

ESSAYS ON BIASED NEWSVENDOR ORDERING BEHAVIOR BASED ON LABORATORY
EXPERIMENTS

JAIME ANDRÉS CASTAÑEDA ACEVEDO



Università della Svizzera italiana

Facoltà di scienze economiche, Institute of Management

Lugano

August 2013

ESSAYS ON BIASED NEWSVENDOR ORDERING BEHAVIOR BASED ON LABORATORY
EXPERIMENTS

JAIME ANDRÉS CASTAÑEDA ACEVEDO

Dissertation submitted in partial fulfillment of the requirements for the degree Dottore in Scienze
economiche

Supervised by

Prof. Paulo Gonçalves, Ph.D.



Università della Svizzera italiana

Facoltà di scienze economiche, Institute of Management

Lugano

August 2013

ABSTRACT

A behavioral approach studying inventory ordering decisions in Newsvendor settings dates back to the early 2000s. Two systematic biased behavioral patterns have been identified since: a pull-to-center effect, or the tendency to order too many costly (low-safety stock) products and too few cheap (high-safety stock) products relative to the optimal stock level; and a demand chasing bias, or the tendency to adjust inventory ordering quantities towards prior demand realizations. Through three essays, this dissertation extends behavioral research in Newsvendor settings by exploring decision making behavior in structurally similar decisions and testing two different debiasing strategies. Essay 1 develops the Innovator model, an analog to the Newsvendor model for New Product Development projects, and explores project complexity level and resource allocation decision biases. This study finds that project complexity level and resource allocation decision biases resemble Newsvendor biases. Essay 2 proposes a debiasing mechanism that builds on cognitive dissonance theory to stress differences in items' importance and safety stock levels in joint decisions as a way to debias Newsvendor ordering decisions for critical items. This study finds that joint consonant decision frameworks reduce to some extent biased Newsvendor ordering behavior, whereas joint dissonant decision frameworks increase it. Finally, essay 3 tests a Newsvendor extension that backlogs unmet demand and compares it to the traditional Newsvendor model that loses unmet demand. This study finds that backorders help achieving better inventory ordering decision making in terms of both profits and product availability relative to lost sales.

SOMMARIO

Una prospettiva comportamentale per lo studio di decisioni d'inventario in strutture *Newsvendor* risale ai primi anni del 2000. Due comportamenti distorti sistematici sono stati identificati: l'effetto *pull-to-center*, ovvero la tendenza a ordinare un numero eccessivo di prodotti costosi (*low-safety stock*) e troppo pochi di prodotti economici (*high-safety stock*) rispetto al livello d'inventario ottimale; e l'errore *demand chasing*, ovvero la tendenza ad aggiustare le decisioni d'inventario in base alla domanda dei periodi precedenti. In tre saggi, questa tesi amplia la letteratura comportamentale su strutture *Newsvendor*, esplorando il processo decisionale in strutture di decisioni simili e testando due strategie diverse di correzione degli errori. Il primo saggio sviluppa l'*Innovator model*, un analogo del *Newsvendor model* per progetti di sviluppo di nuovi prodotti, ed esplora errori nelle decisioni sul livello di complessità di un progetto e sull'allocazione delle risorse. Questo studio rivela che questi errori rispecchiano le distorsioni *Newsvendor*. Il secondo saggio propone un meccanismo di correzione degli errori che si basa sulla teoria della dissonanza cognitiva e che evidenzia differenze nell'importanza e nel livello di *safety stock* degli oggetti per decisioni congiunte d'acquisto. In questo modo possono essere ridotti gli errori *Newsvendor* per decisioni d'acquisto critiche. Questo studio rivela che decisioni comuni concordanti riducono in qualche misura il comportamento d'inventario *Newsvendor* distorto, mentre decisioni comuni dissonanti lo aumentano. In fine, il terzo saggio esplora un'estensione del *Newsvendor model* che considera la domanda non servita accumulata degli ordini arretrati e la confronta con il *Newsvendor model* tradizionale che invece non considera la domanda non servita. Questo studio rivela che gli ordini arretrati aiutano a prendere migliori decisioni d'inventario sia in termini di profitti sia di disponibilità del prodotto rispetto al *Newsvendor model* tradizionale.

NOTES ON SOFTWARE AND DOCUMENTATION

The experimental Newsvendor settings in chapters 2 and 3 were programmed and run by the author in Forio Business Simulations (www.forio.com). The experimental Newsvendor settings in chapter 4 were programmed and run by the author in z-Tree (Fischbacher, 2007). The data coming from the experiments were compiled in Microsoft Office Excel version 2007.

Bootstrap confidence intervals were computed by the author in Stata/MP version 11.2. Wilcoxon Rank-Sum and Signed-Rank tests were done manually by the author in Microsoft Excel version 2007. Regression models were run by the author in Stata/MP version 11.2.

Forio and z-Tree source codes and data files are available upon request for documentation purposes.

ACKNOWLEDGEMENTS

First of all, I want to express my gratitude to Paulo, a wonderful advisor and an even more wonderful human being. Thanks for having given me the opportunity to come to Lugano and work under your supervision. Many thanks also for all the support you gave me and the time we spent together developing ideas. Your brightness and clarity of thinking are things I aspire to have. Thanks also for all the confidence you placed on me. Last but not least, thanks for having treated me with such kindness and respect. The way you approach and treat people is remarkable and is something I strive for.

I want to thank Nitin for all his support in developing some of the ideas expressed in this dissertation. Many thanks also to Karen for her sharp comments on how to improve further the dissertation. I want to thank also both Paulo and Nitin for their advices to develop further my academic career. I also want to express my gratitude to all IMA Faculty. All the training and support I received from you helped me sharpening my intellectual skills. Many thanks also to Sebastián, Mohammad and Gloria for all their support in running some of the experiments reported in this dissertation. Finally, thanks also to all the administrative staff at USI who have helped me.

IMA is a great workplace. I want to thank all the colleagues and friends I made here. Min, Soorjith, Karthik, Pooya, Chanchal, Marco, Martina, Sayed, Alessio, Mohammad, Zoltan, Rebeka, Margarita, Sebastián, Manos, Ivona, Marco (the new one), and Stefania (not IMA, but pretty close)... Thank you for all the shared words, the support you all gave me and your kindness whenever I requested your help or advice.

I want to thank also Santiago and Claire for accommodating me when I first got here. Without your help I do not know what would have become of my Lugano experience.

Lugano... I made really good friends here whom I will never forget. I want to thank all of you for having given me the opportunity to share with you and made my last four years truly rewarding. Many thanks also for all the fun we had together here at Lugano and all over Switzerland and Europe. Those are times I will carry always with me.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
1.1. OVERVIEW	1
1.2. STRUCTURE OF THE DISSERTATION	2
CHAPTER 2: AMBITIOUS DESIGN GOALS AND STRETCHED RESOURCE ALLOCATION: INVESTIGATION OF MANAGERIAL BIASES UNDER INNOVATION UNCERTAINTY	5
2.1. INTRODUCTION	6
2.2. INNOVATION UNCERTAINTY	8
2.3. THE INNOVATOR MODEL	9
2.4. RESEARCH HYPOTHESIS AND LABORATORY EXPERIMENT	12
2.4.1. Innovators behavior	12
2.4.2. Experimental design	14
2.4.3. Experimental procedure	15
2.5. RESULTS	16
2.5.1. Pull-to-center effect	16
2.5.2. Threshold chasing bias	20
2.6. DISCUSSION	23
CHAPTER 3: IMPACT OF JOINT DECISIONS AND COGNITIVE DISSONANCE ON PREPOSITIONING (NEWSVENDOR) DECISIONS	29
3.1. INTRODUCTION	30
3.2. THEORY AND HYPOTHESES DEVELOPMENT	33
3.2.1. Prepositioning of emergency supplies as a Newsvendor problem	33
3.2.2. Newsvendor's pull-to-center effect	34
3.2.3. Joint decision making as a debiasing mechanism	36
3.3. LABORATORY EXPERIMENT	41
3.3.1. Basic design	41
3.3.2. Notation	42
3.3.3. Treatments	42
3.3.4. Experimental procedure	43
3.4. RESULTS	45
3.4.1. Newsvendor biases in inventory prepositioning decisions	45
3.4.2. Baseline results: separate decision treatments	47
3.4.3. Cognitive dissonant joint treatment's results	49

