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

Brent B. Moritz
Ph.D. Candidate
Operations & Management Science Department
Curtis L. Carlson School of Management
University of Minnesota
321 19th Avenue South
Minneapolis, MN 55455-0413 USA
Voice: 612-626-9764
Email: morit022@umn.edu

Arthur V. Hill
The John and Nancy Lindahl Professor
Operations & Management Science Department
Curtis L. Carlson School of Management
University of Minnesota
321 19th Avenue South
Minneapolis, MN 55455-0413 USA
Voice: 612-624-4015
Email: ahill@umn.edu

Karen Donohue
Associate Professor
Operations & Management Science Department
Curtis L. Carlson School of Management
University of Minnesota
321 19th Avenue South
Minneapolis, MN 55455-0413 USA
Voice: 612-625-6320
Email: donoh008@umn.edu

September 30, 2009

WORKING PAPER
Please do not quote or distribute this document without permission from the authors.
Cognition and Individual Differences in the Newsvendor Problem: Behavior Under Dual Process Theory

Abstract

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.

Keywords: Newsvendor, Behavioral Operations, Dual Process Theory, Cognitive Reflection Test

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. 2009; 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. 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 $c_a = p - c + g$ 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

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

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

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

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 \textit{ex post} inventory error, while evidence for demand chasing is weak or non-significant (Schweitzer and Cachon 2000). Bloomfield \textit{et al.} (2007) found that, on average, individuals do not make appropriate changes to the order quantity based on cost and profit. They also find that some of the same behavioral factors found in the newsvendor problem appear in situations where inventory is replenished over time. 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 \textit{et al.} (2009) 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.

2.1 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 basis for this theory draws on work from Simon (1987), who highlights the combined nature of managerial decision-making:

\begin{quote}
It is doubtful that we will find two types of managers (at least, of good managers), one of whom relies almost exclusively on intuition, the other on analytic techniques. More likely, we will find a continuum of decision making styles involving an intimate combination of the two kinds of skill. We will also find that
\end{quote}
the nature of the problem to be solved will be a principle determinant of the mix. (Simon, 1987 p. 61)

Following Simon (1987), many researchers in the cognitive sciences have proposed dual process theories of reasoning. These 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.1 Table 1 is our synthesis of much of this cognitive science literature and compares these two cognitive approaches for decision-making.

Table 1: Comparison of Dual Process Decision-Making

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

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

---

1 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.
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 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 \textit{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) 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: The CRT Instrument**

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>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?</td>
<td>____ cents</td>
</tr>
<tr>
<td>Q2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?</td>
<td>____ minutes</td>
</tr>
<tr>
<td>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?</td>
<td>____ days</td>
</tr>
</tbody>
</table>

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.

### 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 $Q^*$. 

\[ H2: \text{Individuals with higher cognitive reflection will have order quantities closer to the optimal quantity.} \]

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:

\[ H3: \text{Individuals with higher cognitive reflection will have lower order quantity variance.} \]

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. (2009) 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:

\[ H4a: \text{Individuals with higher cognitive reflection will exhibit less anchoring on the mean.} \]

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.

\[ H4b: \text{Individuals with higher cognitive reflection will exhibit less anchoring on the prior period demand.} \]

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.

\[ H4c: \text{Even when individual anchoring tendencies are included, individuals with higher cognitive reflection will have higher expected profit.} \]
Figure 1 illustrates the proposed relationship between Hypotheses H4a, H4b, and H4c and performance.

Figure 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.

\[ H6: \text{When compared with college major, total business experience in years, and managerial position, cognitive reflection is a stronger predictor of expected profit performance in the newsvendor problem.} \]

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 $\sigma_D = 20$). After the newsvendor simulation was completed, each respondent was given the CRT.

5. Analysis and Results

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. 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 3 shows the sample demographics and breakdown by CRT score and by firm.

Table 3  Respondent Demographics and Split by CRT Score

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

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 $\Pi(Q^*) = 940.00/week$. Figure 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

---

$^2$ More detail regarding the parameters of the problem and one of the instruction screens is shown in Appendix 1.

$^3$ Appendix 2 presents a graph of the expected profit function.
order between \( Q^* \) and \( \mu_d \), 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.

