**	0.01
σ_c (random intercept, condition)	0.03	0.06
σ_s (random intercept, dataset)	0.05	0.07
σ_i (random intercept, individual)	0.36**	0.02

Notes. ** $p \leq 0.01$, † $p \leq 0.1$.

Consistent with expectation, learning occurs with experience in the simulation such that individuals forecast more quickly in increasing periods. Additionally, there is small though statistically significant correlation showing that individuals who score higher in CRT (≥2) answer more quickly on average, and that greater error in forecasts cause individuals to take longer on their next forecast.

Taken together, these results allow us to compare performance in absolute terms, by comparing absolute forecast error and distance from the "predicted optimal time" after including learning curve effects and random intercepts by seed, condition and individual. For our particular demand realizations, there appears to be a right amount of time for each particular seed and condition, a function of the conditions and learning effect, as simpler conditions with less change and noise require less analysis time than more complex and noisy conditions and individuals appear to take less time with experience in making forecasts. We compare the deviation between actual times and the predicted optimal time (Δ_t = actual – predicted time) and compare performance relative to absolute deviation as shown in (8).

$$\left|E_t\right| = a_0 + a_1\Delta_t + a_2\Delta_t^2 + a_3\text{CRT} + a_4\ln(\text{decision no.}) + \text{control vars} + \text{random effects} + \varepsilon_t \quad (8)$$

We also include a learning effect on error performance. The effect of learning in (7) was focused on learning relative to increased speed of decision making, while in this case we focus on learning relative to the accuracy of decisions. Again, we again estimate a random intercepts model by condition, demand seed and individual, and include the same control factors as before. The results are shown in Table 4.5.

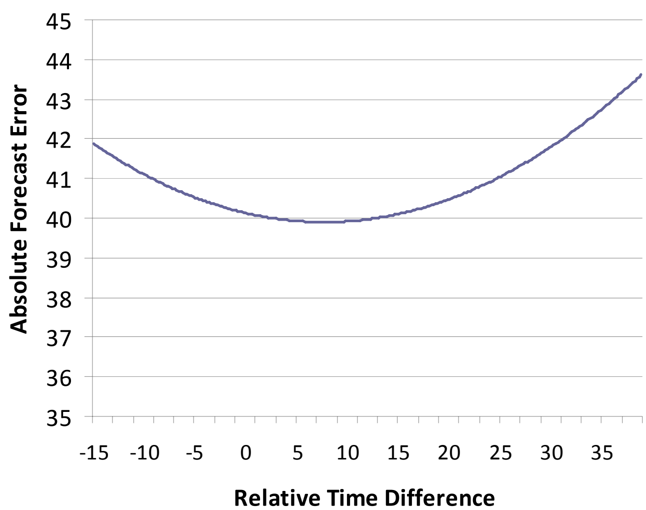
Table 4.5: Individual Differences in Performance (Absolute Forecast Error)

Variable	Estimate	Standard Error
Δ_t	-0.059	(0.039)
Δ_t^2	0.004*	(0.002)
CRT Score	-0.550**	(0.168)
ln (decision no.)	0.094	(0.204)
Area of Study		
Business: Finance/Accounting	-0.032	(0.574)
Business: Other	-0.073	(0.609)
Education, Psychology or Social Sciences	0.609	(0.536)
Engineering or Physical Sciences	0.068	(0.623)
Law, Dentistry or Medicine	-1.325*	(0.644
Other	0.381	(0.522)
No Operations Course Taken	-0.542	(0.478)
Age	-0.025	(0.022)
Work Experience	0.502	(0.374)
Education	-0.271	(0.218)
Gender	-1.242**	(0.329)
Constant	40.127**	(7.545)
σ_c (random intercept, condition)	17.885*	(0.36)
σ_s (random intercept, dataset)	5.946**	(0.00)
σ_i (random intercept, individual)	0.000	(0.19)
N	11,876	

Notes. ** $p \leq 0.01$, *$p \leq 0.05$. *N* is reduced after eliminating any excessively long decision times (≥40 sec, more than 3.5 deviations from the mean time)