3.4.4.	Cognitive consonant joint treatment's results	52
3.4.5.	Summary of results	54
3.5.	DISCUSSION	55
	CHAPTER 4: INVENTORY ORDERING DECISIONS IN A SINGLE ECHELON: THE EFFECT OF BACKORDERS	63
4.1.	INTRODUCTION	63
4.2.	NEWSVENDOR PROBLEM WITH BACKORDERS	67
4.2.1.	Normative implications	67
4.2.2.	Behavioral implications	70
4.3.	LABORATORY EXPERIMENT	72
4.3.1.	Experimental design	73
4.3.2.	Experimental procedure	74
4.4.	RESULTS	76
4.4.1.	Normative hypotheses	76
4.4.2.	Behavioral hypotheses	79
4.4.3.	Robustness check to outliers	83
4.5.	DISCUSSION	84
	CHAPTER 5: CONCLUSIONS	91
5.1.	CONTRIBUTIONS	91
5.2.	LIMITATIONS AND FUTURE WORK	92
	BIBLIOGRAPHY	95

LIST OF FIGURES

Figure 2.1. Average complexity level and resource allocation behaviors	17
Figure 2.2. Average complexity level and resource allocation adjustment behaviors period-to-period	21
Figure A2.2.1. Sample of game screen	28
Figure 3.1. Average inventory prepositioning behaviors in separate decision treatments	49
Figure 3.2. Average inventory prepositioning behaviors in dissonance-related treatments	51
Figure 3.3. Average inventory prepositioning behaviors in consonance-related treatments	53
Figure A3.2.1. Sample of game screen	60
Figure 4.1. Optimality behavior to increasing values of the purchasing cost c	68
Figure 4.2. 95% bootstrap confidence intervals of average inventory ordering behaviors	77
Figure 4.3. 95% bootstrap confidence intervals of average order adjustment behaviors after a shortage	80
Figure A4.2.1. Sample of game screen	90

LIST OF TABLES

Table 2.1. Parallel between the Newsvendor and the Innovator models	12
Table 2.2. Treatments, notation, and number of participants	16
Table 2.3. 95% bootstrap confidence intervals of average complexity level and resource allocation behaviors	18
Table 2.4. Fixed-effects panel regression of learning to avoid the pull-to-center effect	20
Table 2.5. Wilcoxon Rank-Sum tests of threshold chasing bias	22
Table 2.6. Fixed-effects panel regression of learning to avoid the threshold chasing bias	23
Table 3.1. Treatments, notation, and number of participants	44
Table 3.2. 95% bootstrap confidence intervals of average inventory prepositioning behaviors	47
Table 3.3. Summary of hypothesis tests	55
Table 4.1. Treatments, notation, and number of participants	75
Table 4.2. Regression of <i>larger inventory system effect</i> in low-safety stock conditions	78
Table 4.3. Regression of <i>larger inventory system shortage effect</i> in low-safety stock conditions	81
Table 4.4. Fixed-effects panel regression of misperceptions of feedback	82

LIST OF APPENDICES

Appendix 2.1. Sample of written instructions (C_H)	26
Appendix 2.2. Sample of game screen (C_H)	28
Appendix 3.1. Sample of written instructions (joint decision treatment)	58
Appendix 3.2. Sample of game screen ($T6$)	60
Appendix 3.3. Statistical tests	60
Appendix 4.1. Sample of written instructions (B_H)	87
Appendix 4.2. Sample of game screen (B_H)	90

CHAPTER 1

INTRODUCTION

1.1. OVERVIEW

The Newsvendor model is an inventory ordering decision making model under demand uncertainty. The model dates back to Edgeworth (1888), who considered the amount of cash to keep at a bank as a product whose inventory should be controlled to satisfy random cash withdrawals from depositors, and to Arrow *et al.* (1951), who incorporated demand uncertainty to the study of inventory control policies. The model captures the problem a manager faces when she has to order a product that has to be sold during a season without knowing the product's demand for that season. It assumes that the item perishes before the next season and that unmet demand is lost. Accordingly, when the manager orders more than the demand, she must dispose of the remaining stock at a loss. And when the manager orders less than the demand, she loses sales opportunities. It is well-known that the solution of the problem is a base stock or order-up-to policy that balances the costs of ordering too little against the costs of ordering too much (Cachon and Terwiesch, 2009).

Despite the model's long history, a behavioral approach to study Newsvendor ordering decisions is fairly more recent, dating back to Schweitzer and Cachon (2000), who studied how individuals make Newsvendor ordering decisions in a controlled laboratory (lab) experiment. Schweitzer and Cachon's (2000) seminal study found that individuals make biased inventory ordering decisions. In particular, they ordered less than the optimum when the costs associated with the ordered item and the demand process called for larger orders (high-safety stock or high-profit setting), whereas they ordered more than the optimum when the costs associated with the ordered item and the demand process called for smaller orders (low-safety stock or low-profit setting); this bias is also known as the pull-to-center effect. In addition, the authors also found that when individuals did change inventory ordering decisions period to period, the changes tended to be in direction of the prior demand realization; this bias is also known as demand chasing.

The pull-to-center effect and demand chasing have been replicated in subsequent Newsvendor experiments. The pull-to-center effect has been proved robust to extended experience (Benzion *et al.*, 2010; Bolton *et al.*, 2012), sharpened payoff differentials (Bolton and Katok, 2008; Bostian *et al.*, 2008; Feng *et al.*, 2011), and improved outcome feedback (Bostian *et al.*, 2008; Lurie and Swaminathan, 2009), among others. Demand chasing has been proved robust to demand distribution (Benzion *et al.*, 2008), non-operations frames (Kremer *et al.*, 2010), and financial risk-taking behaviors (de Véricourt *et al.*, 2013), among others.

This dissertation contributes to the literature on Behavioral Operations Management (Bendoly *et al.*, 2006; Bendoly *et al.*, 2010; Gino and Pisano, 2008; Loch and Wu, 2007) by looking at biased Newsvendor ordering behavior from three perspectives. First, it seeks to understand if insights from the Newsvendor model and biased Newsvendor ordering behavior can be translated to structurally similar decisions, in particular to complexity level (or scope) and resource allocation decisions in New Product Development (NPD) projects under innovation uncertainty. Second, it examines if a new debiasing mechanism that builds on Festinger's (1957) cognitive dissonance theory can effectively debias and strengthen the pull-to-center effect. Finally, it builds on an existing extension of the Newsvendor model to the case of backorders (Bulinskaya, 1964) and examines whether backlogging unmet demand instead of losing sales can effectively debias inventory ordering behavior.

1.2. STRUCTURE OF THE DISSERTATION

This dissertation is composed of three essays that explore biased decision making behavior in Newsvendor settings. Each essay was written to eventually be sent for publication. Hence, there is some repetition of the Newsvendor setting and some literature review in each essay. However, each essay addresses a different aspect of biased Newsvendor decision making behavior. The ensuing three chapters present the three essays with their corresponding findings. Lastly, a final chapter presents concluding remarks and discusses limitations and opportunities for future research. A brief description of the three essays and their findings is given below.

In the first essay (chapter 2), we develop a stylized model of NPD decision making under innovation uncertainty. Assuming a single stage-gate innovation pipeline under a single uncertainty source, the developed model is analogous to the Newsvendor model. We operationalize the model under project complexity level and resource uncertainty separately and test it in a lab experiment. We find that project complexity level and resource allocation biases resembled those observed in Newsvendor experiments. In particular, we observe the pull-to-center effect, i.e., individuals tend to under react when innovation costs and uncertainty call for ambitious scopes or more resources, whereas they tend to overreact when innovation costs and uncertainty call for less ambitious scopes or fewer resources. In addition, we also observe the threshold (demand) chasing bias, i.e., individuals tend to chase uncertainty thresholds realized in previous innovation efforts. These results suggest that NPD managers may under perform in demanding markets, limiting their a priori likelihood of success; and over perform in less challenging markets, a priori dedicating more resources than those required for success.

In the second essay (chapter 3), we study joint decision making as a potential debiasing mechanism for the pull-to-center effect. In particular, we join or bundle to items that differ in their perceived importance and safety stock condition. Building on cognitive dissonance (Festinger, 1957; Simon *et al.*, 1995) arguments, we pose that bundling a high-importance high-safety stock item with a low-importance low-safety stock item (consonance) reduces the bias for the high-importance item. Alternatively, we pose that bundling a high-importance low-safety stock item with a low-importance high-safety stock item (dissonance) increases the bias for the high-importance item. We test this new debiasing mechanism in a lab experiment in which we compare joint inventory ordering decisions to corresponding baseline inventory ordering decisions (no bundling). We find support for our predictions, suggesting that joint consonant and dissonant decision frameworks may help achieving higher product availability (or customer service satisfaction) and/or profits for critical items.