**Figure 2  Average Order Quantity by Respondent**

![Graph showing average order quantity by respondent.](image)

**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 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
### 3.1 Expected Profit

<table>
<thead>
<tr>
<th>CRT Group</th>
<th>Average Exp. Profit</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$907.64</td>
<td>1.701</td>
</tr>
<tr>
<td>1</td>
<td>$915.25</td>
<td>1.795</td>
</tr>
<tr>
<td>2</td>
<td>$921.42</td>
<td>1.577</td>
</tr>
<tr>
<td>3</td>
<td>$921.69</td>
<td>2.196</td>
</tr>
</tbody>
</table>

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

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

### 3.2 Order Quantity

<table>
<thead>
<tr>
<th>CRT Group</th>
<th>Average Order Qty.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>107.543</td>
<td>1.004</td>
</tr>
<tr>
<td>1</td>
<td>111.878</td>
<td>1.060</td>
</tr>
<tr>
<td>2</td>
<td>113.406</td>
<td>0.931</td>
</tr>
<tr>
<td>3</td>
<td>114.625</td>
<td>1.296</td>
</tr>
</tbody>
</table>

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

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

### 3.3 Order Quantity Variance

<table>
<thead>
<tr>
<th>CRT Group</th>
<th>Average Variance</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>207.9</td>
<td>16.82</td>
</tr>
<tr>
<td>1</td>
<td>166.9</td>
<td>17.75</td>
</tr>
<tr>
<td>2</td>
<td>132.7</td>
<td>15.59</td>
</tr>
<tr>
<td>3</td>
<td>136.3</td>
<td>21.72</td>
</tr>
</tbody>
</table>

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

Pairwise Multiple Comparisons (Tukey HSD):  
$\mu_0 > \mu_1, p = 0.091$  
$\mu_0 > \mu_2, \mu_3, p \leq 0.001$  
$\mu_1 > \mu_2, \mu_3, p = n/s$  
$\mu_2 > \mu_3, p = n/s$
CRT = 3. The results show that performance generally increases with CRT score, though there is no statistical difference between the highest-performing groups (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.

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:

\[ Q_i = \mu_D + \alpha_i (Q^* - \mu_D) + \epsilon_i \]  

(3)

We use OLS regression to calculate the average score \( \alpha_i \) for each respondent \( i \) that best reflects the deviation in their order quantity in a given period \( (Q_i) \) from the average demand \( (\mu_D) \). As shown in Figure 4, we find that low CRT groups anchored on the mean more often than high CRT groups \( (F = 8.537, p \leq 0.001) \). H4a is supported.

Figure 4 Comparison of Mean Anchoring Tendency by CRT Group

<table>
<thead>
<tr>
<th>CRT Group</th>
<th>Average Anchoring (1-(\alpha))</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.602</td>
<td>0.048</td>
</tr>
<tr>
<td>1</td>
<td>0.382</td>
<td>0.053</td>
</tr>
<tr>
<td>2</td>
<td>0.325</td>
<td>0.045</td>
</tr>
<tr>
<td>3</td>
<td>0.205</td>
<td>0.051</td>
</tr>
</tbody>
</table>

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 \( (1- \alpha) \) to make interpretation easier and to be directionally consistent with \( \beta_i \) and \( \delta_i \).
We next operationalize the second anchoring heuristic (demand chasing / Prior Q Anchoring) as a linear partial adjustment model (again following Bostian et al. 2008):

\[ Q_i = Q_{i-1} + \beta_i(D_{i-1} - Q_{i-1}) + \epsilon_i \]  

(4)

where subject \( i \) considers the previous order \( (Q_{i-1}) \) and adjusts the order quantity based on the realized demand in the previous period \( (D_{i-1}) \). Again using OLS regression, we calculate \( \beta_i \) for each respondent \( i \) that measures the propensity to chase demand. As shown in Figure 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 5  Comparison of Anchoring and Adjustment Tendency (Chasing) by CRT Group**

<table>
<thead>
<tr>
<th>CRT Group</th>
<th>Average Chasing (( \beta ))</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.458</td>
<td>0.028</td>
</tr>
<tr>
<td>1</td>
<td>0.379</td>
<td>0.030</td>
</tr>
<tr>
<td>2</td>
<td>0.299</td>
<td>0.026</td>
</tr>
<tr>
<td>3</td>
<td>0.247</td>
<td>0.029</td>
</tr>
</tbody>
</table>

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

Pairwise Multiple Comparisons (Tukey HSD):
- \( \mu_0 < \mu_1 \)  \( p = n/s \)
- \( \mu_0 < \mu_2, \mu_3 \)  \( p \leq 0.001 \)
- \( \mu_1 < \mu_2 \)  \( p = n/s \)
- \( \mu_1 < \mu_3 \)  \( p = 0.005 \)
- \( \mu_2 < \mu_3 \)  \( p = n/s \)

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 6.

