Journal of Operations Management 31 (2013) 72–85

Contents lists available at SciVerse ScienceDirect

Journal of Operations Management

journal homepage: www.elsevier.com/locate/jom

Invited research paper

Individual differences in the newsvendor problem: Behavior and cognitive reflection

Brent B. Moritz[a,*], Arthur V. Hill[b,1], Karen L. Donohue[b,2]

[a] Department of Supply Chain & Information Systems, Smeal College of Business, Penn State University, 469 Business Building, University Park, PA 16082, USA
[b] Supply Chain & Operations Department, Curtis L. Carlson School of Management, University of Minnesota, 321 19th Avenue South, Minneapolis, MN 55455-0413, USA

ARTICLE INFO

Article history:
Available online 7 December 2012

Keywords:
Newsvendor problem
Behavioral Operations
Cognitive reflection

ABSTRACT

Previous research has shown that when solving a newsvendor problem, individuals systematically and persistently deviate from the profit maximizing quantity. This paper investigates the relationship between cognitive reflection and newsvendor decision making, testing experienced supply chain professionals and subjects affiliated with a university business school in a newsvendor experiment. We find that in high and medium critical ratio environments, individuals with higher cognitive reflection exhibit a lower tendency to chase demand. We also find that cognitive reflection is related to task outcome measures including average expected profit, average order quantity and order quantity variance, and that cognitive reflection is a better predictor of performance than college major, years of experience, and managerial position. These results suggest that cognitive reflection contributes to an understanding of newsvendor decision-making behavior.

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 is to select an order quantity between the profit-maximizing optimal quantity and the mean demand. Subsequent work has tested different explanations for this average behavior (Kremer et al., 2010; Su, 2008) and examined how this 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 censored demand (Rudi and Drake, 2009) and group decision making (Gavirneni and Xia, 2009). Several studies have observed wide variation in ordering between individuals, but few studies have documented causal factors to explain this individual variation.

Much of the prior newsvendor research has reported average results, which implicitly assumes that decision makers are homogeneous. 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) identified unobserved heterogeneity, also known as individual differences. In operations management, Doerr et al. (2004) highlighted worker heterogeneity and its impact on the variability of performance in assembly lines. Within inventory research, several studies point to the importance of individual differences in decision makers. Croson and Donohue (2006) called for theoretical research that incorporates individual biases. Similarly, Bendoly et al. (2006) highlighted the importance of finding out if humans act in accordance with how they are modeled, and specifically if differences are systematic and predictable. Cantor and Macdonald (2009) showed that individuals primed for abstract problem solving performed better than those primed for a concrete approach. 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 called for theory to explain individual variance in newsvendor-type decisions.

The goal of this paper is to apply theory from judgment and decision-making to explain some of the individual variation observed in a newsvendor task. We hypothesize that the individual-specific construct of cognitive reflection, as measured by the Cognitive Reflection Test (CRT) (Frederick, 2005), is related to chasing behavior and task outcome. We test this in three experimental studies across a range of critical ratio conditions using both experienced practitioners and students. In Study 1, we focus on the high critical ratio setting since this is the most representative scenario

* Corresponding author. Tel.: +1 814 863 7243.
E-mail addresses: bmoritz@psu.edu (B.B. Moritz), ahill@umn.edu (A.V. Hill), donoh008@umn.edu (K.L. Donohue).
[1] Tel.: +1 612 624 4015.
[2] Tel.: +1 612 625 6320.

0272-6963/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jom.2012.11.006

faced by our industry partners. Our subjects are from a pool of experienced supply chain managers/analysts from three Fortune 500 firms. Using experienced professionals also allows us to test the impact of individual characteristics that are often considered in hiring decisions, such as college major, years of professional service, and managerial position. In Study 2, we evaluate outcomes in high, medium and low critical ratio conditions in a controlled laboratory environment using subjects from a business school subject pool. In Study 3, we further test the low critical ratio setting using experienced professionals from a fourth firm.

We find that individuals with high cognitive reflection are less likely to chase demand in high and medium critical ratio settings, but observe no significant relationship between cognitive reflection and chasing behavior in low critical ratio settings. Similar results hold for task outcomes, including expected profit, average deviation from the optimal order quantity and order variance. The data also shows that cognitive reflection is a better predictor of task outcome than other individual characteristics such as college major, years of experience, or managerial position.

Section 2 begins with an introduction to the theory underlying this research. Section 3 develops the research hypotheses and the experiment. Sections 4–6 report the results, and Section 7 summarizes the findings and suggests opportunities for further research.

2. Theory development

In the newsvendor model, a decision maker is faced with the task of 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 cost of having one too few units relative to demand (underage) or one too many units relative to demand (overage) are then $c_u = p - c + g$ and $c_o = c - s$, respectively. 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 solution to a newsvendor problem is to choose the order quantity that maximizes expected profit

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

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

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

where $F^{-1}(.)$ is the inverse of the cumulative distribution function for demand and $c_u/(c_u + c_o)$ is the critical ratio.

Although the newsvendor problem has a long history of published research (Edgeworth, 1888), deviations from the optimal order quantity are frequently observed in both experimental and industrial environments (e.g., Fisher et al., 1994). In repeated newsvendor contexts, persistent and well-documented deviations from the optimal order quantity include a tendency to over-order in a low critical ratio setting and under-order in a high critical ratio setting. Previous research suggests a number of possible explanations for this behavior. Even when the distribution of demand is known, average ordering behavior is somewhat consistent with heuristics such as anchoring on the mean, while prior evidence for consistent demand chasing is weak or non-significant (Schweitzer and Cachon, 2000). Bostian et al. (2008) observed that for some individuals, order quantity decisions are consistent with use of a demand-chasing heuristic. Similarly, Kremer et al. (2010) reported that, in some cases, demand chasing is significant at the individual level. Feng et al. (2011) found significant differences between American and Chinese subjects in both anchoring on the mean and demand chasing. In a more complex multi-echelon setting, Bloomfield et al. (2007) found that order quantities selected by individuals were not sufficiently sensitive to relative costs. Bloomfield et al. (2007) also found that some of the same behavioral factors in the newsvendor problem also occur 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. However, they observed anecdotally that the tendency for “too-quick” conclusions based on small samples seemed to vary widely between individuals. Olivares et al. (2008) examined newsvendor decision making in a healthcare setting and devised structural estimates of the mismatch cost ratios implied by observed inventory decisions. They found that individuals place greater weight on more tangible overage costs (e.g., idle operating room capacity) than on less tangible underage costs (e.g., staff overtime). While many papers describe potential heuristics and preferences that individuals might use to solve a newsvendor problem, little theory has emerged to explain or predict the observed heterogeneity between individuals.

2.1. Cognitive reflection and 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. While a number of possible heuristics have been proposed to explain newsvendor behavior, our research posits that cognitive reflection (Frederick, 2005) provides a theoretical foundation for understanding and explaining a portion of the individual heterogeneity observed in newsvendor decisions. Cognitive reflection refers to the tendency of an individual to let his or her System 2 process moderate, override, or endorse an initial System 1 response. Rooted in dual process theories of decision making (Stanovich and West, 2000), these two systems are parallel cognitive approaches activated when an individual solves a problem. System 1 processes are typically described as intuitive, tacit, contextualized and rapid while System 2 processes are reflective, analytical and rely on abstract reasoning. While there is a large body of literature detailing aspects of these two approaches, and not all scholars agree on terminology and all details of the two processes, (e.g., Evans, 1984, 2008; Hammond, 1996; Sloman, 1996; Stanovich and West, 2000; Kahneman and Frederick, 2002; Kahneman, 2011), the key concept is that these are different cognitive processes that are simultaneously active in decision making. The two systems work together, with System 1 generating suggestions for System 2 to consider in the forms of “impressions, intuitions, intentions and feelings” which, if endorsed by System 2, turn into beliefs and voluntary actions (Kahneman, 2011). Kahneman (2011) notes the degree of cognitive reflection may vary by individual, as well as with task environment and experience.

In some newsvendor decision contexts, intuitive, descriptive, and experiential decision inputs (typically governed by System 1) may have a role in generating the profit-maximizing order quantity. This occurs when limited relevant historical demand data is available to characterize future demand. For example, some individuals may have a particularly keen intuitive sense for predicting future demand for fashionable and trendsetting items. Similarly, while it is often difficult to estimate lost goodwill (g), some individuals may have skills or experience in making such an estimate. In such settings, System 1 may play a larger role in the decision process.

Table 1
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 min 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

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 min and 24 days, respectively.