Finally, in the third essay (chapter 4), I test behaviorally Bulinskaya's (1964) Newsvendor extension to the case of backorders and compare it to the traditional lost sales Newsvendor model. Consistent with a theoretical comparison of both inventory systems, I find that backorders drive

individuals' inventory ordering quantities upwards compared to lost sales. In addition, consistent with reference dependence and misperceptions of feedback, I also find that individuals react to shortages in a stronger manner when unmet demand is backlogged than when is lost and underweight backorders when making inventory ordering decisions, respectively. These results suggest that suppliers may benefit in terms of product availability (or customer service satisfaction) and/or profits by backlogging rather than losing unmet demand.

CHAPTER 2

AMBITIOUS DESIGN GOALS AND STRETCHED RESOURCE ALLOCATION: INVESTIGATION OF MANAGERIAL BIASES UNDER INNOVATION UNCERTAINTY

(with Paulo Gonçalves and Nitin Joglekar)

ABSTRACT

We develop a stylized decision making model to inform decision making in New Product Development (NPD) under innovation uncertainty. The model incorporates the possibility of setting ambitious (or, alternatively, less aggressive) design goals or stretched (or, alternatively, less restricted) resource allocations. Delivering on these goals enhances market payback, but also creates product launch risks. Our setup is analogous to the Newsvendor model for ordering inventory. We operationalize and test the model experimentally under complexity and resource uncertainty separately. Results show that decision making biases resemble those observed in Newsvendor settings. On the one hand, we observe the pull-to-center effect; that is, individuals tend to under react when innovation costs and uncertainty call for either more resources or ambitious scopes, and they tend to overreact when innovation costs and uncertainty call for either fewer resources or ambitious scopes. On the other hand, we also observe the threshold chasing bias; that is, individuals tend to chase uncertainty thresholds realized in previous innovation efforts. Findings suggest that NPD managers may underperform in demanding markets, limiting their a priori likelihood of success; and overperform in less challenging markets, a priori dedicating more resources than those required for success.

Keywords: Behavioral Operations Management, Laboratory Experiments, Innovation Uncertainty, New Product Development, Newsvendor Biases, Newsvendor Model.

2.1. INTRODUCTION

New Product Development projects face significant uncertainty impacting their success in the market place. Managing uncertainty to reduce the risk of project failure is a key challenge faced by NPD managers (Cooper, 2003). Uncertainty sources are diverse and include, among others: customer, technological, market, and resource uncertainty (Cooper, 2003; Krishnan and Ulrich, 2001; Moenaert and Souder, 1990; Mullins and Sutherland, 1998; Thomke, 2008). To the extent that information about the uncertainty sources is available and adequate —there is a list of possible events, their probabilities, and their impact on project payoff—, NPD managers can rely on traditional project management tools such as task scheduling and risk management to address them (Pich *et al.*, 2002). Activity network techniques such as Critical Path Method (CPM) and Program Evaluation and Review Technique (PERT) have been widely used for decades for project planning and project management. Risk management, scenario-planning and simulation extend activity network techniques, by identifying possible but uncertain events and planning for them.

However, information about the uncertainty sources is frequently incomplete or inaccurate. NPD managers are not aware of all possible unanticipated events (Schrader *et al.*, 1993) and do not fully understand the impact that their decisions may have on project performance (Pich *et al.*, 2002). Accordingly, empirical research suggests a number of approaches to manage projects in highly uncertain environments. For instance, research shows that iterative prototyping and testing can help NPD projects progress towards acceptable results (Lynn *et al.*, 1996; Thomke, 1998). Alternatively, managers can work on parallel trials to develop multiple solutions, choosing the best one once their outcomes are observable (Beinhocker, 1999; Sobek *et al.*, 1999).

Similarly, analytical approaches suggest how to model information about the uncertainty sources in order to optimize a given project outcome. For instance, Pich *et al.* (2002) consider a general model that maps a network of activities and a set of influence factors to a project payoff. The uncertain nature of the set of influence factors and the complexity of the map structure make necessary to consider policies —instead of a pre-specified network of activities— that identify in advance a complete set of

actions which are triggered by signals. Under this approach, a policy is then optimal if it maximizes the conditional expected payoff given a signal. Following decision sciences, Loch and Terwiesch (2005) consider a model of preliminary information in which information is treated in the form of probabilities and outcomes. The model also considers events or aggregated set of outcomes and actions that the information receiving party takes. Knowing that earlier actions incur lower costs, but are associated with higher uncertainty, and later actions can use more information, but are more costly, the manager's problem is then to time actions in the presence of uncertainty so as to minimize expected costs.

Although these approaches seek to inform decision making under innovation uncertainty, they are too complex and hence derivation of optimal policies is extremely difficult, rendering difficult to quantitatively support managerial decisions (Loch and Terwiesch, 2005; Pich *et al.*, 2002). In addition, these approaches have not been tested behaviorally to the best of our knowledge and hence a clear understanding of how managers behave under innovation uncertainty is lacking (Loch and Wu, 2007).

Addressing such gaps, this work makes two contributions. First, it adapts the seminal Newsvendor model for perishable inventory to NPD decision making under innovation uncertainty. By drawing a parallel between NPD decision making under innovation uncertainty and Newsvendor decision making, we derive a foundational model, which we name the Innovator model, to suitably inform decision making in NPD under innovation uncertainty. Our simple model provides a normative solution for an NPD manager deciding on the complexity level (or scope) of a project and on its resource allocation under complexity and resource uncertainty, respectively. Second, it explores decision biases NPD managers may be prone to. Previous research on judgment and decision theory has shown that individuals are prone to a number of decision biases (Kahneman, 2003; Tversky and Kahneman, 1974). Cooper (2003) recognizes that NPD managers face similar challenges to those faced by other individuals, which suggests that their decisions may deviate from optimum. By running decision experiments applying the Innovator model under either complexity or resource uncertainty, we show that NPD managers may be prone to the well-known pull-to-center effect and demand (threshold) chasing bias commonly observed in behavioral studies of the Newsvendor model.

The rest of the paper is organized as follows: Section 2.2 provides an explanation of innovation uncertainty, emphasizing endogenous uncertainty sources. Section 2.3 draws an analogy between decisions in NPD under innovation uncertainty and Newsvendor settings, developing an analog to the Newsvendor model, which we name the Innovator model. Based on behavioral studies of the Newsvendor model, section 2.4 develops a research hypothesis and presents the laboratory (lab) experiment designed to test the Innovator model under resource and complexity uncertainty separately. Section 2.5 presents the main results and hypothesis tests. Finally, section 2.6 summarizes the work and discusses the main findings, implications, limitations, and opportunities for future research.

2.2. INNOVATION UNCERTAINTY

Uncertainty is a prevalent issue in innovation. The NPD literature highlights some common sources of uncertainty: customer, technology, market, and resource uncertainty (Cooper, 2003; Krishnan and Ulrich, 2001; Moenaert and Souder, 1990; Mullins and Sutherland, 1998; Thomke, 2008). Customer uncertainty is related with the inability of customers to fully specify all of their needs because they either face uncertainty themselves or cannot articulate their needs on products that do not yet exist (Thomke, 2008). Technological uncertainty arises because there is a lack of knowledge about the availability and performance of new technology (Cooper, 2003). Market uncertainty is related with the absence of information about the market opportunities that a new technology offers (Mullins and Sutherland, 1998). And resource uncertainty is related with the absence of information about the human, financial, and technical resources needed to create the innovation (Cooper, 2003; Moenaert and Souder, 1990; Mullins and Sutherland, 1998).

Customer uncertainty, technological uncertainty, and market uncertainty all originate from the external environment (Moenaert and Souder, 1990). Due to their exogenous nature, these sources of uncertainty are harder for NPD managers to control. In contrast, resource uncertainty originates internally and hence it is arguably easier for NPD managers to measure and control (Jauch and Kraft, 1986). This is so even though resource uncertainty may be impacted by external sources of uncertainty. For instance, while significant market uncertainty may limit the amount of resources NPD

managers may choose to allocate in a specific market (Thomke, 2008), they still retain control over the decision. Hence, the Resource Innovator model deals initially with resource uncertainty.