**Figure 6  Mediation Model and Results**

<table>
<thead>
<tr>
<th>Path</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>-0.072</td>
<td>0.014</td>
<td>-5.646</td>
<td>0.000</td>
</tr>
<tr>
<td>$a_2$</td>
<td>-0.126</td>
<td>0.022</td>
<td>-5.695</td>
<td>0.000</td>
</tr>
<tr>
<td>$b_1$</td>
<td>-28.667</td>
<td>2.543</td>
<td>-11.272</td>
<td>0.000</td>
</tr>
<tr>
<td>$b_2$</td>
<td>-14.204</td>
<td>1.462</td>
<td>-9.715</td>
<td>0.000</td>
</tr>
<tr>
<td>$c$</td>
<td>5.533</td>
<td>0.753</td>
<td>7.359</td>
<td>0.000</td>
</tr>
<tr>
<td>$c'$</td>
<td>1.682</td>
<td>0.624</td>
<td>2.694</td>
<td>0.008</td>
</tr>
</tbody>
</table>

$R^2 = 0.513$, $R_{adj}^2 = 0.508$, $F = 108.52 (3,309)$ $p \leq 0.001$

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.

### 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.

$$U_i(Q_{it}) = \Pi(Q_{it}) - \delta_i R(Q_{it}) + \epsilon_i,$$  \hspace{1cm} (5)$$

where expected regret in period $t$ is defined as $R(Q_{it}, D_t) = |Q_{it} - D_t|^2$. The $\delta_i$ parameter is the implied degree of induced regret for each individual found through a double exponential maximum likelihood logit model. As shown in Figure 7, analyzing ex post regret $\delta_i$, 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 7 Comparison of Minimizing Ex Post Inventory Error by CRT Group

<table>
<thead>
<tr>
<th>CRT Group</th>
<th>Average Error Preference ($\delta_i$)</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.482</td>
<td>0.044</td>
</tr>
<tr>
<td>1</td>
<td>0.358</td>
<td>0.048</td>
</tr>
<tr>
<td>2</td>
<td>0.246</td>
<td>0.040</td>
</tr>
<tr>
<td>3</td>
<td>0.205</td>
<td>0.046</td>
</tr>
</tbody>
</table>

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.

### 5.5 Differences in Performance: Alternative Explanations

We investigated alternative explanations of performance including college major\(^4\), 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.

---

\(^4\) 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 6 Comparison of Expected Profit by College Major and CRT Group

<table>
<thead>
<tr>
<th>College Major</th>
<th>Complete Sample</th>
<th>CRT Group and College Major</th>
<th>E[π]</th>
<th>SE</th>
<th>F (df, df error)</th>
<th>Pairwise Multiple Comparisonsa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>E[π]</td>
<td>SE</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Liberal Arts</td>
<td>15</td>
<td>$913.61</td>
<td>5.287</td>
<td></td>
<td>$890.52</td>
<td>7.722</td>
</tr>
<tr>
<td>Business: Accounting/Finance</td>
<td>20</td>
<td>$921.11</td>
<td>3.569</td>
<td></td>
<td>$913.36</td>
<td>-</td>
</tr>
<tr>
<td>Business: Marketing/Management</td>
<td>142</td>
<td>$913.78</td>
<td>1.295</td>
<td>$909.56</td>
<td>2.265</td>
<td>$911.38</td>
</tr>
<tr>
<td>Business: Supply Chain / Operations</td>
<td>57</td>
<td>$919.84</td>
<td>2.065</td>
<td>$906.33</td>
<td>4.397</td>
<td>$917.71</td>
</tr>
<tr>
<td>Education / Social Sciences</td>
<td>12</td>
<td>$919.63</td>
<td>4.500</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Engineering / Physical Sciences</td>
<td>36</td>
<td>$925.13</td>
<td>2.674</td>
<td>-</td>
<td>$915.52</td>
<td>5.548</td>
</tr>
<tr>
<td>Other</td>
<td>24</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