Rather than test context-specific parameter estimation, our research focuses on individual differences in selecting the order quantity assuming the parameters are available and stable (i.e., where context-specific knowledge has limited importance). This is the newsvendor decision context that is most commonly studied in the Behavioral Operations literature and where higher cognitive reflection is more likely to be beneficial. The newsvendor model in Eqs. (1) and (2) assumes that the cost and demand parameters are fixed; a static order quantity policy is optimal and does not change with demand realizations in prior periods. To determine the order quantity that maximizes expected profit, the decision maker must accurately apply the cost parameters (or at least the critical ratio) and understand the demand distribution. With this information and the logic of Eq. (2), the decision maker can solve for Q^*. To our knowledge no previous research has tested cognitive reflection in the repeated newsvendor context, so it serves as a natural starting point as a behavioral explanation of observed individual decision making behavior. If cognitive reflection is not related to task outcomes in this setting, it is less likely to be relevant in other, more complex environments.

2.2. Measuring cognitive reflection

Kahneman and Tversky (1982) suggest studying systematic errors in reasoning because those errors may expose cognitive limitations or reveal the processes and procedures governing statistical or logical inference. Such observations might highlight System 1 features that create an error, or suggest reasons why an error was not overridden by a System 2 process (Kahneman and Frederick, 2002). Frederick (2005) concluded that the CRT is a measure of the ability to avoid acting based on the initial response to a problem. Frederick (2005) observed that individual decision makers differ in terms of cognitive reflection and developed the Cognitive Reflection Test (CRT) to measure this tendency. The CRT consists of three quantitative items (Table 1) where the obvious, impulsive answer is incorrect. The correct answers are only found if each respondent moderates their System 1 intuition with a System 2 approach to override their initial response. Frederick defined the CRT score for an individual as the sum of the number of correct responses on the Cognitive Reflection Test.

Within cognitive psychology, additional research is investigating the relationship between intelligence and cognitive reflection in decision making. The results indicate that cognitive reflection differs from IQ and that lower cognitive reflection correlates with certain heuristics and biases. Frederick (2005) found positive correlation between the CRT and proxies for IQ such as the SAT, ACT, and the Wonderlic Personnel Test. However, he also found that the CRT is correlated with an individual's choice of a more patient temporal option, more risk seeking option in gains, and less risk seeking option in losses. These individual time and risk tendencies are not strongly correlated with IQ. Toplak et al. (2011) found that cognitive reflection was a predictor of syllogistic reasoning, inhibition, superstitious thinking, as well as a number of other heuristics and biases. They also found that cognitive reflection was able to predict substantial variance that was not related to intelligence, and noted that the CRT accounted for over twice as much unique variance in rational thinking as IQ. They also noted that the CRT had a unique ability to predict performance on a heuristics and bias composite beyond that of cognitive ability, executive functioning, or thinking dispositions. This is similar to the results in Oechssler et al. (2009), who found that higher cognitive reflection was consistent with lower incidences of biased decision making. In particular, individuals with low cognitive reflection were more likely to exhibit the conjunction fallacy and to demonstrate conservatism toward the midpoint in a standard probability judgment exercise. Cesarini et al. (2012) found that individuals with high CRT scores are resistant to behavioral anomalies such as the illusion of control, insensitivity to sample size and representativeness. More broadly, Stanovich (2011) calls for development of new instruments that would measure aspects of rational thought (including cognitive reflection) that are missed by existing conventional IQ tests. Taken together, these studies show that CRT scores, while correlated with IQ, also capture other important aspects of rational decision making. It is beyond the scope of this paper to definitively separate the psychological processes of cognitive reflection and other individual differences—this is a task receiving considerable attention by the cognitive science community. However, an interesting avenue for future research would be to compare cognitive reflection to other measures of individual differences in predicting performance on the newsvendor task.

More importantly, typical errors in newsvendor decisions, such as demand chasing, anchoring on the mean, or underweighting the margin, are examples of rational thinking errors typical of miserly processing. In such behaviors, a decision maker defaults to the first answer available or focuses on an available anchor, and subsequently fails to correct that first response through System 2 override. Another class of miserly processing error, framing, has also been investigated in a newsvendor context by Schultz et al. (2011). Given that prior research has highlighted the unique and powerful relationship between cognitive reflection and individual performance in heuristics and biases tasks, our research compares cognitive reflection with newsvendor decision making.

3. Hypotheses and experimental design

Based on Frederick (2005), individuals who score low on the CRT are prone to answer with the first idea that comes to mind. These individuals are thought to exhibit a weaker System 2 supervisory function. In the context of the newsvendor problem, a weaker System 2 supervisory function could lead to the use of more intuitive heuristics, such as chasing demand, or sub-optimal but readily available answers, such as ordering the mean demand. Such behavior may result in less order stability or lower average profit.

Our first hypothesis compares demand chasing behavior with cognitive reflection. Chasing behavior in a newsvendor context is typically defined as the tendency of an individual to anchor on the previous order quantity and adjust toward the most recent demand realization (Bostian et al., 2008). The amount of chasing is captured by the strength of adjustment toward that demand realization (see Section 4.2). While demand chasing is a plausible decision shortcut, adjusting toward the most recent demand is suboptimal in our decision context since the demand distribution is stable (i.e., has no trends). Because individuals with higher cognitive reflection may be less likely to answer with the first response that comes to mind, we hypothesize they will be less likely to chase demand than individuals with lower cognitive reflection.

H1 (COGNITIVE REFLECTION and DEMAND CHASING). When making repeated newsvendor decisions, individuals with higher cognitive reflection will exhibit less chasing of prior period demand.

Our second hypothesis predicts a relationship between individual cognitive reflection and newsvendor task outcome. In the context of a repeated newsvendor decision, possible initial responses might be to order the mean demand (thus underweighting the margin implications), or order based on the realized demand or left over inventory from a prior period. We hypothesize that since individuals with higher cognitive reflection may be more likely to override such heuristics, their orders may have lower variance across periods and achieve higher average performance.

H2 (COGNITIVE REFLECTION and TASK OUTCOME). When making repeated newsvendor decisions, individuals with higher cognitive reflection will have higher expected profit, order quantities closer to the optimal quantity, and have less variance in their order quantities.

While these outcome measures are not orthogonal (as they are all based on the observed order quantity), each measure addresses different aspects of performance. For example, a decision maker who orders Q^* each period will achieve a higher long-term expected profit than a second decision maker who orders exactly the mean demand each period. However, both individuals would have the same (zero) variance. Cognitive reflection may also influence task outcome even when the impact of demand chasing is included. We test for this second order effect through a mediation model described in Section 6.

Another question of interest is whether other *ex ante* individual characteristics, such as those available during hiring decisions, are linked to cognitive reflection or help explain task outcome. For example, decision makers with certain college majors (e.g., engineering, supply chain, and finance) may be more inclined to use System 2 processing than those with other majors. Similarly, Bolton et al. (2008) found that years of experience was related to lower performance in the newsvendor problem, while a higher managerial position was related to higher performance. We test for these alternate explanations in Section 4.4.

We ran several computer-based newsvendor experiments to test these hypotheses. The experimental design followed that of prior newsvendor experiments, where respondents made order quantity decisions over several periods, with three important differences. First, the simulated demand was normally distributed rather than uniformly distributed. This was important for external validity since demand is rarely uniformly distributed in practice and our subject pool of professionals (Studies 1 and 3) was more accustomed to this type of demand. Additionally, this choice followed Su's (2008) recommendation of studying decision biases under non-uniform demand.

Second, we designed the instrument for remote online access for Studies 1 and 3 to obtain results from industry professionals across multiple firms and locations. None of the industry respondents were compensated based on individual performance because of the difficulty and expense of developing a meaningful compensation scheme for professional subjects across multiple locations. However, each firm was given a confidential benchmarking report comparing their performance to the other firms. The potential loss of experimental control and lack of direct compensation is specifically addressed in Study 2, where the experiment was performed in a controlled laboratory environment with cash compensation based on performance in the newsvendor exercise.

Third, because we are particularly interested in individual differences, we chose to focus on a limited number of treatments and used a larger subject pool to ensure that a sufficient number of subjects within each CRT category were available for analysis. After consultation with our industry partners, we chose to place primary emphasis on the high critical ratio condition since it is most similar to the environment of the firms in our research. Consequently, Study 1 had the largest number of participants and is featured prominently in our results.