In addition, complexity has been identified as a central contributor to project uncertainty (Hobday, 1998; Pich *et al.*, 2002; Tatikonda and Rosenthal, 2000) and associated with the reasons why NPD projects are often late, over budget, or lacking scope (Kim and Wilemon, 2003; Tatikonda and Rosenthal, 2000). Kim and Wilemon (2003) summarize the different definitions of complexity provided in the NPD literature, defining it as the challenges posed by the different number of technologies/components/functions in development efforts and the nature of organizational tasks that individuals and organizations face in carrying out NPD programs. From this definition, one can arguably infer that complexity arises from within the organization, both from the characteristics of the product being developed (Griffin, 1997; Murmann, 1994) and from the different number of tasks that need to be carried out to develop the product (Tatikonda and Rosenthal, 2000). Because of its endogenous nature, complexity may be easier for NPD managers to control even though it may be exacerbated by external sources of uncertainty. For instance, significant customer uncertainty can lead to constantly changing product specifications, making it difficult for managers to assess the desired product functionality and total number of development tasks. Still, NPD managers retain control over the decision potentially addressing a subset of customer requirements. Hence, the Complexity Innovator model deals initially with complexity uncertainty.

2.3. THE INNOVATOR MODEL

In the Newsvendor model (Arrow *et al.*, 1951), a manager places an order quantity q at unit cost w facing an uncertain demand D in a single selling season. Once D is realized, the manager sells each unit at price $p > w$. If q exceeds D , then D units are sold and $q - D$ units can be salvaged for $s < w$. That is, there is a unit overage cost $c_o = w - s$. If D exceeds q , then q units are sold and the potential profit from selling $D - q$ units is forgone. That is, there is a unit underage cost $c_u = p - w$. For simplicity, and following previous Newsvendor experiments, we assume no salvage value. That is, $c_o = w$. In sum, the order quantity q results in a realized period profit:

$$\pi(q, D) = (p - w)D - C(q, D) \quad (2.1)$$

where

$$C(q, D) = c_o(q - D)^+ + c_u(D - q)^+ \quad (2.2)$$

is the mismatch cost when ordering q . The normative solution that minimizes mismatch cost equals the maximizer of profits. If D is a random variable with pdf f , the expected mismatch cost can be expressed as a function of the order quantity q :

$$E[C(q, D)] = c_o \int_0^q (q - x)f(x)dx + c_u \int_q^\infty (x - q)f(x)dx \quad (2.3)$$

It is well-known that the normative solution Q^* that minimizes expected mismatch cost is an order-up-to or base-stock policy that balances overage and underage costs, which is characterized by the following expression:

$$F(Q^*) = \frac{c_u}{c_u + c_o} \quad (2.4)$$

where F is D 's cdf. Schweitzer and Cachon (2000) define a product as a high-profit (or high-safety stock) product when $F(Q^*) \geq 1/2$ and as a low-profit (or low-safety stock) product otherwise.

Summarizing, the Newsvendor model is characterized by three main components—a single decision (q), an uncertain parameter (D), and a cost structure (c_o and c_u)—and a relatively simple optimal policy that balances overage and underage costs given a distribution of the uncertain parameter.

Assuming a single stage-gate innovation pipeline, the proposed Innovator model for NPD decision making under innovation uncertainty is characterized by the same structure. In NPD under complexity uncertainty, a manager decides on the complexity level C or scope of a project (e.g. the functionality

of a software program) before observing the functionality threshold C_T the project must meet to be successful in the market (e.g. functionality the software must have at time of launch). Analogously, in NPD under resource uncertainty, a manager decides on the amount of resources R to allocate in a project (e.g. the number of engineer-hours with a fixed productivity in terms of tasks/hour) before knowing the total number of tasks R_T the project will require before launch (e.g. planned work and unplanned rework to program all the functionality the software requires).

Letting p be the unit revenue associated with a project and w the associated unit development cost, the cost structure of the Innovator model is analogous to that of the Newsvendor model. In the Complexity Innovator model, and following the Newsvendor logic, the project cost depends only on the decided complexity level of the project. That is, project cost is independent of the functionality threshold required for product success, and equals wC . A unit overage cost $c_o = w$ is then incurred whenever the decided complexity level C is more than the functionality threshold C_T required for product success (e.g. the software program has more functionality than required). That is, the organization builds more functionality than required for product success, incurring unnecessary costs for the extra work $C - C_T$. In contrast, and again following the Newsvendor logic, project revenue depends on whether the project is successful. That is, project revenue depends on whether the project meets the functionality threshold, and equals $(p - w)C$ if $C \leq C_T$ and $(p - w)C_T$ if $C > C_T$. A unit underage cost $c_u = p - w$ is then incurred whenever C is less than the C_T required for product success. That is, not all functionality is built and the product is launched lacking functionality, preventing the organization from realizing potential profits from $C_T - C$.

Similarly, in the Resource Innovator model, the project cost depends only on the amount of resources allocated or the number of tasks completed. That is, project cost is independent of the amount of resources necessary to complete all tasks the project will require before launch, and equals wR . A unit overage cost $c_o = w$ is then incurred whenever the allocated resources R are more than enough to complete all required tasks R_T (e.g. allocating 100 engineer-hours when 90 engineer-hours are enough to program all the software functionalities). That is, the organization allocates more resources than required to complete all tasks, incurring unnecessary cost for the extra resources $R - R_T$.

In contrast, and again following the Newsvendor logic, project revenue depends on whether all required tasks are completed. That is, project revenue depends on whether all tasks the project will require before launch are completed, and equals $(p - w)R$ if $R \leq R_T$ and $(p - w)R_T$ if $R > R_T$. A unit underage cost $c_u = p - w$ is then incurred whenever R is not enough to complete R_T . That is, not all tasks are completed and the product is launched with defects, preventing the organization from realizing potential profits from $R_T - R$.

Assuming that NPD managers decide on C or R under the assumed c_u and c_o and NPD organizations collect information about past NPD efforts to learn about the previous uncertainty sources C_T or R_T (McCarthy *et al.*, 2006), the Newsvendor structure in (2.1)-(2.4) can inform complexity level or resource allocation decisions under innovation uncertainty, providing managers with a normative complexity level or resource allocation given by (2.4).

Table 2.1 summarizes the parallel between the Newsvendor model and the proposed Complexity and Resource Innovator models.

Table 2.1. Parallel between the Newsvendor and the Innovator models.

	Newsvendor model	Complexity Innovator model	Resource Innovator model
Decision	q	C	R
Uncertain parameter	D	C_T	R_T
Cost structure	c_o, c_u	c_o, c_u	c_o, c_u

2.4. RESEARCH HYPOTHESIS AND LABORATORY EXPERIMENT

2.4.1. Innovators behavior

Lab experiments on the Newsvendor model have mainly reported two decision biases: level bias and adjustment bias. Level bias refers to individuals' average tendency to order away from the normative order quantity (Rudi and Drake, 2011). It is commonly reported in terms of the *pull-to-center* effect, which refers to the average tendency of individuals to order between the normative solution and the

mean demand (Bostian *et al.*, 2008). That is, individuals tend to order too few of high-profit products and too many of low-profit products (Schweitzer and Cachon, 2000).

In a seminal behavioral study of the Newsvendor model, Schweitzer and Cachon (2000) found evidence of the pull-to-center effect in both high- and low-profit products. A number of other studies have provided further support for the effect, showing it is robust to the demand distribution (Benzion *et al.*, 2008, 2010), sharpened payoff differentials addressing flat maximum concerns and their impediments to learning (Bolton and Katok, 2008; Bostian *et al.*, 2008), and improved outcome feedback (Bolton and Katok, 2008; Bostian *et al.*, 2008; Lurie and Swaminathan, 2009), among others.

Given the structural similarity between the Newsvendor and both Complexity and Resource Innovators, it is then reasonable to expect complexity level and resource allocation behaviors consistent with the pull-to-center effect. This leads to the first hypothesis:

HYPOTHESIS 1: Complexity level decisions will fall between the mean functionality threshold and the optimal complexity level in both profit conditions. Similarly, resource allocation decisions will fall between the mean number of tasks that need to be completed before launch and the optimal resource allocation in both profit conditions.

In addition, we do not have any reason to expect differences in complexity level and resource allocation behaviors in the same profit condition given the structural similarity between the Complexity and Resource Innovators in the same profit condition. This leads to the second hypothesis:

HYPOTHESIS 2: In the same profit condition, complexity level behavior will be similar to resource allocation behavior.

Adjustment bias refers to individuals' average tendency to adjust order quantities period-to-period (Rudi and Drake, 2011). It is commonly reported in terms of the *demand (threshold) chasing* bias, which refers to the average tendency of individuals to adjust orders towards the prior demand realization (Schweitzer and Cachon, 2000). That is, when individuals adjust orders period-to-period, they tend to do so more frequently towards than away from prior threshold realizations.

In the seminal behavioral study of the Newsvendor model, Schweitzer and Cachon (2000) found also evidence of threshold chasing behavior in both high- and low-profit products. Threshold chasing behavior has been less studied than the pull-to-center effect; however, subsequent studies have also provided further support for the bias, showing it is robust to the demand distribution (Benzion *et al.*, 2008), non-operations frames (Kremer *et al.*, 2010), and financial risk-taking behaviors (de Véricourt *et al.*, 2013), among others.