There are a number of observations from Table 4.5. First, there is a quadratic effect of decision time (Δ_t^2) and performance, suggesting that a large deviation from the optimal predicted time from decision is detrimental to performance. Decisions that are either too long or too short show an increase in absolute forecast error (See Figure 4.3 below). Second, as before, individuals with higher CRT scores are more likely to have a lower forecast error. Third, no additional learning appears to improve performance. Fourth, some of the demographic variables are significant, as women tend to have higher performance while individuals majoring in law, dentistry or medicine tend to have poorer performance. Summarizing the results across conditions and data seeds, we provide a visual representation of absolute forecast error in Figure 4.3:

Figure 4.4 Curvilinear Results of Response Time and Performance



This figure illustrates the results described in Table 4.5, showing the curvilinear effect of time difference on performance. While the performance (absolute forecast error) is relatively flat in the range of 0-10 seconds of deviation from the predicted optimal time, performance declines if individuals answer too quickly (a negative relative time difference) or take too long (a positive relative time difference). This supports H3.

Lastly, we split out the results of time difference conditioned on whether subjects had a relative time difference greater than or less than zero relative to the predicted optimal time. We would expect that across 50 forecasts, individual forecast data points may sometimes take longer than the predicted optimal time, and sometimes shorter than the predicted optimal time. We do this to determine how individual performance in cognitive reflection interacts with the optimal time. The results are shown in Table 4.6[1]:

Table 4.6: Performance Conditional on Relative Time Difference

Conditional on $\Delta_t > 0$

	Estimate	Standard Error	p-value
CRT Score	-0.26	0.13	0.04
Constant	4.72	0.18	< 0.01

Conditional on $\Delta_t < 0$

	Estimate	Standard Error	p-value
CRT Score	0.14	0.06	0.01
Constant	-2.68	0.11	< 0.01

The upper panel of Table 4.6 shows that individuals who have a high cognitive reflection score are less likely to take longer than the optimal time (-0.26, p = 0.04). The lower panel of Table 4.6 shows that individuals who have a high cognitive reflection score are less likely to take shorter than the optimal time (0.14, $p \leq 0.01$). Results from this analysis including decision time show that individuals with high cognitive reflection answer closer to the predicted optimal time across the range of time series forecasting environments tested in this analysis. H4 is supported.

[1] Clearly we could have split the sample at other points in the performance curve in figure 4.3 as well, such as < 0 and >10, but this may not add any new information and may be more difficult to interpret.

4.5 Conclusion

This research has found that decision makers follow an error-response model consistent with single exponential smoothing. Individuals with high cognitive reflection (Frederick 2005) demonstrate high performance by having lower forecast errors. Consistent with the system neglect hypothesis (Massey and Wu 2005), individuals with high cognitive reflection demonstrate less system neglect. This was especially true in conditions characterized by high noise. In addition, these individuals also have lower trembling hands error. Not surprisingly, task difficulty influences the time to reach a decision (Rustichini 2008). However, taking more time to reach a decision or answering too quickly is detrimental to performance, at least in time series environments that lack underlying trend or seasonal components. Individuals were quick to include estimates trend where none was present. In addition, individuals with high cognitive reflection performed better relative to the optimal time to generate a forecast and were less likely to either answer too quickly or take too long.

These results are subject to certain limitations: First, our time series had no underlying trend or seasonal component. This allowed for precise mathematical comparisons with a normative standard, but is not necessarily representative of all real world time series. However, methods to de-trend or de-seasonalize historical data are available, so these results are still valid in such environments. Our demand data was not serially correlated beyond the random walk (with precisely controlled noise as well). However, actual customer demand may be "lumpy" and is often serially correlated due to short-term changes in customer tastes, preferences or macro considerations that may influence demand. While additional research could include such factors, this may lack context supporting the inclusion of potentially realistic data. Third, we used a somewhat limited set of measures of individual difference. Our choice was guided by what we could practicably obtain and which had a theoretical relationship to forecasting, though there might be other individual-specific factors (such as IQ, personality type, mathematical ability, etc.) which would be more effective. This could be the subject of future research. Additionally, many forecasts are made by groups. Our research was focused on individual forecasters, and future research could emphasize the role of group interaction on time series forecasting behavior.