4. Study 1 – analysis and results

Study 1 was carried out with supply chain managers and analysts who regularly made inventory or supply chain planning decisions. The participants, employed at one of three Fortune 500, supply-chain-intensive firms, took part in a high critical ratio setting. This setting was defined by $p = 4$, $c = 2$, $g = 8$, $s = 0$, which implies $c_u = 10$ and $c_o = 2$, a critical ratio of 0.83, and $Q^* = 119.4$ units. Demand was randomly generated from the normal distribution with $\mu_D = 100$ and $\sigma_D = 20$. A single demand stream was used to ensure that we could compare results across firms with sufficient statistical power. For this demand realization, the observed average and standard deviation of demand were $\bar{D} = 103.5$ and $s_D = 24.9$. See Appendix A for more information on the experiment instructions.

The process was initiated in each firm by a senior executive who sent an e-mail to potential respondents through an existing employee distribution list, inviting them to participate and promising confidentiality. The experiment proceeded as follows: First, each respondent provided basic demographic information about their work experience and education. Next, each respondent made repeated order quantity decisions for one product in a simulated retail environment. At the end of each period (a virtual week), respondents received an updated report with the actual (realized) customer demand, the number of units discarded or short, and a calculation of their financial performance for the week. The task instructions included a clear statement of the cost and demand distribution parameters, and emphasized the stable demand distribution (i.e., no trends or correlation over time). This approach allowed our research to focus on individual differences in a stable problem context. In other words, differences in ordering behavior cannot be attributed to differences in the demand realizations, but may be attributed to individual differences in ability to estimate the critical ratio and appropriately apply it with the demand distribution. After the newsvendor simulation was completed, each respondent was given the CRT.

4.1. Summary observations

A total of 319 professionals participated in this study. As a check against respondents answering randomly and completing the exercise very quickly, we recorded the elapsed time to complete the entire instrument.[3] We removed three subjects who completed the exercise unusually quickly (less than half the average time) and three subjects because their order quantities exactly matched demand in every period, yielding a final sample of $n = 313$ practitioners. Following Frederick (2005), each individual was classified into one of four CRT groups based on their answers to the instrument. Table 2 shows sample demographics and CRT scores by firm. Although the distribution of CRT scores in Study 1 includes a number of low scores, the professionals performed well overall. On average, the practitioners had higher mean CRT scores than the 3428 individuals reported in Frederick (2005, Table 1). Specifically, of the eleven sample populations in Frederick (2005), only the student population from MIT had higher average CRT scores than the practitioners in our study.

[3] Post hoc, we also regressed elapsed time against CRT and performance but found no relationship.

Table 2
Respondent demographics and split by CRT score – Study 1.

Firm	Number of respondents	Avg. years of professional experience	S.D. 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

As in previous studies, the average order quantity deviated toward the mean demand with $\bar{Q} = 112.4$ units versus $Q^* = 119.4$ (see Fig. B1 in Appendix B for more detail). The average expected profit across all respondents was \$916.80/week compared with the optimal expected profit of $\Pi(Q^*)$ = \$940.00/week. The range of average order quantities varied from 96 to 160 with an interquartile range (106, 117) and a median of 110. While 83% of respondents had an average order quantity between Q^* and μ_D, the bias toward the mean varied between individuals.

4.2. Differences in use of demand chasing by CRT group

To test for differences in demand chasing behavior, we fit the data to a demand chasing heuristic represented as a linear partial adjustment model (following Bostian et al., 2008).

$$Q_{it} = Q_{it-1} + \beta_i(D_{t-1} - Q_{it-1}) + \varepsilon_{it} \tag{3}$$

This model assumes that 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}). The β parameter represents the strength of the tendency to move toward the most recent demand, with $\beta_i = 1$ implying complete demand chasing (i.e., setting the order quantity exactly equal to the last period demand). Using OLS regression, we estimated β_i for each respondent, and compared performance using a GLM ANOVA procedure.[4] As shown in Fig. 1, the magnitude of adjustment in the direction of the last period demand (i.e., chasing) differed by group ($F = 5.667$ (3, 307) $p \leq 0.001$), with low CRT groups exhibiting higher chasing behavior. This supports H1.

Contrary to previous research that shows only limited support for the chasing heuristic, many respondents appeared to exhibit chasing behavior by moving their responses in the direction of the most recent demand. This may be an appropriate strategy if demand has an underlying trend, but in our experiment demand was drawn from a stationary distribution with no trend or autocorrelation. The key finding of interest is that the magnitude of chasing varied based on the individual's level of cognitive reflection. In particular, the low CRT group exhibited nearly 50% more chasing than the highest CRT groups, as measured by the average beta value from Eq. (3).

The demand chasing metric (3) could return more than expected false positives (Type I errors) if no demand chasing is present (Lau and Bearden, in press). Hence, we also evaluated the correlation between order quantity (Q_t) and prior demand (Q_{t-1}) for each subject i. This analysis found evidence of widespread chasing behavior, as 86% of the experienced practitioners (270 of 313) exhibited positive correlation; the correlation was greater than 0.5 for 76% (238 of 313). GLM ANOVA showed that the results differed by CRT group ($F = 8.554$ (3, 309) $p \leq 0.001$) and that correlation decreased as CRT scores increased. While correlation can detect chasing behavior, it does not capture the magnitude of the change in order quantity: If demand goes up, an increase in order quantity is consistent with positive correlation if it is one unit or one hundred units. Eq. (3) captures "order quantity-weighted" chasing behavior and therefore we use it for the remainder of the paper.

4.3. Task outcome by CRT score

We tested the impact of cognitive reflection on task outcome, again using GLM ANOVA.[5] As shown in Fig. 2, average expected profit, average order quantity, and order quantity variance all move closer to their optimal values with increasing cognitive reflection. Task outcome differs significantly by CRT group for each performance measure: average expected profit ($F = 15.833$ (3, 307) $p \leq 0.001$), average order quantity ($F = 11.384$ (3, 307) $p \leq 0.001$), and order quantity variance ($F = 4.552$ (3, 307) $p \leq 0.01$). Performance generally increases with CRT score, particularly between the low CRT group (CRT = 0) and the high CRT groups (CRT = 2 and CRT = 3), though there is no statistically significant difference between the two highest-performing groups (CRT = 2 and CRT = 3). Further analysis (not shown) indicates that these performance results were similar between each of the three firms. In summary, in a high critical ratio setting, individuals from the lowest CRT categories have a lower average expected profit, an average order size further from the optimal quantity, and a higher order quantity variance than those in the highest CRT groups. This supports H2.

4.4. Differences in expected profit: alternative explanations

We next investigated whether other *ex ante* individual characteristics that are often considered in a hiring decision are linked to cognitive reflection or help explain differences in task outcome. For this analysis we focused on three characteristics, college major, years of experience, and managerial position, and one outcome measure, expected profit. The results are presented in Appendix C. Briefly, we found that college major is related to expected profit, ($F = 3.605$ (6, 299) $p = 0.002$), with Business (Accounting/Finance) and Engineering/Physical Science majors performing better than other majors. However, if cognitive reflection is included in the analysis, the effect of major is not significant ($F = 1.709$ (6, 278) $p = 0.119$) while cognitive reflection is significant ($F = 9.310$ (3, 278) $p \leq 0.001$). In addition, within a particular major, average expected profit increases with CRT score, 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) $p = 0.100$).

Turning to business experience (Table C2), previous research suggests that expected profit decreases with years of experience (Bolton et al., 2008). Our research found no significant relationship between average expected profit and years of experience ($F = 0.710$ (3, 309) $p = 0.547$) or using a regression of actual (non-categorized) years of experience (not shown). Additionally, as before, CRT was

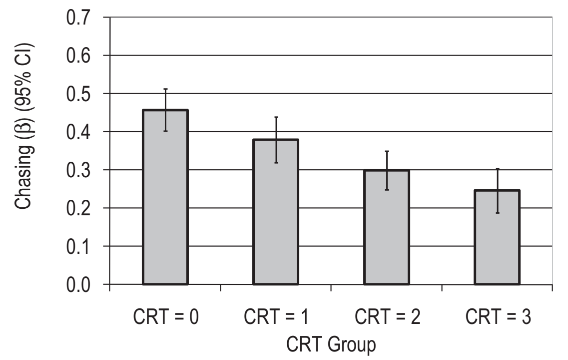
[4] Consistent with their higher proportion of high CRT respondents, Firm C performed slightly better than the other firms and we control for firm effects where appropriate.

[5] As an alternative, we also fit a linear mixed model (LMM), which yielded similar results and no change in the conclusions.