Given the structural similarity between the Newsvendor and both Complexity and Resource Innovators, it is then reasonable to expect complexity level and resource allocation adjustment behaviors consistent with the threshold chasing bias. This leads to the third hypothesis:

HYPOTHESIS 3: Complexity level adjustments will be directed more frequently towards than away from prior functionality thresholds in both profit conditions. Similarly, Resource allocation adjustments will be directed more frequently towards than away from prior number of tasks that need to be completed before launch in both profit conditions.

Unlike the pull-to-center effect, the threshold chasing bias does not have corresponding threshold chasing regions. Hence, and despite the structural similarity between the Complexity and Resource Innovators in the same profit condition, we do not make claims about the similarity in complexity level and resource allocation adjustment behaviors in the same profit condition.

2.4.2. Experimental design

Following previous Newsvendor experiments (Rudi and Drake, 2011; Schweitzer and Cachon, 2000), we set unit project revenue at $p = 12$ and manipulate unit development cost c . In particular, we set unit development cost for high-profit projects at $c = 3$, and for low-profit projects at $c = 9$. Following also previous Newsvendor experiments (e.g., Bolton and Katok, 2008; Schweitzer and Cachon, 2000), we consider uniformly distributed functionality thresholds $C_T \sim U(0, 100)$ and number of task to be completed $R_T \sim U(0, 100)$. The distributions imply a mean or $E[C_T] = E[R_T] = 50$. All individuals

experienced realizations from the same set of threshold values, controlling for the impact of threshold realizations on decision making behavior.

In the Resource Innovator individuals decide on the amount of resources to allocate, and they receive feedback on number of tasks. In this setting, the *productivity of resources* measures the number of tasks that can be completed per resource. For simplicity, we assume that the productivity of resources equals 1 task per resource. Hence, there is a straightforward conversion between allocated resources and number of tasks. For example, if an individual allocates *50 resource units*, the number of tasks completed is $1 \text{ task/resource} \cdot 50 \text{ resource units} = 50 \text{ tasks}$.

The described parameterization implies a normative complexity level and allocation of resources of 75 percent complexity level and engineer-hours in the high-profit Innovators, respectively, and a normative complexity level and allocation of resources of 25 percent complexity level and engineer-hours in the low-profit Innovators, respectively.

To explore Innovators biases, the experiment hence considers a 2×2 full factorial between-subjects design. The factors are Innovator setting, viz Complexity Innovator and Resource Innovator (C, R), and profit condition, viz high and low (H, L). Notation-wise, X_i , with $X \in (C, R)$, refers to Complexity Innovator (C) or Resources Innovator (R), where $i \in (H, L)$ refers to a high-profit (H) or a low-profit (L) condition. For example, C_H refers to the high-profit Complexity Innovator, whereas R_L to the low-profit Resources Innovator. Thus, the experiment considers four treatments:

T1: high-profit Complexity Innovator (C_H)

T2: low-profit Complexity Innovator (C_L)

T3: high-profit Resources Innovator (R_H)

T4: low-profit Resources Innovator (R_L)

2.4.3. Experimental procedure

A total of 55 individuals participated in the experiment. All participants were undergraduate and graduate students from management-related disciplines enrolled in US and Swiss universities. The

experiment was programmed and run with Forio Business Simulations (www.forio.com). Table 2.2 shows the treatments with their corresponding notation and number of participants.

Table 2.2. Treatments, notation, and number of participants.

		<i>Profit condition</i>	
		High	Low
<i>Innovator setting</i>	Complexity	$T1 (C_H)$ 14	$T2 (C_L)$ 14
	Resources	$T3 (R_H)$ 14	$T4 (R_L)$ 13

Participants arrived and were given the instructions (see Appendix 2.1). Participants had time to ask clarifying questions before initiating the experiment. After having read the instructions and answered any clarifying questions, an assistant initiated the experiment. Participants made 50 consecutive decisions for a single project each decision round aiming at maximizing cumulative profits. Each decision round began with the participant deciding on the complexity level or resource allocation, after which the C_T or R_T realization and the corresponding realized profits were revealed. Information on unit project revenue p and unit development cost c was available in the decision screen at any time. Participants had also access to historical information about the outcomes in previous projects, including previous decisions, uncertainty factor realizations, profits, and total cumulative profits (Appendix 2.2 shows a snapshot of the game screen). Monetary rewards were not used to incentivize participants.

2.5. RESULTS

2.5.1. Pull-to-center effect

Before showing the formal hypothesis tests, we first show an overview of the average complexity level and resource allocation behaviors. For instance, average complexity level behavior in C_H is given by averaging *average complexity levels across rounds for each participant* across the number of participants in C_H . Figure 2.1 suggests that average complexity level and resource allocation behaviors

in the high-profit Innovator settings exhibit the pull-to-center, whereas low-profit Innovator settings exhibit a stronger pull-to-center effect. In addition, Figure 2.1 also suggests there are no differences between Innovator settings in the same profit condition. Following, we present the formal hypothesis tests.

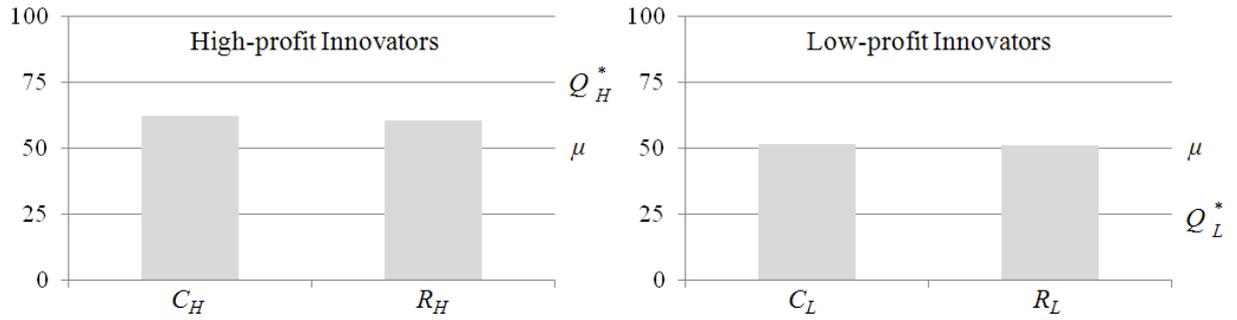


Figure 2.1. Average complexity level and resource allocation behaviors.

We first formally test whether complexity level and resource allocation behaviors exhibit the pull-to-center effect in both profit conditions (Hypothesis 1). Table 2.3 shows 95% bootstrap confidence intervals around the population of participants' average decisions for all Innovator settings.¹ Table 2.3 shows that the results of the high-profit Innovator settings are consistent with the pull-to-center effect since average complexity level and resource allocation behaviors are contained within the high-profit pull-to-center region. Table 2.3 also shows that the results of the low-profit Innovator settings show an asymmetric pull-to-center effect since average complexity level and resource allocation behaviors are not fully contained within the low-profit pull-to-center region. Such an asymmetry is common in low-profit conditions (Bolton and Katok, 2008; Bostian *et al.*, 2008; Schweitzer and Cachon, 2000). In addition, a Wilcoxon Rank-Sum test comparing average deviations from the optimum between high- and low-profit Innovator settings shows that the effect is in fact stronger in low-profit conditions ($W =$

¹ We report bootstrap confidence intervals since the sample sizes do not guarantee that the samples conform to the assumptions needed to report standard confidence intervals.

902, $z = 2.46$, $p\text{-value}_{1\text{ tail}} = 0.0070$, $r = 0.33$).² Hence, we observe average complexity level and resource allocation behaviors consistent with the pull-to-center effect in high-profit Innovator settings and with an asymmetric pull-to-center effect in low-profit Innovator settings, providing support for Hypothesis 1.

Table 2.3. 95% bootstrap confidence intervals of average complexity level and resource allocation behaviors.