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≤0.001

a Only cells with n ≥ 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[π] = $940.00, E[π]Q=mean=$904.25

e Pairwise multiple comparison of group means (Bonferroni)
Table 7 Comparison of Expected Profit by Years of Experience and CRT Group

<table>
<thead>
<tr>
<th>Years Exp.</th>
<th>Complete Sample</th>
<th>CRT Group and Years of Experience</th>
<th>E[π]</th>
<th>SE</th>
<th>F (df, dferror)</th>
<th>Pairwise Multiple Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n E[π]</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>0 - 1 Year</td>
<td>21</td>
<td>$918.27 4.421</td>
<td>-</td>
<td>-</td>
<td>$913.49 5.335</td>
<td>$918.61 4.939</td>
</tr>
<tr>
<td>1-5 Years</td>
<td>65</td>
<td>$916.45 1.858</td>
<td>$910.71 4.118</td>
<td>$912.93 3.684</td>
<td>$924.54 3.113</td>
<td>$917.62 3.460</td>
</tr>
<tr>
<td>5 - 15 Years</td>
<td>107</td>
<td>$918.55 1.468</td>
<td>$908.18 2.459</td>
<td>$919.34 3.077</td>
<td>$922.39 2.694</td>
<td>$924.29 3.478</td>
</tr>
<tr>
<td>&gt;15 Years</td>
<td>120</td>
<td>$915.10 1.357</td>
<td>$904.19 2.700</td>
<td>$912.64 3.134</td>
<td>$919.81 2.505</td>
<td>$923.79 2.744</td>
</tr>
</tbody>
</table>

F (df, dferror) = 0.710 (3, 309), p=n/s

F(df, dferror)c = 1.038 (3, 297), p=n/s

F (df, dferror)c = 6.080 (3, 297), p≤0.001

a Only cells with n ≥ 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[π] = $940.00, E[π]mean = $904.25
e Pairwise multiple comparison of group means (Bonferroni)

Table 8 Comparison of Expected Profit by Managerial Position and CRT Group

<table>
<thead>
<tr>
<th>Managerial Position</th>
<th>Complete Sample</th>
<th>CRT Group and Managerial Position</th>
<th>E[π]</th>
<th>SE</th>
<th>F (df, dferror)</th>
<th>Pairwise Multiple Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n E[π]</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Non-Managers</td>
<td>245</td>
<td>$915.90 0.949</td>
<td>$907.66 1.682</td>
<td>$913.56 1.955</td>
<td>$922.26 1.717</td>
<td>$921.01 2.074</td>
</tr>
<tr>
<td>Managers</td>
<td>68</td>
<td>$917.19 2.021</td>
<td>$900.89 5.765</td>
<td>$921.40 4.231</td>
<td>$919.66 3.114</td>
<td>$926.81 3.114</td>
</tr>
</tbody>
</table>

F (df, dferror) = 4.949 (1, 311) p=0.026

F(df, dferror)a = 0.233 (1, 305) p=n/s

F (df, dferror)a = 11.726 (3, 305) p≤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]mean = $904.25
d Pairwise multiple comparison of group means (Bonferroni)
limited by the data available for each combination of interest. In looking at expected profits between groups (Table 6), major is a significant predictor of performance ($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 = \text{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 7), 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 ($F = 0.710 \ (3, 309) \ p = \text{n/s}$) 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 8). 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 ($F = 0.233 \ (1, 305) \ p = \text{n/s}$), 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.
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.

References


Appendix 1: Experimental Setup

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 (Figure A1). Additional detail is available upon request.

\[
\begin{align*}
\mu_D &= 100 \text{ gallons/week} & c_u &= $10.00/\text{gallon} \\
\sigma_D &= 20 \text{ gallons/week} & c_o &= $2.00/\text{gallon} \\
p &= $4.00/\text{gallon} & g &= $8.00/\text{gallon} \\
c &= $2.00/\text{gallon} & s &= $0.00/\text{gallon}
\end{align*}
\]

**Figure A1.** 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: Expected Profit

Figure A2 shows the expected profit function (equation (1)) for the experimental parameters shown in Appendix 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.

Figure A2. The Expected Profit Function $\Pi(Q)$