This research has implications in the field of behavioral operations and in actual supply chains. In behavioral operations, a number of studies have had subjects manage inventory or respond to sequential demand in both single and multi-echelon environments. The frequent observation is that individuals make sub-optimal decisions in many of these contexts. While this is a robust finding, it will become increasingly important to understand the difference between system neglect errors in demand forecasting and errors grounded in a lack of understanding of other factors (such as costs, profits and customer service) that may influence an inventory decision.

In particular, the importance system neglect hypothesis cannot be overstated: Individuals have a tendency to place too much weight on a recent signal and neglect the system that generated the signal. However, the tendency of system neglect in time series forecasting is not the same for all individuals. Cognitive reflection is related to system neglect, and the CRT test, and this attribute is predictive of forecasting performance. High cognitive reflection lowers individual tendency for randomness in decision making. From a personnel selection perspective, this suggests that crucial forecasting tasks for the organization should be performed by forecasters with a high cognitive reflection. From a system design perspective, the most recent data point is likely to be very salient for participants, so placing more emphasis on the system are likely to improve performance. Lastly, the time to reach a decision appears to be a driving factor of performance. If decision makers choose too quickly, they perform poorer, while taking a large amount of extra time also did not improve performance. Cognitive reflection is related to the time to reach a decision, and this impacts performance.

Chapter 5

Conclusion

This research is a study of inventory decision making focused in two key areas. First, past research has lacked underlying theory that would explain the observed behavioral deviations from normative, profit maximizing inventory decisions. Second, the results are focused on individual-specific attributes where possible. More specifically, the results point to cognitive reflection in high and medium margin predicting performance, while cognitive dissonance specific to customer service appears to drive behavior in low margin settings. These results in inventory ordering decisions also are impacted by forecasting performance. The explanations for these observed behaviors are based on theory from psychology and judgment and decision making and have been experimentally tested with both practitioners and individuals affiliated with the business school subject pool.

In essay one (chapter two), cognitive reflection is shown to correlate with performance in a high-margin setting. Individuals with higher cognitive reflection tended to have higher expected profit, order closer to the optimal quantity and had lower order quantity variance. Consistent with those performance indicators, individuals with high cognitive reflection exhibited lower tendency to anchor on the mean, less propensity to chase demand and appeared to exhibit less preference to minimize *ex post* inventory error. These results also held when we considered other explanations of observed behavior, including major, years experience and managerial position. However, these same results did not appear to be the case in low-margin settings.

Essay two (chapter three) contrasts the results based on margin settings and cognitive dissonance. In essay two, the results show that individual-level cognitive reflection predicts performance only in medium and high margin settings. However, cognitive dissonance specific to customer service expectations appears to be strong in low margin settings but not in other margin conditions. This suggests that individuals are reluctant to plan to disappoint some customers, even when it would maximize their expected profits to do so. Following cognitive dissonance theory, individuals should seek to reduce this dissonance by ordering closer to the mean demand in low margin settings, which appears to explain at least a portion of the "pull-to-center" effect that has been noted in prior research such as Bostian *et al.* 2008. Consistent with dissonance theory, individuals appear to trivialize the importance of satisfying customers in the low-margin setting, but do not exhibit a corresponding increase in the importance of satisfying customers in a high margin setting.

While these results hold in a newsvendor inventory decision task, some of the same behavioral factors are active in the forecasting task as discussed in essay three (chapter four). In many cases, a forecasting task can be considered an antecedent to the inventory ordering task. In fact Schweitzer and Cachon (2000) highlight the importance of managerial judgment in forecasting while suggesting that the inventory decision can be automated. Essay three is consistent with prior research which shows that individuals exhibit system neglect by over-reacting to demand signals when demand is more stable, while individuals under-react when demand is less stable. However, individual cognitive reflection mitigates a portion of this error: Individuals with high cognitive reflection make forecasts closer to the optimal value, especially in environments where there is a higher amount of noise. Additionally, individuals with high cognitive reflection take closer to the "right" amount of time across the range of time series forecasting environments outlined in this study.