	\bar{x}	<i>SE</i>	<i>CI</i>	<i>Pull-to-center region</i>
C_H	62.36	3.25	[55.64 68.16]	[50 75]
R_H	60.40	2.96	[55.35 66.96]	
C_L	51.57	3.97	[44.51 59.29]	[25 50]
R_L	51.35	2.72	[46.66 57.44]	

We next test whether complexity level and resource allocation behaviors are similar in the same profit condition (Hypothesis 2). We compare populations of participants' average decisions between Innovator settings in the same profit condition. Wilcoxon Rank-Sum tests show that we cannot statistically rule out the possibility that average complexity level and resource allocation behaviors are similar in both high- ($W = 190$, $z = -0.60$, $p\text{-value}_{2\text{ tails}} = 0.5503$, $r = -0.11$) and low-profit Innovator settings ($W = 186$, $z = 0.19$, $p\text{-value}_{2\text{ tails}} = 0.8461$, $r = 0.04$).³ Because such similarities are stated as null hypotheses, the best we can do is to fail to reject them. These results do not automatically allow us to accept the similarities since we could be making a type II error. However, observing the effect sizes —denoted as r — provides us with an objective measure of the importance of the effects (Field, 2009). The effect sizes $r = -0.11$ and $r = 0.04$ have a low-to-middle and a low importance in absolute

² We report a Wilcoxon Rank-Sum tests since the pooled sample sizes (28 and 27 observations for $C_H + R_H$ and $C_L + R_L$, respectively) do not guarantee that the samples conform to the assumptions needed to report an unpaired t -test. However, a non-reported unpaired t -test shows qualitatively the same results.

³ We report Wilcoxon Rank-Sum tests since the sample sizes do not guarantee that the samples conform to the assumptions needed to report unpaired t -tests.

value since they are below the cut-off values 0.30 and 0.10, respectively (Field, 2009). These results suggest that complexity level and resource allocation behaviors are similar in the same profit condition. Hence, the results are fairly consistent with Hypothesis 2.

Learning

Previous Newsvendor experiments have shown mixed evidence regarding learning to avoid the pull-to-center effect (Benzion *et al.*, 2010; Bolton *et al.*, 2012), yet we also test for it. We compare the population of participants' average deviations from the optimum in the first 10 rounds to that of the last 10 rounds within each Innovator setting. Wilcoxon Rank-Sum tests show there is significant evidence of learning to avoid the pull-to-center effect in C_L ($W = 167$, $z = -1.65$, $p\text{-value}_{1\text{ tail}} = 0.0491$, $r = -0.31$), whereas no evidence in the remaining Innovator settings (C_H : $W = 195.5$, $z = -0.34$, $p\text{-value}_{1\text{ tail}} = 0.3652$, $r = -0.07$ — R_H : $W = 187.5$, $z = -0.71$, $p\text{-value}_{1\text{ tail}} = 0.2382$, $r = -0.13$ — R_L : $W = 162.5$, $z = -0.67$, $p\text{-value}_{1\text{ tail}} = 0.2525$, $r = -0.13$).

Within each Innovator setting, we also run a fix-effects panel regression model of the form:

$$|C_{H,i,t+1} - C_H^*| = \beta_0 + \beta_1 t + \beta_2 \text{Over}_{i,t} + \beta_3 \text{Under}_{i,t} + v_i + \varepsilon_{i,t}, \quad t = 1, \dots, 49 \quad (2.5)$$

where the dependent variable captures participants' tendency to get closer to the optimal decision over time, t refers to round, $\text{Over}_{i,t}$ and $\text{Under}_{i,t}$ refer to the amount of over- and under-functionality of participant i in round t , respectively, and serve as a control for threshold chasing effects, v_i is the participants' effect, and $\varepsilon_{i,t}$ is the error term. Evidence of learning is provided by a significant and negative round coefficient. Table 2.4 shows highly significant evidence of learning to avoid the pull-to-center effect in C_L , whereas significant evidence of learning in both R_H and R_L .

Taken together, we observe evidence of learning to avoid the pull-to-center effect in C_L , whereas poor evidence in both R_H and R_L , which is consistent with the mixed evidence presented in previous Newsvendor experiments (Benzion *et al.*, 2010; Bolton *et al.*, 2012).

Table 2.4. Fixed-effects panel regression of learning to avoid the pull-to-center effect.^a

	C_H	R_H	C_L	R_L
β_0 (Constant)	22.12* (1.3483)	18.92* (1.6753)	31.56* (1.5253)	36.25* (1.829)
β_1 (Round)	-0.0403 (0.0381)	0.1004 [†] (0.0475)	-0.1147* (0.0434)	0.1101 [†] (0.0523)
β_2 (Over)	0.0340 (0.0244)	-0.0394 (0.0301)	0.0487 (0.0328)	0.0608 [‡] (0.0335)
β_3 (Under)	-0.0902 [†] (0.0382)	0.2103* (0.0397)	-0.0313 (0.0346)	-0.1727* (0.0432)
R^2	0.2459	0.1971	0.4042	0.2376
F	4.99	17.30	4.43	12.55
p -value	0.0020	0.0000	0.0043	0.0000

^a Standard errors in parentheses.

* Highly significant, [†] Significant, [‡] Marginally significant.

2.5.2. Threshold chasing bias

Before showing the formal hypothesis test, we first show an overview of the average adjustment behavior period-to-period. For instance, average adjustment behavior towards prior functionality thresholds in C_H is given by averaging *the number of complexity level adjustments towards the prior functionality threshold for each participant* across the number of participants in C_H . Regardless of the profit condition, Figure 2.2 suggests that when participants did change decisions period-to-period, they did it more frequently towards than away from prior functionality thresholds and number of tasks that need to be completed before launch in the Complexity and Resource Innovators, respectively.

Following, we present the formal hypothesis test.

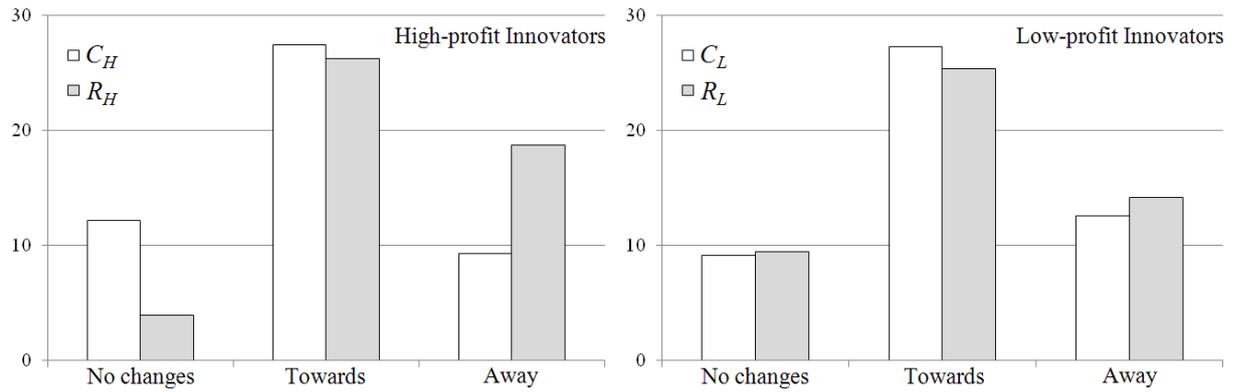


Figure 2.2. Average complexity level and resource allocation adjustment behaviors period-to-period.

We now test whether complexity level and resource allocation adjustment behaviors period-to-period are consistent with the threshold chasing bias in both profit conditions (Hypothesis 3). We compare the population of participants' number of adjustments towards prior threshold realizations to that of adjustments away from them within each Innovator setting. Wilcoxon Rank-Sum tests in Table 2.5 show that the average number of adjustments towards prior threshold realizations is significantly larger than the average number of adjustments away from them in all Innovator settings, providing support for Hypothesis 3.⁴

⁴ We report Wilcoxon Rank-Sum tests since the sample sizes do not guarantee that the samples conform to the assumptions needed to report unpaired t -tests.

Table 2.5. Wilcoxon Rank-Sum tests of threshold chasing bias.

	\bar{x}	W	z	$p\text{-value}_{1\text{ tail}}$	r	<i>Support</i>
C_H^{Tw}	27.5	117.5	-3.93	0.0000	-0.74	High
C_H^{Aw}	9.3					
R_H^{Tw}	26.3	144	-2.71	0.0034	-0.51	High
R_H^{Aw}	18.8					
C_L^{Tw}	27.3	123.5	-3.65	0.0001	-0.69	High
C_L^{Aw}	12.6					
R_L^{Tw}	25.4	129.5	-2.36	0.0092	-0.46	High
R_L^{Aw}	14.2					

Learning

Newsvendor experiments have tested learning to avoid the threshold chasing bias to a lesser extent than learning to avoid the pull-to-center effect. Nevertheless, previous Newsvendor experiments suggest that individuals show a tendency to avoid the threshold chasing bias over time (Benzion *et al.*, 2008; Schweitzer and Cachon, 2000). Accordingly, we compare the population of participants' number of adjustments towards prior threshold realizations in the first 10 rounds to that of the last 10 rounds within each Innovator setting. Wilcoxon Rank-Sum tests show there is significant evidence of learning to avoid the threshold chasing bias in both C_H and C_L (C_H : $W = 162.5$, $z = -1.86$, $p\text{-value}_{1\text{ tail}} = 0.0314$, $r = -0.35$ — C_L : $W = 160.5$, $z = -1.95$, $p\text{-value}_{1\text{ tail}} = 0.0254$, $r = -0.37$), marginal evidence in R_H ($W = 174.5$, $z = -1.31$, $p\text{-value}_{1\text{ tail}} = 0.0952$, $r = -0.25$), and no evidence in R_L ($W = 165$, $z = -0.54$, $p\text{-value}_{1\text{ tail}} = 0.2951$, $r = -0.11$).