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

CRT Group	*Average Chasing (β)*	*Std. Dev.*
0	0.458	0.232
1	0.379	0.229
2	0.299	0.237
3	0.247	0.238

CRT Group Significance
F = 5.667(3, 307) $p \le 0.001$ R^2 = 0.055



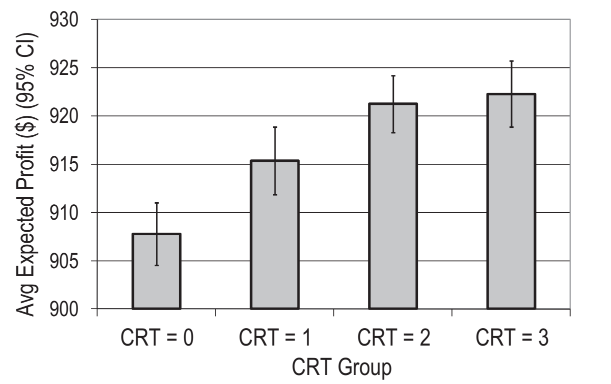
[a] Controlling for firm effect.

Fig. 1. Anchoring and adjustment tendency (chasing) by CRT group – Study 1.

2a Average Expected Profit[a]

CRT Group	*Average Exp. Profit*	*Std. Dev.*
0	$907.79	14.8
1	$915.38	14.6
2	$921.27	14.5
3	$922.29	14.7

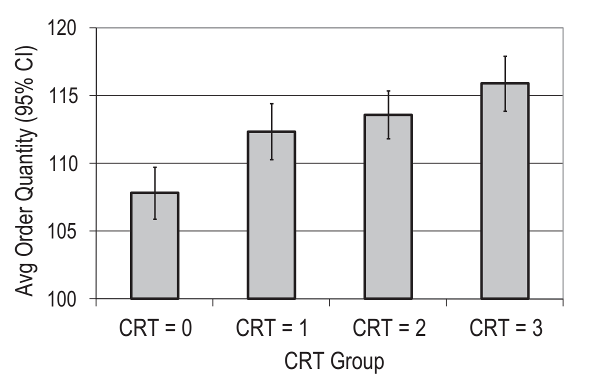
CRT Group Significance
F = 15.833 (3, 307) $p \le 0.001$



2b Average Order Quantity[a]

CRT Group	*Average Order Qty.*	*Std. Dev.*
0	107.83	8.77
1	112.36	8.60
2	113.61	8.63
3	115.91	8.73

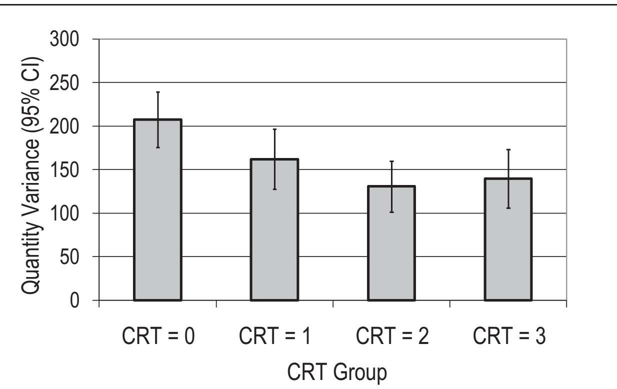
CRT Group Significance
F = 11.384 (3, 307) $p \le 0.001$



2c Order Quantity Variance[a]

CRT Group	*Average Variance*	*Std. Dev.*
0	207.6	144.9
1	162.1	143.2
2	130.6	142.5
3	139.8	145.1

CRT Group Significance
F = 4.552 (3, 307), $p \le 0.004$



[a] Controlling for firm effect

Fig. 2. Comparison of task outcome by CRT group – Study 1.

Table 3
Experimental parameters – Studies 2 and 3.

Critical ratio	Price (p)	Cost (c)	Goodwill (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
Low (Study 3)	$3.00	$2.50	$0.00	$0.50	$2.50	0.17	80.6

Table 4
Overall results – Studies 2 and 3.

	Critical ratio	n	Mean	S.D.	Comparison of group means[a]
Average expected profit ($)	High	56	501.9	14.8	$F = 38.656$ (2, 173)
	Medium	56	146.7	3.9	$p < 0.001$
	Low	60	36.6	6.2	
	Low (Study 3)	46	21.1	6.0	
Average order quantity	High	56	105.6	7.2	$F = 18.828$ (2, 173)
	Medium	60	100.9	6.6	$p < 0.001$
	Low	60	97.5	7.5	
	Low (Study 3)	46	101.0	4.0	
Average order quantity variance	High	56	210.0	194.8	$F = 0.621$ (2, 173)
	Medium	60	174.6	149.2	$p = 0.621$
	Low	60	192.4	167.7	
	Low (Study 3)	46	142.2	118.7	

[a] F-test by critical ratio for Study 2 only; (Study 3 differs by critical ratio and demand distribution).

a robust predictor of task outcome across the range of experience ($F = 6.080$ (3, 297) $p \leq 0.001$).

Lastly, we compared the performance of managers versus non-managers (individual contributors, typically identified as analysts) (Table C3). Taken separately, managers had higher average expected profit than non-managers ($F = 4.989$, (1, 311) $p = 0.026$). However, when cognitive reflection is included in the model, the average expected profit of managers and non-managers was not statistically different ($F = 0.233$ (1, 305) $p = 0.630$), although CRT is again significant ($F = 11.726$ (3, 305) $p \leq 0.001$). We conclude that managers achieved higher average expected profit than non-managers in our study, but that CRT is a better predictor.[6]

5. Studies 2 and 3 – analysis and results

While it is valuable to use practicing supply chain professionals as the subject pool to increase external validity, this design characteristic of Study 1 came with the loss of some experimental control, including the ability to provide an incentive-compatible payout to participants. To test the robustness of our results, we conducted Study 2 in an incentive-compatible controlled laboratory setting, drawing 176 subjects from a business school subject pool. This study also expanded the number of critical ratio settings considered; otherwise, the experiment was similar to Study 1. We collected demographics and administered the CRT instrument before having subjects experience a high, medium, or low critical ratio newsvendor setting. The simulated demand was generated from a normal distribution with mean demand $\mu_D = 100$ and standard deviation of demand $\sigma_D = 30$. For the actual realized sample, the average demand was $\bar{D} = 97.0$ and the standard deviation was $s_D = 27.1$. As part of the instructions for their respective critical ratio conditions (parameters shown in Table 3), subjects were told that they would be compensated in cash based on their newsvendor decision performance. Subject payoffs ranged from $11 to $15; the exact payoff formula was adjusted based on the critical ratio condition and the average cash compensation was $12.70 per subject. We also compared task outcome by critical ratio condition, as shown in Table 4. Average order quantity increased with the critical ratio, which suggests that subjects were directionally responsive to the critical ratio conditions (see also Fig. B2 in Appendix B).

To gain more insight into the low critical ratio setting, we also conducted a third study with practitioners. This allows us to disentangle the impact of critical ratio (high versus low) from differences in subject pool (students versus practitioners). As in Study 1, the participants in Study 3 were practicing professionals from a fourth Fortune 500 firm that regularly made inventory or inventory-related decisions. Forty-six subjects participated from this firm; they averaged eight years of experience, and as in Study 1 they were not directly compensated based on performance. Study 3 used the same demand distribution (100, 20) and the same demand realizations as Study 1; the only substantial difference between Studies 1 and 3 was the critical ratio setting (0.83 in Study 1 and 0.17 in Study 3) (see Fig. B3 in Appendix B). Post-hoc, we also looked for any systematic differences between Study 1 and Study 3. While we drew from a different firm for Study 3, all four firms were highly regarded Fortune 500 firms and had a significant supply chain component in their businesses. There were no obvious differences by years of experience, education or managerial position.

5.1. Differences in use of demand chasing by CRT group

Study 1 showed that cognitive reflection is related to demand chasing behavior, which can be observed *ex post* by comparing the

[6] We also investigated performance by functional group within a firm, but these results were limited by sample size and company-specific titles. More importantly, where data were available, higher cognitive reflection also indicated higher performance within a functional group.