These results have several implications for academics and practitioners. First, individual-specific cognitive reflection predicts performance in both forecasting and inventory decisions. This may impact hiring decisions for supply chain positions. While other factors may influence performance, if all else is equal, the clear recommendation would be to hire individuals that exhibit high cognitive reflection. Second, individuals are influenced by their perceptions of the importance of customer service, particularly in low-margin settings. This may cause individuals to be reluctant to plan to disappoint customers in a low margin setting, even when that would maximize expected profits for the firm. Managers should consider this factor in how they structure reward systems and recognition of performance. Third, there may be an impact of how information is presented to the decision maker: Simply presenting this information as a moving average of demand rather than including the most recent demand may change performance relative to order quantities.

There are several opportunities for future research. One future research idea involves studying supply chain behavior under other demand distributions. Most prior behavioral studies of supply chain behavior sample demand from artificial distributions: The uniform distribution is found frequently in literature, in part because it does not require any special information to calculate a percentage of the distribution. However, uniformly distributed demand is rare in practice, and most actual supply chain demand has some sort of central tendency. Hence, this dissertation has focused on normally-distributed data for the experiments, a recommendation from Su (2008) and others. This could be extended with other demand distributions.

Beyond the distribution of demand, in many actual supply chain contexts, there is autocorrelation of demand between periods, as well as underlying trends. Individuals see this sort of response all the time in nature: Objects in motion tend to move in that same direction unless acted upon by another force. In a supply chain context, high demand for a product in one period is often associated with high demand in upcoming period(s). However, most supply chain literature focuses on independent random draws from a distribution, rather than allowing for autocorrelation. This issue appears to be worth further investigation, because apparent behavioral anomalies under the assumption of a random draw might prove to be appropriate heuristics if demand is correlated, or if there is an underlying trend or seasonal pattern.

Additionally, there are other research approaches that could be employed to investigate inventory and forecasting decisions on a neurocogitive level. It may be possible to use fMRI studies of brain function to investigate neural activity, and such activity may provide additional insight into observed behavior. Such studies have provided insight into economic decision making and moral reasoning, though as yet published research does not appear to include forecasting or inventory ordering decisions more common in supply chain settings. In particular, it may be interesting to explore the difference between how an individual is compensated or seeks to mitigate cognitive dissonance when compared to decisions that would maximize expected profits for the firm. Given the impact of cognitive reflection and cognitive dissonance, experiments in differing margin contexts may encourage activity in different parts of the brain.

An additional area for future research is to further explore the relationship between the forecasting task and the inventory ordering task under different margin conditions. It is possible that forecasts will be biased as a way of reducing cognitive dissonance, especially in low margin settings. If the normative forecast suggests using the mean of the demand distribution as the best estimate of future demand, individuals who also must place an order for low-margin inventory may display a downward-biased forecast. Though clearly illusory, this type of behavior might be one way of adding a consonant condition to support a lower order quantity decision. In addition, there may be opportunities to improve forecasts and inventory quantity decisions by controlling the time devoted to the forecasting task, either by preventing individuals from answering too quickly or taking too long.

These future research ideas tend to focus on further explaining behaviors and conditions which cause those behaviors. Another area for future research is to develop specific training that would mitigate non-optimal behavior. The research presented in this dissertation applies theory to investigate behavior. These theories may influence the type of training that may work best, though the efficacy of specific methods to mitigate behavior remains an open question.

Other than inclusion of risk aversion, the performance implications of individual behavior and cognitive dissonance have yet to be applied in more complex models of supply chain behavior. In a simple supply chain, a manufacturer may have a low margin, a distributor a medium margin and the retailer may have a high margin. The built-in biases that each individual decision maker may exhibit may impact the overall performance, and perhaps these may be considered when modeling behavior across the supply chain. Performance may especially be impacted by cognitive dissonance regardless of the specific margin conditions.