However, a decline in the number of adjustments towards prior threshold realizations does not necessarily capture a decline in the magnitude of decision adjustments period-to-period. Within each Innovator setting, we hence run a fix-effects panel regression model in which the only difference with respect to (2.5) is the dependent variable $|C_{H,i,t+1} - C_{H,i,t}|$, which captures participants' tendency to reduce the absolute change in complexity levels between two consecutive rounds over time. Table 2.6

shows significant evidence of a decline in the magnitude of decision adjustments period-to-period in C_L only.

Whereas the number of adjustments towards prior threshold realizations tends to decline over time, the magnitude of the adjustments does not. Taken together, we observe poor evidence of learning to avoid the threshold chasing bias.

Table 2.6. Fixed-effects panel regression of learning to avoid the threshold chasing bias.^a

	C_H	R_H	C_L	R_L
β_0 (Constant)	11.31* (1.4634)	13.63* (1.7870)	9.41* (1.3876)	13.77* (1.8270)
β_1 (Round)	-0.0557 (0.0414)	0.0184 (0.0507)	-0.0837† (0.0395)	0.0006 (0.0522)
β_2 (Over)	0.1379* (0.0265)	0.0008 (0.0321)	0.1890* (0.0299)	0.0160 (0.0334)
β_3 (Under)	0.2920* (0.0414)	0.0767‡ (0.0423)	0.1404* (0.0315)	0.0793‡ (0.0431)
R^2	0.2758	0.1448	0.1872	0.1549
F	18.41	1.45	15.97	1.19
p -value	0.0000	0.2274	0.0000	0.3129

^a Standard errors in parentheses.

* Highly significant, † Significant, ‡ Marginally significant.

2.6. DISCUSSION

We proposed a stylized analytical model to inform NPD decision making under innovation uncertainty. Assuming a single stage-gate innovation pipeline under a single uncertainty source, the proposed Innovator model draws a close parallel with the Newsvendor model, a traditional Operations Management model for ordering inventory under stochastic demand. The study shows how decisions in our stylized Innovator model are analytically equivalent to decisions in the traditional Newsvendor model. The analogy allows applying insights from the Newsvendor model to NPD decision making

under innovation uncertainty. Hence, our research seeks to inform decision making under uncertainty in NPD settings.

By the same token, our research also suggests that NPD managers may be prone to the same decision making biases commonly observed in Newsvendor settings. Operationalizing the Innovator model under complexity and resource uncertainty separately, a lab experiment showed that decision making biases in NPD settings resemble those observed in Newsvendor experiments. In particular, we observed a pull-to-center effect. That is, for high-profit projects, individuals under react when innovation costs and uncertainty require more ambitious scopes or resources. In contrast, for low-profit projects, they overreact when innovation costs and uncertainty require less ambitious scopes or resources. In addition, the effect is stronger in low-profit projects. That is, the overreaction is stronger than the under reaction.

Under the proposed operationalizations of the Innovator model, high-profit Innovator settings require setting high complexity levels (ambitious design goals) or allocating a large amount of resources (stretched resource allocations) in order to launch a successful product to the market place. In contrast, low-profit Innovator settings require setting low complexity levels or allocating a small amount of resources in order to launch a successful product to the market place. Pull-to-center effect results then suggest that NPD managers may underperform in demanding markets (high-profit Innovator settings), limiting their a priori likelihood of success, and over perform in less challenging markets (low-profit Innovator settings), a priori investing more effort than that required for success. In addition, the asymmetric pull-to-center effect in low-profit Innovator settings suggests that less challenging markets may pose more survival threats to NPD managers and their organizations.

Moreover, we observed no differences between Complexity and Resource Innovators in the same profit condition, suggesting that a poor understanding of the structure of the problem rather than a specific uncertainty type drives poor managerial decision making behavior under innovation uncertainty. That is, individuals seem to fail to find a balance between over and under development costs under innovation uncertainty in general. These results may hold also for NPD managers since purchasing managers exhibit a similar behavior in the analogous Newsvendor model (Bolton *et al.*,

2012). In addition, we also observed a threshold chasing bias. That is, individuals tend to chase uncertainty thresholds realized in previous innovation efforts. These results suggest that NPD managers may be affected by a recency bias that makes individuals to place a higher weight on recent events to the detriment of an understanding of the structure of the task at hand. Taken together, these results suggest that NPD managers may poorly understand decision making under innovation uncertainty and the influence of threshold cues may prevent them from engaging in understanding it in the first place. The mixed results regarding learning to avoid both the pull-to-center effect and the threshold chasing bias further reinforce the previous point.

Although Newsvendor research has shown that in some cases individuals recognize the structure of the problem, it has also shown that they may still fail to convert such information into good Newsvendor decision making (Cui *et al.*, 2013; Gavirneni and Isen, 2010). However, Bolton *et al.* (2012) showed that Newsvendor task training helps individuals convert such information into improved Newsvendor decision making. Hence, this suggests that training NPD managers in the Innovator model may help them mitigate biased behavior.

Overall, our research suggests that the Innovator model can be used as a building block to study NPD decision making under innovation uncertainty, bringing special attention to the study of managerial biases in NPD settings. However, we also acknowledge the novelty of the application of the Innovator model and the challenges that it imposes. For instance, we proposed a stylized analytical model by assuming a single stage-gate innovation pipeline under a single uncertainty source. These simplifying assumptions bring attention to external validity concerns typical of most experimental studies. Hence, future work could expand the Innovator model by incorporating more than one development stage, several uncertainty sources simultaneously, or relax both assumptions in a systematic manner to study the effect of increasing complexity levels of the innovation setting on decision making.

Also, managers in real situations may not decide as did our sample of both undergraduate and graduate students from management-related disciplines. Although there is not systematic evidence indicating that managers perform better than students in Newsvendor settings (Bolton *et al.*, 2012), it

may be the case that NPD settings impose additional challenges to those typical of Newsvendor settings (e.g., potentially riskier environments), and thus managers may decide differently in an Innovator setting. Hence, future work could test the Innovator model with a sample of managers. In addition, monetary rewards were not used to incentivize participants. On the one hand, it could be argued that monetary rewards would improve results. On the other hand, the fact that behavior was fairly consistent with behavior observed in previous Newsvendor experiments may cast doubt on this observation. Nevertheless, future research could use monetary rewards for the sake of experimental rigor and analyze whether the use of incentives makes a significant difference.

Similarly, other contextual factors potentially important in NPD settings such as incentive systems and group decision process were not taken into consideration to avoid introducing confounding factors into the analysis and run a clean test of the Innovator model. Future work could explore how these and other contextual factors influence Innovators' behavior.

Notwithstanding these limitations, endeavors like this make the tradeoff inherent to experiments, i.e., the advantages of experiments for controlling confounding factors and establishing cause-and-effect relationships vs. the lack of external validity, both necessary and acceptable. The Innovator model is a first step to formally explore how NPD managers make decisions under different types of innovation uncertainty. We believe there are significant research opportunities ahead along this same line.

Appendix 2.1. Sample of written instructions (C_H)

INSTRUCTIONS

The purpose of this session is to study how people make decisions in product development (PD) efforts.

DESCRIPTION OF THE GAME

You are a senior PD manager deciding the complexity level of the projects that your company will launch. Your complexity decisions (e.g., the functionality of a software program) influence the likelihood that a project may pass (or not) the threshold established in the screening process set by the company's vice president (VP).

For each project, your complexity decision must be made before you know for certain what the VP's screen level is. Based on past projects, however, you know that the screen is uniformly distributed between 1% and 100%. That is, the screen level is equally likely to take any value from 1% to 100%. Moreover, projects are independent of each other. That is, complexity decisions made in one project do not carry over and do not affect other projects.