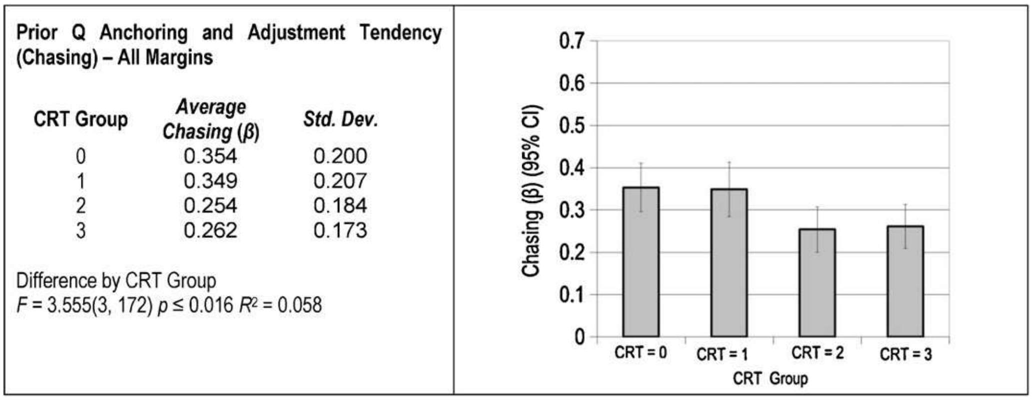


Fig. 3. Chasing by CRT group – Study 2.

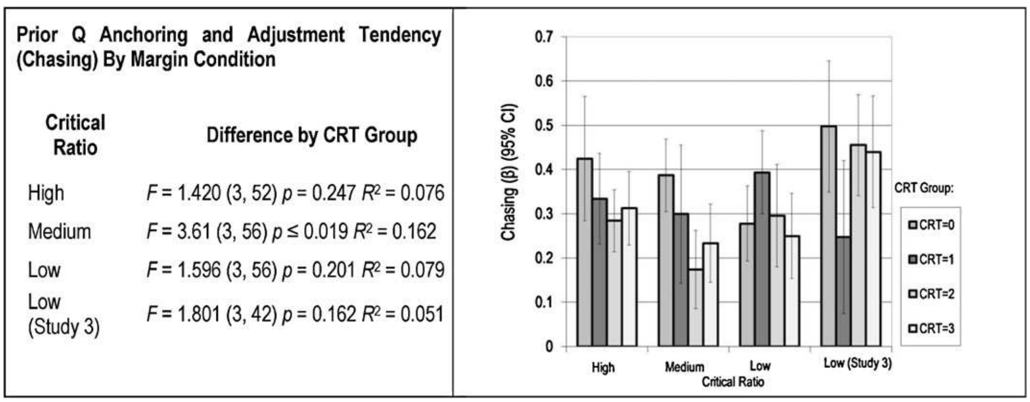


Fig. 4. Chasing by CRT group and margin – Studies 2 and 3.

Table 5
Cognitive reflection and task outcome by critical ratio (Studies 2 and 3).

Critical ratio	Outcome measure	Optimal value[a]	CRT = 0	CRT = 1	CRT = 2	CRT = 3	Comparison of group means by group[b]
High	*n*		11	14	18	13	
	Average expected Profit	$523.73	$488.01	$498.60	$507.60	$509.27	$F = 7.312$ (3, 52), $p < 0.001$
	Average order quantity	120	100.7	104.5	107.5	108.2	$F = 3.159$ (3, 52), $p = 0.032$
	Average order variance	0	340.2	277.0	123.0	148.1	$F = 4.569$ (3, 52), $p = 0.006$
Medium	*n*		18	10	14	18	
	Average expected profit	$152.13	$145.02	$146.66	$148.35	$147.99	$F = 2.669$ (3, 56), $p = 0.056$
	Average order quantity	100	97.1	101.3	99.8	105.3	$F = 5.960$ (3, 56), $p = 0.001$
	Average order variance	0	238.8	196.3	137.1	126.6	$F = 2.211$ (3, 56), $p = 0.097$
Low	*n*		18	16	14	12	
	Average expected profit	$46.40	$36.33	$36.12	$35.17	$39.26	$F = 1.036$ (3, 56), $p = 0.384$
	Average order quantity	80	96.7	98.8	98.5	95.7	$F = 0.533$ (3, 56), $p = 0.662$
	Average order variance	0	231.7	184.2	222.3	109.4	$F = 1.509$ (3, 56), $p = 0.222$
Low (Study 3)	*n*		10	7	16	13	
	Average expected profit	$35.01	$21.15	$21.76	$21.56	$19.35	$F = 0.505$ (3, 42), $p = 0.674$
	Average order quantity	80	99.4	101.8	100.8	102.0	$F = 0.932$ (3, 42), $p = 0.434$
	Average order variance	0	171.0	76.0	130.1	170.4	$F = 1.242$ (3, 42), $p = 0.306$

[a] Optimal values based on the demand distribution and critical ratio parameters.
[b] ANOVA results for each measure by critical ratio condition.

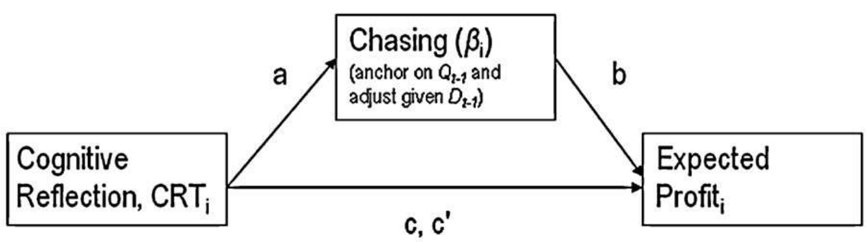
observed order quantities with prior period demand. To examine whether this relationship holds for other critical ratio conditions, we extended the chasing results from Study 1 (Section 4.3) to Studies 2 and 3. First, we report the overall chasing results (across all margin conditions) for Study 2, as shown in Fig. 3. These results indicate that on average, chasing is occurring in Study 2, and individuals with low CRT scores exhibit higher average chasing behavior, supporting H1. However, the results are less clear if segmented by critical ratio condition and by CRT scores, as shown in Fig. 4. While the average beta is highest for individuals with low CRT scores in the high and medium critical ratio conditions, the only statistically significant difference is in the medium critical ratio condition. There is no difference in chasing behavior in the low critical ration conditions. The results by critical ratio and CRT group may be somewhat limited by the smaller sample size (see Table 5) of subjects in that particular combination. In Study 1, the total number of participants was 313, with a range from 67 to 94 in each CRT category. Here the total number of participants is at most 60 per treatment with a range from 7 to 18 in each CRT category. Note that this smaller sample size was a consequence of our objective to provide a robustness test of our findings from Study 1 (focused on a high critical ratio setting and professional decision makers) across a number of different settings. It appears that H1 is partially supported for medium and high critical ratio environments with student participants, but is not supported for the low critical ratio settings.

5.2. Task outcome by CRT score

Next, we compared task outcome based on cognitive reflection for each of the three critical ratio conditions in Study 2 and for Study 3, as shown in Table 5. As discussed above, behavior appears to be influenced by the margin condition. In the High and Medium critical ratios of Study 2, individuals with higher cognitive reflection have higher expected profit, tend to order closer to the optimal order quantity, and have lower variance in their order quantities across the duration of the experiment, supporting H2. However, in the low critical ratio setting, there is no difference in outcome for either the student subject pool from Study 2 or the practitioners (Study 3). In the low critical ratios, one consistent observation is that the average order quantity is close to the mean demand, suggesting that individuals are reluctant to order below the mean even when doing so would maximize their expected profit. We conclude that cognitive reflection is correlated with task outcome in high and medium critical ratio contexts, but not in low critical ratio settings. These results partially support H2.