While there may be other important areas for future research, hopefully this dissertation has started to answer some questions about why individuals behave as they do when faced with the supply chain tasks of ordering inventory and forecasting demand.

The results show that individuals are heterogeneous with respect to the decisions they make. Based on the experimental results, it appears as though cognitive reflection, cognitive dissonance and system neglect may provide important guidance in explaining the behavior of supply chain decision makers.

References

Andreassen, P. B. and S. J. Kraus. 1990. Judgmental Extrapolation and the Salience of Change. *Journal of Forecasting* **9**(4), 347-372.

Aronson, E. 1997. Review: Back to the Future: Retrospective Review of Leon Festinger's "A Theory of Cognitive Dissonance". *American Journal of Psychology.* **110**(1), 127-137.

Baron, R.M. and D.A. Kenney. 1986. The Moderator-Mediator Variable in Social Psychological Research: Conceptual, Strategic and Statistical Considerations. *Journal of Personality and Social Psychology.* **51**(6), 1173-1182.

Barry, D. M. and G. F. Pitz. 1979. Detection of Change in Nonstationary, Random Sequences. *Organizational Behavior and Human Performance* **24**, 111-125.

Bendoly, E., K. Donohue and K. L. Schultz. 2006. Behavior in Operations Management: Assessing Recent Findings and Revisiting Old Assumptions. *Journal of Operations Management* **24**(6), 737–752.

Berkowitz, L. and E. Donnerstein. 1982. External Validity is More Than Skin Deep. *American Psychologist.* **37**(3), 245-257.

Bloomfield, R.J., F. Gino and S.M. Kulp. 2007. Behavioral Causes of the Bullwhip Effect in Single Supply Chains, Working Paper, Cornell University.

Bolger, F. and N. Harvey. 1993. Context-Sensitive Heuristics in Statistical Reasoning. *Quarterly Journal of Experimental Psychology* **46**A(4), 779-811.

Bolton, G.E. and E. Katok. 2008. Learning by Doing in the Newsvendor Problem: A Laboratory Investigation of the Role of Experience and Feedback, *Manufacturing and Service Operations Management*. **10**(3), 519-538.

Bolton, G.E., A. Ockenfels and U. Thonemann. 2008. Managers and Students as Newsvendors: How Out-of-Task Experience Matters. University of Cologne: Working Paper Series in Economics No. 39.

Bostian, AJ A., C.A. Holt and A.M. Smith. 2008. Newsvendor "Pull-to-Center" Effect: Adaptive Learning in a Laboratory Experiment. *Manufacturing and Service Operations Management.* **10**(4), 590-608.

Brehm, J. 1956. Postdecision Changes in the Desirability of Alternatives. *Journal of Abnormal and Social Psychology.* **52**, 384-389.

Cacioppo, J.T. and R.E. Petty. 1982. The Need for Cognition. *Journal of Personality and Social Psychology.* **41**(1), 116-131.

Carbone, R., W. Gorr. 1985. Accuracy of Judgmental Forecasting of Time Series. *Decision Sciences.* **16,** 153-160.

Clark, A. 1998. Being There: Putting Brain, Body and World Together Again. Cambridge, MA: The MIT Press.

Cooper, J. 2007. *Cognitive Dissonance: Fifty Years of a Classic Theory.* Los Angeles: SAGE Publications.

Cooper, J., and R.H. Fazio. 1984. A New Look at Dissonance Theory. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, **16**, 199-206.

Croson R. and K. Donohue. 2006. Behavioral Causes of the Bullwhip Effect and the Observed Value of Inventory Information. *Management Science.* **52**(3), 323-336.

Croson, R., K. Donohue, E. Katok, J. Sterman. 2009. Order Stability in Supply Chains: Coordination Risk and the Role of Coordination Stock. Working paper.