6. Mediation model – all studies

Based on these results, we further explored the relationship between cognitive reflection and task outcome when chasing behavior is considered. Our empirical findings supported the



Study Name *Critical Ratio*	Model Summary	Path	Coefficient[a]	Standard Error	*t*	*p*
Study 1 *High*	$R^2 = 0.364$	a	-0.072	0.013	-5.646	≤ 0.001
	$R_{adj}^2 = 0.360$	b	-29.748	2.898	-10.264	≤ 0.001
	$F = 88.836$ (2, 310)	c	5.533	0.752	7.359	≤ 0.001
	$p \leq 0.001$	c'	3.395	0.683	4.970	≤ 0.001
Study 2 *High*	$R^2 = 0.383$	a	-0.019	0.011	-1.773	0.082
	$R_{adj}^2 = 0.359$	b	-28.009	8.927	-3.137	0.002
	$F = 16.416$ (2, 54)	c	3.396	0.729	3.925	≤ 0.001
	$p \leq 0.001$	c'	2.861	0.728	3.925	≤ 0.001
Study 2 *Medium*	$R^2 = 0.357$	a	-0.033	0.011	-3.116	0.003
	$R_{adj}^2 = 0.335$	b	-10.177	2.199	-4.628	≤ 0.001
	$F = 15.847$(2, 57)	c	0.569	0.206	2.757	≤ 0.008
	$p \leq 0.001$	c'	0.233	0.192	1.214	0.230
Study 2 *Low*	$R^2 = 0.170$	a	-0.009	0.011	-0.838	0.406
	$R_{adj}^2 = 0.141$	b	-13.009	3.856	-3.374	≤ 0.001
	$F = 5.828$ (2, 57)	c	0.171	0.353	0.483	0.631
	$p \leq 0.005$	c'	0.050	0.327	0.153	0.879
Study 3 *Low*	$R^2 = 0.526$	a	0.007	0.016	0.444	0.659
	$R_{adj}^2 = 0.500$	b	-18.004	2.659	-6.769	≤ 0.001
	$F = 23.536$(2, 43)	c	-0.312	0.397	-0.784	0.437
	$p \leq 0.001$	c'	-0.185	0.280	-0.659	0.514

[a] The coefficients can be interpreted as in regression; the change in output per unit increase in the input.

Fig. 5. Mediation model and results – all studies.

relationship between cognitive reflection and chasing behavior (H1) and task outcome (H2). However, we wanted to see if the outcome was only about chasing. Said another way, we investigated if cognitive reflection and ordering behavior is exclusively a chasing story or if there are other aspects of rational thinking errors possibly captured by cognitive reflection that impact newsvendor results outside of chasing. To do this, the mediation model (Fig. 5) captures how cognitive reflection may impact expected profit both indirectly through chasing (path a) and directly (path c′) beyond *ex post* assessments of chasing (notation as in Baron and Kenney, 1986). We tested the path significance of the mediation model in each critical ratio specifically by utilizing the bootstrap test and syntax developed by Preacher and Hayes (2008).

To summarize the main findings, the results support partial mediation in Study 1 (high critical ratio) and Study 2 (high and medium critical ratios), but not in the low critical ratio studies. Where mediation was present, path a was also significant which indicates that cognitive reflection is negatively related to chasing. This is consistent with our prior results (Section 4), supporting H1. Further, and as expected, the analysis showed a consistently strong negative impact of demand chasing on expected profit (note the negative signed path b) for all studies, implying that if demand chasing is present, it consistently leads to lower expected profit irrespective of the critical ratio.

Specific to each critical ratio, and focusing first on Study 1, the c path is significant, which is consistent with prior results supporting H2. (This is also the comparison baseline for the direct effect after inclusion of the mediator). The c′ path is particularly interesting. The c′ path is significant ($t = 4.970$, $p < 0.001$), confirming that the CRT is related to the *ex ante* tendency to chase (path a) and is also related to *ex post* performance in its own right. This result also holds for the high critical ratio condition of Study 2, where the c′ path remains significant ($t = 3.925$, $p < 0.001$) along with paths a and c. However, the results are less clear in the medium and low critical ratio conditions. In the medium critical ratio condition, CRT is still related to demand chasing (path a) and the direct effect on expected profit (path c) is significant, which provides support for H1 and H2. However, the direct effect on expected profit when chasing is included (path c′) is no longer significant, although this may be partially due to a smaller sample size and the relatively high overall strength of chasing. These first-order effects are not significant in the low critical ratio conditions, which is consistent with our prior finding that H2 is not supported in the low critical ratio cases. In sum, at least in the high and medium critical ratio conditions, we observe that individuals with low cognitive reflection scores exhibit a stronger tendency to chase demand and have lower expected profit. However, because of the partial mediation results, other aspects of rational thinking captured by the CRT are also related to newsvendor decision outcomes.

7. Conclusions

This study connects cognitive reflection to newsvendor behavior, and is a step toward answering Bolton and Katok's (2008) call for behavioral theory with respect to order quantity misjudgments. Consistent with cognitive reflection, we believe that errors in newsvendor decision making can be at least partially attributed to System 1 versus System 2 processing behaviors: the results show that in high and medium-critical ratio newsvendor environments, individuals with higher cognitive reflection are statistically less likely to chase demand, which supports H1. In such settings, the outcomes from these individuals are closer to the normative values. In addition, in high critical ratio settings, cognitive reflection also directly impacts expected profit even when the impact of demand chasing is already considered in a mediation model. Lastly, based on the results with experienced practitioners, we investigated other possible covariates such as college major, years of experience, and managerial position. While some of these factors are modest predictors in their own right, none is better than cognitive reflection in predicting expected profit.

The fact that our empirical studies found no significant relationship between cognitive reflection and task outcome in the low critical ratio conditions suggests that other behavioral factor(s) may guide ordering behavior in these settings. Study 1 focused on a high critical ratio environment because it is frequently observed in many supply chain contexts and was consistent with the business environment of our practitioner subjects. It is difficult to find common supply chain environments where service levels are intentionally set below 50%, as even low-margin businesses often have minimum customer service expectations. Since our practitioner subjects in Study 3 also had work experience in a high service level environment, we were not entirely surprised that they were also reluctant to order below the mean demand. However, our student subjects in Study 2 had little or no work experience in high (or low) service-level business environments and behaved the same way. Based on several follow-up discussions with industry participants, we speculate that a low critical ratio setting may be more likely to activate cognitive dissonance (Festinger, 1957) because the goal of profit maximization is in direct conflict with expressed or implied goals of satisfying customer demand and avoiding lost goodwill. In effect, individuals may be including objectives other than maximizing profit. In low-margin settings, individuals may be reluctant to order below the mean demand (thus planning to stock-out more than 50% of the time), and the potential dissonance between conflicting goals may guide behavior in such settings. Further research, beyond the scope of this paper, is needed to test this conjecture.

We believe the current research has potential implications for practitioners. The results suggest that, all else being equal, it may be desirable to assign inventory analysts with higher cognitive reflection tendencies to environments where customer demand is stable but stochastic. On average, individuals with high CRT scores appear to chase less in high and medium critical ratio newsvendor environments, and are no different than those with low CRT scores in low critical ratio settings. Clearly cognitive reflection should never be the only criteria in assignment decisions and we do not discount the importance of other factors such as domain knowledge, creativity or interpersonal skills that are not assessed by the CRT. These results may also be useful in developing training programs to help mitigate suboptimal ordering behaviors, perhaps by making individuals more aware of an inclination to use easily available heuristics such as demand chasing. While clearly a subject for future research, these results may also influence the design of decision support systems that interact with human decision makers. For example, if the expected user profile is likely to have lower CRT scores, decision support software could place additional emphasis on a rolling average of demand rather than highlighting the prior period, possibly reducing unwarranted chasing.

This study has several limitations. This research is based on an experimental task environment, using a design that is similar to other newsvendor studies. The performance of practicing managers on this simplified newsvendor problem may not reflect their performance in more complex inventory decision environments where the demand distribution and cost parameters are not known with certainty. Although the experiment was empirically grounded, practicing managers and analysts may be able to perform better with specific training, experience, or the use of task-specific software that we are unable to replicate in an experiment. Finally, we have not eliminated the possibility that other individual factor(s) or measurement instrument(s) are better predictors of task outcome.

These findings also suggest a number of possible follow-up studies. In addition to further study on cognitive dissonance or other behavioral influences in low critical ratio settings, this research could also be applied to other, more complex supply chain contexts such as multi-stage inventory problems. We expect that the heterogeneity present in this simple repeated newsvendor setting will also impact outcomes in more complex settings where replenishment decisions are made across multiple periods, such as those subject to the bullwhip effect. Similarly, it may be interesting to extend this research to supply chain decision contexts with unobserved/censored demand, non-stationary or correlated demand distributions, unclear or changing costs, discounting mechanisms, or group decision making. It is possible that individuals with low CRT scores would perform better in settings where demand chasing actually improves task outcome (e.g., when demands are positively correlated over time). Further, differences in actual and experimental incentive structures, the time to make a decision, or the duration between decisions may be worth additional study. It would also be interesting to explore decision making in limited information contexts where tacit knowledge, recognition or experience may have added value.

Appendix A. Experimental setup and details

The experimental design was similar to that of prior published newsvendor experiments where respondents made order quantity decisions over several periods. The experiment asked respondents to manage the inventory at a small retailer. Respondents were asked to place one inventory order per period (i.e., week). For Studies 1 and 3 (experienced industry participants), the exercise was run for 12 weeks, while for Study 2 (business school subject pool), the exercise was for 25 weeks. Demand was drawn from a normal distribution, with demand and cost parameters as described in sections 4 and 5. 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 demand history is shown below in Fig. A1. Additional detail is available upon request.