Dickhaut, J., Rustichini, A. and Smith, A. A Neuroeconomic Theory of the Decision Process. *Proceedings of the National Academy of Sciences,* 106(52), December, 2009.

Dijksterhuis, A. Bos, M. W. Nordgren and L. F. van Baaren, R. B. 2006. On Making the Right Choice: The Deliberation-Without-Attention Effect. *Science.* **311**(5763), 1005-1007

Doerr, K.H., T. Freed, T.R. Mitchell, C.A. Schriesheim, X. Zhou. 2004. Work Flow Policy and Within-Worker and Between-Workers Variability in Performance. *Journal of Applied Psychology*. **89**(5), 911-921.

Edgeworth, F.Y. 1888. The Mathematical Theory of Banking. *Journal of the Royal Statistical Society*. **51**(1), 113-127.

Epley, N., T. Gilovich. 2001. Putting adjustment back in the anchoring and adjustment heuristic. *Psychological Science*. **12**(5), 391-396.

Epstein, S. 1994. Integration of the Cognitive and Psychodynamic Unconscious. *American Psychologist.* **49**(8), 709-724.

Festinger, L. 1957. *A Theory of Cognitive Dissonance.* Evanston, IL: Row-Peterson.

Fildes, R. P. Goodwin, M. Lawrence, K. Nikolopoulos. 2009. Effective Forecasting and Judgmental Adjustments: An Empirical Evaluation and Strategies for Improvement in Supply-Chain Planning. *International Journal of Forecasting.* **25**, 3-23.

Fischbacher, U. 2007. z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics* **10**(2), 171–178.

Frederick, S. 2005. Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*. **19**(4), 25-42.

Gavirneni, S. and Y. Xia. 2009. Anchor Selection and Group Dynamics in Newsvendor Decisions—A Note. forthcoming, *Decision Analysis.*

Gigerenzer, G. 2007. *Gut Feelings: The Intelligence of the Unconscious.* New York: Penguin Books.

Gino, F. and G. Pisano. 2008. Toward a Theory of Behavioral Operations. *Manufacturing Service Operations Management.* **10**(4), 676–691.

Griffin, D. and A. Tversky. 1992. The Weighing of Evidence and the Determinants of Confidence. *Cognitive Psychology.* **24**, 411-435.

Hammond, K.R. 1996. *Human Judgment and Social Policy*. New York: Oxford University Press.

Hutchinson, J.W., W.A. Kamarura and J.G. Lynch, Jr. 2000. Unobserved Heterogeneity as an Alternative Explanation for "Reversal" Effects in Behavioral Research. *Journal of Consumer Research.* **27**, 324-344.

Harrison, P. J. 1967. Exponential Smoothing and Short-Term Sales Forecasting. *Management Science.* **13**(11), 821-842.

Harvey, N. 2007. Use of Heuristics: Insights From Forecasting Research. *Thinking & Reasoning.* **13**(1), 5-24.

Hogarth, R.M. and S. Makridakis. 1981. Forecasting and planning: an evaluation. *Management Science.* **27**(2), 115-138.

Kahneman, D and A. Tversky. 1982. On the Study of Statistical Intuitions. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 493-508). New York: Cambridge University Press.

Kahneman, D. and S. Frederick. 2002. Representativeness Revisited: Attribute Substitution in Intuitive Judgment. In *Heuristics and Biases: The Psychology of Intuitive Judgment,* Thomas Gilovich, Dale W. Griffin, Daniel Kahneman eds. New York: Cambridge University Press.

Kremer, M., B. Moritz and E. Siemsen. 2010. Demand Forecasting Behavior: System Neglect and Change Detection. Working Paper: University of Minnesota.

Kremer, M., S. Minner and L.N. Van Wassenhove. 2010. Do Random Errors Explain Newsvendor Behavior? *Manufacturing and Service Operations Management.* Articles in Advance, April 9, 2010.

Lawrence, M. J., R. H. Edmundson, and M. J. O'Connor. 1985. An Examination of the Accuracy of Judgmental Extrapolation of Time Series. *International Journal of Forecasting.* **1**, 25-35.