Studies 1 and 3 (involving industry participants) took place over a two-week window at each firm. Before the experiment was closed, at least one reminder notice was sent out. For Study 1, response rates by firm were 53.8%, 50.4% and 25.1%, respectively, for an overall response rate of 36.6%. For Study 3, the response rate was 70.7%. Given the anonymous nature of the instrument in Studies 1 and 3, we looked for systematic bias in our data. We provided firm-specific details about the number of responses by department to each firm. Based on this data, senior managers at each firm indicated that participation was representative of their organization.

Appendix B. Distribution of responses

See Figs. B1–B3 .

Appendix C. Task outcome by CRT group

See Tables C1–C3.

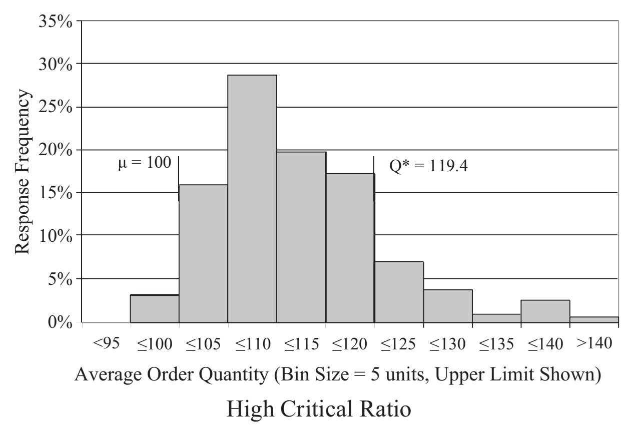


Fig. B1. Average order quantity by respondent – Study 1.

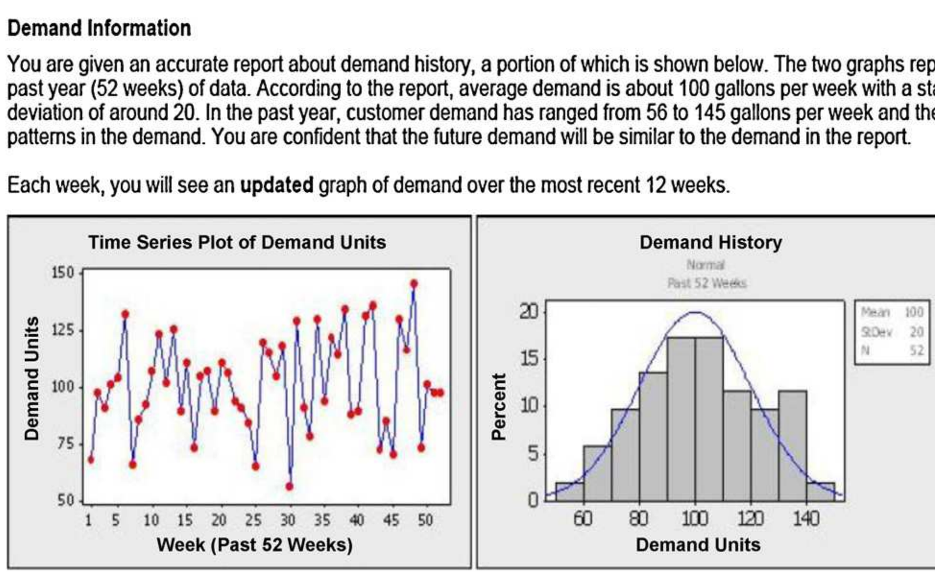


Fig. A1. Excerpt of instructions to participants.

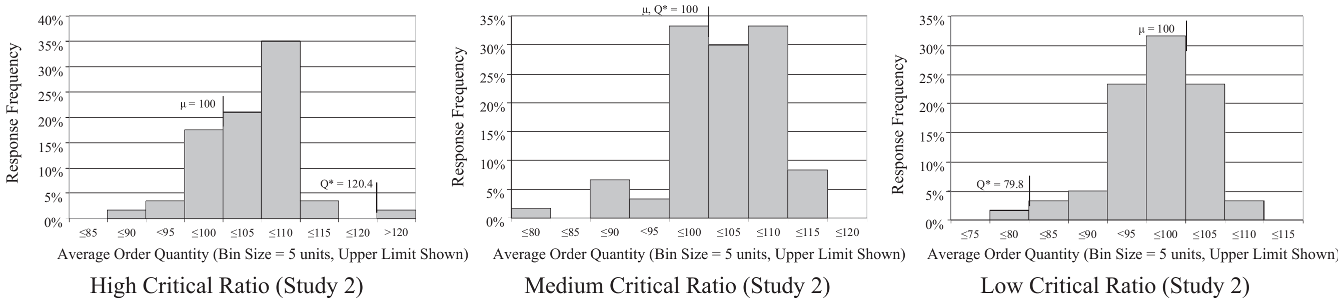


Fig. B2. Average order quantity by respondent – Study 2.

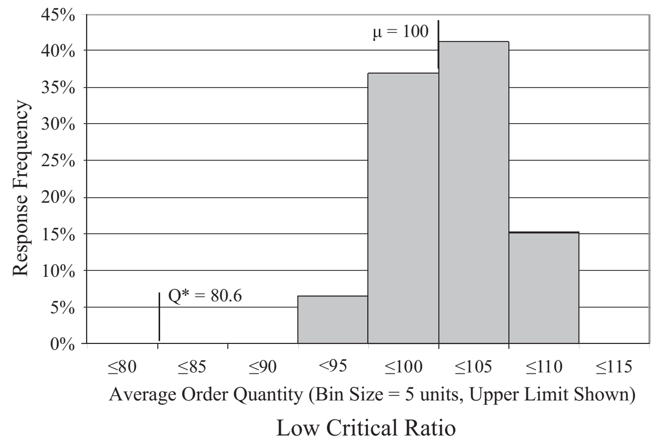


Fig. B3. Average order quantity by respondent – Study 3.

Table C1
Comparison of average expected profit by major and CRT group – Study 1.

College major	Complete sample			Average expected profit by CRT group[a]					
	n	Average expected profit[d]	S.D.	0	1	2	3	F (df, $\mathrm{df_{error}}$)	Pairwise multiple comparisons[e]
Liberal arts	15	$913.61	20.5	$890.52	–	–	–		
Business: Accounting/Finance	20	$921.11	16.0	$913.36	–	$925.85	–		
Business: Marketing/Management	142	$913.78	15.4	$909.56	$911.38	$920.94	$915.23	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	15.6	$906.33	$917.71	$923.00	$926.87	5.406 (3, 53) $p = 0.003$	$\mu_0 < \mu_1$ p = n/s, $\mu_0 < \mu_2$, μ_3 $p = 0.010$ $\mu_1 < \mu_2$, μ_3 p = n/s, $\mu_2 < \mu_3$ p = n/s
Education/Social Sciences	12	$919.63	15.6	–	–	–	–		
Engineering/Physical Sciences	36	$925.13	16.0	–	$915.52	$923.10	$929.05	2.484 (2, 34) $p = 0.100$	$\mu_1 < \mu_2$ p = n/s, $\mu_1 < \mu_3$ $p = 0.100$ $\mu_2 < \mu_3$ p = n/s
Other	24	n/a	n/a	n/a	n/a	n/a	n/a		
F (df, $\mathrm{df_{err}}$) 3.605 (6, 299), $p = 0.002$									
F (df, $\mathrm{df_{err}}$)[b] 1.709 (6, 278), $p = 0.119$				F (df, $\mathrm{df_{error}}$)[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] Average expected profit. Note: Maximum expected profit = $940.00, expected profit at Q = mean is $904.25.
[e] Pairwise multiple comparison of group means (Bonferroni).

Table C2
Comparison of average expected profit by years of experience and CRT group – Study 1.