Lawrence, M. and M. O'Connor. 1992. Exploring Judgmental Forecasting. *International Journal of Forecasting.* **8**, 15-26.

Lawrence, M. P. Goodwin, M. O'Connor, D. Önkal. 2006. Judgmental Forecasting: A Review of Progress Over the Last 25 Years. *International Journal of Forecasting*, **22**, 493-518.

Lee, H.L., V. Padmanabhan and S. Wang. 1997. Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science.* **43**(4), 546-558.

Lurie, N.H. and J.M. Swaminathan. 2009. Is Timely Information Always Better? The Effect of Feedback Frequency on Decision Making. *Organizational Behavior and Human Decision Processes.* **108**(2), 315-329.

Massey, C., G. Wu. 2005. Detecting Regime Shifts: The Causes of Under- and Overreaction. *Management Science.* **51(**6), 932-947.

McNamara, J. M., A. I. Houston. 1987. Memory and the Efficient Use of Information. *Journal of Theoretical Biology*. **125**, 385-395.

Nunnally, J.C. 1978. Psychometric Theory, 2nd ed., McGraw-Hill, New York.

Moritz, B., A. Hill and K. Donohue. 2009. Cognition and Individual Differences in the Newsvendor Problem: Behavior Under Dual Process Theory. Working Paper: University of Minnesota.

Oshikawa, S. 1972. The Measurement of Cognitive Dissonance: Some Experimental Findings. *Journal of Marketing.* **36**(1), 64-67.

Parlar, M. and Z.K. Weng. 1997. Designing a Firm's Coordinated Manufacturing and Supply Decisions with Short Product Life Cycles. *Management Science*. **43**(10), 1330-1344.

Porteus, E.L. 2002. *Stochastic Inventory Theory*. Stanford University Press, Stanford CA.

Rudi, N. and D. Drake. 2009. Level, Adjustment and Observation Biases in the Newsvendor Model. Working Paper: INSEAD, France.

Sanders, N. 1997. The Impact of Task Properties Feedback on Time Series Judgmental Forecasting Tasks. *Omega.* **25**, 135-144.

Sanders, N. and K. B. Manrodt. 2003. The Efficacy of Using Judgmental Versus Quantitative Forecasting Methods in Practice. *Omega.* **31**, 511-522.

Schweitzer, M.E. and G.P. Cachon. 2000. Decision Bias in the Newsvendor Problem with a Known Demand Distribution: Experimental Evidence. *Management Science*. **46**(3), 404-420.

Silver, E.A., D.F. Pyke and R. Peterson. 1998. *Inventory Management and Production Planning and Scheduling.* John Wiley & Sons Publishing: Hoboken, NJ.

Simon, L., Greenberg, J., and Brehm, J. (1995). Trivialization: The Forgotten Mode of Dissonance Reduction. *Journal of Personality and Social Psychology*. **68**, 247-260.

Simon, H.A. 1987. Making Management Decisions: The Role of Intuition and Emotion. *Academy of Management Executive*. **1**(1), 57 – 64.

Sloman, S.A. 1996. The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin.* **119**(1), 3–22.

Stanovich, K.E. and R.F. West. 2000. Individual Differences in Reasoning: Implications for the Rationality Debate?, *Behavioral and Brain Sciences*. **22**(5), 645-726.

Stone, J. and Cooper, J., 2001. A Self Standards Model of Cognitive Dissonance. *Journal of Experimental Social Psychology.* **37**, 228-243.

Su, X. 2008. Bounded Rationality in Newsvendor Models. *Manufacturing and Service Operations Management.* **10**(4), 566-589.

Thaler, R. 1980. Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior and Organization.* **1**(1), 39-60.

Webster, D.M. and A.W. Kruglanski. 1994. Individual Differences and the Need for Cognitive Closure. *Journal of Personality and Social Psychology* **67**(6), 1049-1062.

Appendix 2.1 Experimental Setup (Chapter 2)

The experiment tasked respondents to manage the inventory of milk at a small retailer. Respondents could place only one order per week and were given the relevant cost parameters. The simulation was run for twelve simulated weeks, with the same distribution and cost. Each week, respondents were given feedback on excess inventory or lost sales along with a financial report for the previous week. An excerpt from the instructions showing the history is displayed below. Additional detail is available upon request.

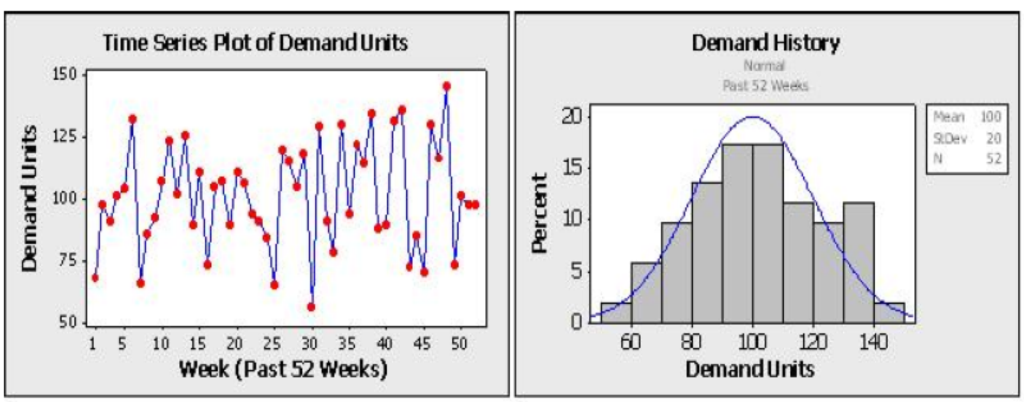
μ_D = 100 gallons/week	c_u = \$10.00/gallon
σ_D = 20 gallons/week	c_o = \$2.00/gallon
p = \$4.00/gallon	g = \$8.00/gallon
c = \$2.00/gallon	s = \$0.00/gallon

Excerpt of Instructions to Participants:

Demand Information

You are given an accurate report about demand history, a portion of which is shown below. The two graphs represent the past year (52 weeks) of data. According to the report, average demand is about 100 gallons per week with a standard deviation of around 20. In the past year, customer demand has ranged from 56 to 145 gallons per week and there were no patterns in the demand. You are confident that the future demand will be similar to the demand in the report.

Each week, you will see an **updated** graph of demand over the most recent 12 weeks.



Appendix 2.2 Expected Profit (Chapter 2)

The figure below shows the expected profit function (equation (1)) for the experimental parameters shown in Appendix 2.1 Many authors (cf. Bolton and Katok 2008) use expected profit as a performance measure because it is independent of the demand realization and, therefore, reduces the role of chance.

Expected Profit Function $\Pi(Q)$

Appendix 3.1 Pre-test Instructions (Chapter 3)

During the simulation exercise, you will be asked to manage inventory at a retail store. Your main task will be to place orders for inventory for a period in advance of knowing the exact customer demand for that period. This inventory is perishable, as inventory from one period cannot be transferred to the next period.

When you arrive at the campus lab, you will be given more instructions regarding the simulation, including how to place orders, relevant costs, and a history of past customer demand. You will be compensated based on your performance. ***Your compensation will be based on minimizing total costs throughout the simulation.***

There are a number of possible decision factors that you might use in deciding on an inventory order quantity:

Prior to the simulation, subjects were asked to both rank and rate (1 = unimportant, 5 = extremely important) the decision factors, including minimizing left over inventory, satisfying the most customers and optimizing costs and profits. Additionally, an open ended question allowed for additional decision factors, but few subjects listed any additional factors.

Appendix 4.1 The Cognitive Reflection Test (Frederick, 2005)

Q1. A bat and a ball cost $1.10 in total. The bat costs $1 more than the ball. How much does the ball cost? ____ cents

Q2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____ minutes

Q3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half the lake? ____ days

Appendix 4.2 Sample Datasets by Condition (Chapter 4)