Years of experience	Complete sample			Average expected profit by CRT group[a]					
	n	Average expected profit[d]	S.D.	0	1	2	3	F (df, df_{error})	Pairwise multiple comparisons[e]
≤1 year	21	$918.27	20.3	–	$913.49	$918.61	$928.78	1.693 (2, 16) p = n/s	
1–5 years	65	$916.45	15.0	$910.71	$912.93	$924.54	$917.62	3.120 (3, 61) p = 0.031	$\mu_0 < \mu_1$, μ_3 p = n/s, $\mu_0 < \mu_2$ $p \le 0.046$, $\mu_1 < \mu_2$ p = 0.087, $\mu_1 < \mu_3$ p = n/s, $\mu_2 < \mu_3$ p = n/s
5–15 years	107	$918.55	15.2	$908.18	$919.34	$922.39	$924.29	7.263 (3, 103) $p \le 0.001$	$\mu_0 < \mu_1$ p = 0.028, $\mu_0 < \mu_2$, μ_3 $p \le 0.001$, $\mu_1 < \mu_2$, μ_3 p = n/s, $\mu_2 < \mu_3$ p = n/s
>15 years	120	$915.10	14.8	$904.19	$912.64	$919.81	$923.79	10.156 (3, 116) $p \le 0.001$	$\mu_0 < \mu_1$ p = n/s, $\mu_0 < \mu_2$, μ_3 $p \le 0.001$, $\mu_1 < \mu_2$ p = n/s, $\mu_1 < \mu_3$ p = 0.051, $\mu_2 < \mu_3$ p = n/s
F (df, df_{err}) 0.710(3, 309), p = 0.599									
F (df, df_{err})[b] 1.038 (3, 297), p = 0.443				F(df, df_{err})[c] 6.080 (3,297) $p \le 0.001$					

[a] Only cells with $n \ge 5$ are reported.
[b] F values for years of experience with CRT included.
[c] F values for CRT with years of experience included.
[d] Average expected profit. *Note*: maximum expected profit = $940.00, expected profit at Q = mean is $904.25.
[e] Pairwise multiple comparison of group means (Bonferroni).

Table C3
Comparison of average expected profit by managerial position and CRT group – Study 1.

Managerial position	Complete sample			Average expected profit by CRT group					
	n	Average expected profit[c]	S.D.	0	1	2	3	F (df, df_{error})	Pairwise multiple comparisons[d]
Non-Managers	245	$915.75	15.8	$907.66	$913.56	$922.26	$921.01	15.076 (3, 241) $p \le 0.001$	$\mu_0 < \mu_1$ p = n/s, $\mu_0 < \mu_2$, μ_3 $p \le 0.001$, $\mu_1 < \mu_2$ p = 0.006, $\mu_1 < \mu_3$ p = 0.057, $\mu_2 < \mu_3$ p = n/s
Managers	68	$920.58	15.8	$900.89	$921.40	$919.66	$926.81	5.264 (3, 64) p = 0.003	$\mu_0 < \mu_1$, μ_2, μ_3 $p < 0.028$, $\mu_1 < \mu_2$, μ_3 p = n/s, $\mu_2 < \mu_3$ p = n/s
F (df, df_{err}) 4.989 (1, 311) p = 0.026									
F (df, df_{err})[a] 0.233 (1, 305) p = 0.705				F (df, df_{error})[b] 11.726 (3, 305) $p \le 0.001$					

We identified 68 individuals as managers based on their titles such as manager, director or vice-president, and 245 individuals as non-managers based on titles such as analyst, buyer, or planner.
[a] F values for managerial position with CRT included.
[b] F values for CRT with managerial position included.
[c] Average expected profit. *Note*: maximum expected profit = $940.00, expected profit at Q = mean is $904.25.
[d] Pairwise multiple comparison of group means (Bonferroni).

References

Baron, R.M., Kenney, D.A., 1986. The moderator–mediator variable in social psychological research: conceptual, strategic and statistical considerations. Journal of Personality and Social Psychology 51 (6), 1173–1182.
Bendoly, E., Donohue, K.L., Schultz, K.L., 2006. Behavior in operations management: assessing recent findings and revisiting old assumptions. Journal of Operations Management 24 (6), 737–752.
Bloomfield, R.J., Gino, F., Kulp, S.M., 2007. Behavioral Causes of the Bullwhip Effect in Single Supply Chains. Working Paper, Cornell University.
Bolton, G.E., Katok, E., 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., Ockenfels, A., Thonemann, U., 2008. Managers and Students as Newsvendors: How Out-of-Task Experience Matters. Working Papers in Economics No. 39, University of Cologne.
Bostian, A.A., Holt, C.A., Smith, A.M., 2008. Newsvendor "pull-to-center" effect: adaptive learning in a laboratory experiment. Manufacturing and Service Operations Management 10 (4), 590–608.
Cantor, D.E., Macdonald, J.R., 2009. Decision-making in the supply chain: examining problem solving approaches and information availability. Journal of Operations Management 27, 220–232.
Cesarini, D., Johannesson, M., Magnusson, P.K.E., Wallace, B., 2012. The behavioral genetics of behavioral anomalies. Management Science 58 (1), 21–34.
Croson, R., Donohue, K.L., 2006. Behavioral causes of the bullwhip effect and the observed value of inventory information. Management Science 52 (3), 323–336.
Doerr, K.H., Freed, T., Mitchell, T.R., Schriesheim, C.A., Zhou, X., 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.
Evans, J.St.B.T., 1984. Heuristic and analytic processes in reasoning. British Journal of Psychology 75 (4), 451–468.
Evans, J.St.B.T., 2008. Dual-process accounts of reasoning, judgment and social cognition. Annual Review of Psychology 59, 255–278.
Feng, T., Keller, L.R., Zheng, X., 2011. Decision making in the newsvendor problem: a cross-national laboratory study. Omega 39 (1), 41–50.
Festinger, L., 1957. A Theory of Cognitive Dissonance. Row-Peterson, Evanston, IL.
Fisher, M.L., Hammond, J.H., Obermeyer, W.R., Raman, A., 1994. Making supply meet demand in an uncertain world. Harvard Business Review 72 (3), 83–93.
Frederick, S., 2005. Cognitive reflection and decision making. Journal of Economic Perspectives 19 (4), 25–42.

Gavirneni, S., Xia, Y., 2009. Anchor selection and group dynamics in newsvendor decisions – a note. Decision Analysis 6 (2), 87–97.

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

Hutchinson, J.W., Kamakura, W.A., Lynch Jr., J.G., 2000. Unobserved heterogeneity as an alternative explanation for "reversal" effects in behavioral research. Journal of Consumer Research 27 (3), 324–344.

Kahneman, D., 2011. Thinking, Fast and Slow. Farrar, Straus and Giroux, New York.

Kahneman, D., Tversky, A., 1982. On the study of statistical intuitions. Cognition 11 (2), 123–141.

Kahneman, D., Frederick, S., 2002. Representativeness revisited: attribute substitution in intuitive judgment. In: Gilovich, T., Griffin, D.W., Kahneman, D. (Eds.), Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge University Press, New York.

Kremer, M., Minner, S., Van Wassenhove, L.N., 2010. Do random errors explain newsvendor behavior? Manufacturing and Service Operations Management 12 (4), 673–681.

Lau, N., Bearden, N. Newsvendor demand chasing revisited. Management Science, http://www.informs.org/Pubs/ManSci/Upcoming-Issues, in press.

Lurie, N.H., Swaminathan, J.M., 2009. Is timely information always better? The effect of feedback frequency on decision making. Organizational Behavior and Human Decision Processes 108 (2), 315–329.

Oechssler, J., Roider, S., Schmitz, P.W., 2009. Cognitive abilities and behavioral biases. Journal of Economic Behavior and Organization 72, 147–152.

Olivares, M., Terwiesch, C., Cassorla, L., 2008. Structural estimation of the newsvendor model: an application to reserving operating room time. Management Science 54 (1), 41–55.

Preacher, K.J., Hayes, A.F., 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior Research Methods 40, 879–891.

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

Schultz, K.L., Robinson, L.W., Thomas, L.J., Schultz, J. 2011. The Use of Framing in Inventory Decisions. Working Paper. University of Alberta, Canada.

Schweitzer, M.E., Cachon, G.P., 2000. Decision bias in the newsvendor problem with a known demand distribution: experimental evidence. Management Science 46 (3), 404–420.

Sloman, S.A., 1996. The empirical case for two systems of reasoning. Psychological Bulletin 119 (1), 3–22.

Stanovich, K.E., West, R.F., 2000. Individual differences in reasoning: implications for the rationality debate? Behavioral and Brain Sciences 22 (5),645–726.

Stanovich, K.E., 2011. Rationality and the Reflective Mind. Oxford University Press, New York.

Su, X., 2008. Bounded rationality in newsvendor models. Manufacturing and Service Operations Management 10 (4), 566–589.

Toplak, M.E., West, R.F., Stanovich, K.E., 2011. The cognitive reflection test as a predictor of performance on heuristics and biases tasks. Memory and Cognition 39, 1275–1289.