Projects that are launched generate on average more profits than projects that fail. Profits for projects that are launched are proportional to the screen level (the VP's assessed market potential). Profits for projects that fail are proportional to the complexity level (the amount of work the company has devoted to the project). The reward for a project launched is 12 francs for each screen unit. The cost for a project is 3 francs for each complexity unit.

GOAL

Your goal is to maximize the profits you make over 50 projects (rounds of decisions).

PLAYING THE GAME

To access the game, follow the link [*game link*].

DECISIONS

Please write down your **Complexity level** decisions in the table provided.

[Decision table]

After completing all your decisions, send the electronic data by email by following the next steps:

- **Right click on the table with your decisions** (the table on the upper right of the screen)
- **Select Copy Data and Paste it in an e-mail**
- **Copy and Paste the link of the simulation** (web address that appears in your Internet browser)
- **Send the e-mail to:** [e-mail address]

Appendix 2.2. Sample of game screen (C_H)

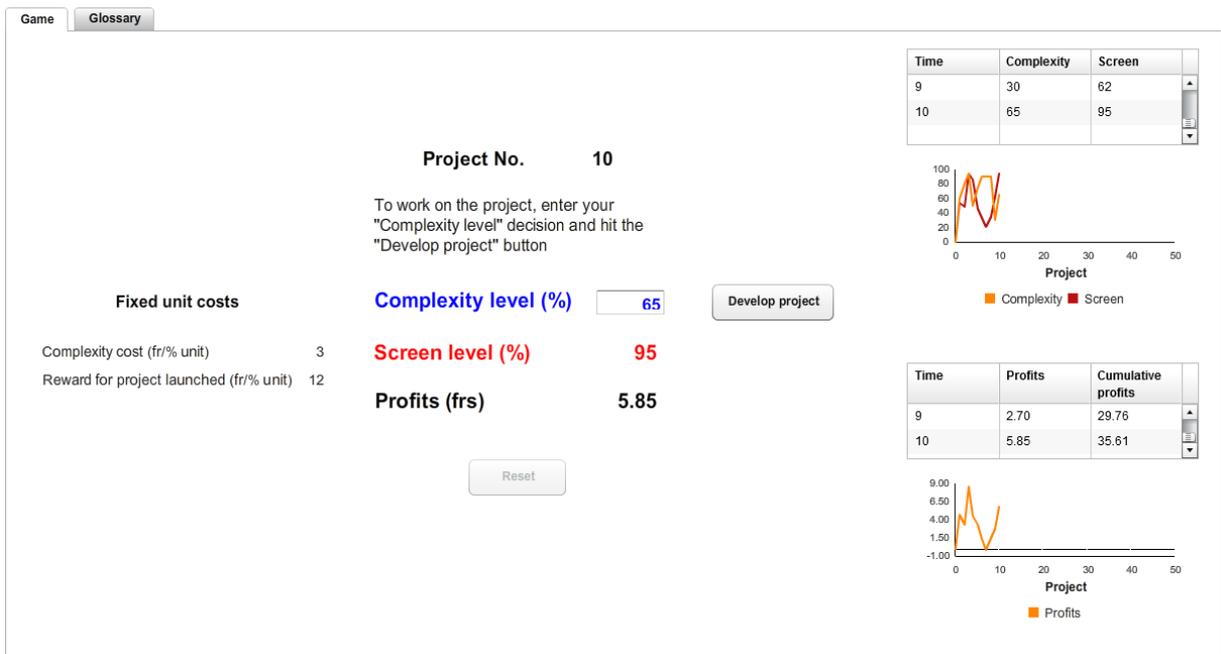


Figure A2.2.1. Sample of game screen.

CHAPTER 3

IMPACT OF JOINT DECISIONS AND COGNITIVE DISSONANCE ON PREPOSITIONING (NEWSVENDOR) DECISIONS

(with Paulo Gonçalves)

ABSTRACT

Prepositioning of emergency supplies is a critical task for the success of humanitarian relief operations. However, little is known about how humanitarian practitioners actually make prepositioning decisions. In a laboratory (lab) experiment based on the Newsvendor model, humanitarian practitioners prepositioned emergency supplies of different importance. When making single item decisions, practitioners' prepositioning behavior shows the pull-to-center effect observed in traditional Newsvendor experiments. When making decisions for two items of different importance, practitioners either increase or reduce the pull-to-center effect. In particular, practitioners made joint prepositioning decisions in either a cognitive dissonant treatment, where a high-importance item in a low-safety stock condition was joined with a low-importance item in a high-safety stock condition; or a cognitive consonant treatment, where a high-importance item in a high-safety stock condition was joined with a low-importance item in a low-safety stock condition. Results show that the importance of emergency items in joint decisions influences prepositioning behavior, with dissonant prepositioning decisions increasing the pull-to-center effect for high-importance items, and consonant prepositioning decisions reducing the pull-to-center effect for high-importance items. Neither dissonance nor consonance influence prepositioning behavior for low-importance items. Our research suggests that cognitive dissonance can influence joint prepositioning decisions in Newsvendor settings.

Keywords: Behavioral Operations Management, Cognitive Dissonance, Debiasing, Inventory Prepositioning, Laboratory Experiments, Newsvendor Model, Pull-to-Center Effect.

3.1. INTRODUCTION

Events triggering humanitarian action such as hurricanes, earthquakes, and acts of terrorism can strike communities with little warning and leave devastation, homeless people and casualties behind. As expressed by the Good Humanitarian Donorship initiative in their meeting held in Stockholm in 2003, “the objectives of humanitarian action are to save lives, alleviate suffering and maintain human dignity during and in the aftermath of man-made crises and natural disasters.” In order to achieve these goals, humanitarian organizations (HOs) provide shelter and assistance to victims of disasters as soon as possible. Under these circumstances, prepositioning of supplies becomes critical because supplies necessary to provide shelter and assistance are readily available when needed (Rawls and Turnquist, 2010).

Prepositioning supplies is not an easy task, however. One of the main difficulties for prepositioning activities is uncertainty about whether or not humanitarian emergencies will occur, and if they do, where and with what magnitude (Rawls and Turnquist, 2010). Examples illustrating these challenges are common. For instance, several manufacturing and retail firms experienced stock-outs in 2004 because they were not prepared to meet the demand caused by the multiple hurricanes that struck southeastern United States. In 2005, these firms again experienced stock-outs because of the extreme demand surge caused by Hurricane Katrina. These experiences motivated firms to be more aggressive in their approach to stocking supplies the following year. However, because of an inactive hurricane season in 2006, excess inventory was commonplace among firms (Taskin and Lodree Jr., 2010).

The inventory control literature is extensive; however, no much work has been done to directly address prepositioning plans for emergency supplies. For instance, based on a case study of a single HO operating a warehouse in Kenya and responding to the south Sudan crisis, Beamon and Kotleba (2006) developed a stochastic inventory control model that determines optimal order quantities and

reorder points for a long-term emergency relief response. Using the Newsvendor model as starting point, Lodree Jr. and Taskin (2008) introduced two variants to account for the uncertainties about (i) the occurrence of an extreme event and (ii) the demand for supplies, equipment, and manpower, comparing the solution of the modified model to the classic one. The difference was interpreted as an insurance premium associated with proactive disaster relief planning, establishing an insurance policy framework that managers can easily relate to in terms of quantifying the risks and benefits associated with stocking decisions related to preparing for disaster relief responses. Going beyond prepositioning activities, Balcik and Beamon (2008) integrated facility location and inventory ordering decisions in a variant of the maximal covering location model, considering also multiple item types, budgetary constraints, and capacity restrictions to determine the number and location of distribution centers in a relief network, and the amount of relief supplies to be stocked at each center.

The studies above are normative in nature, i.e., they provide an optimal solution given certain assumptions of the prepositioning task. To our knowledge, no previous work has directly assessed how people actually make inventory prepositioning decisions for emergency supplies. However, typical lab experiments on profit-based Newsvendor settings are insightful as to how individuals make inventory ordering decisions. They have shown that, on average, individuals' inventory ordering decisions are *lower* than the optimum when a *high-safety stock* is required; and they are *higher* than the optimum when a *low-safety stock* is required. This systematic Newsvendor result is known as the *pull-to-center effect* or the average tendency of individuals to order between the normative solution and the mean demand (Bostian *et al.*, 2008).

The pull-to-center effect has been replicated in several experiments since it was first documented by Schweitzer and Cachon (2000) (e.g., Benzion *et al.*, 2008, 2010; Bolton and Katok, 2008; Bostian *et al.*, 2008). Given the prevalence of this biased inventory ordering behavior, subsequent lab experiments have tested different mechanisms aimed at helping individuals to overcome it. The focus has been mainly on devising mechanisms to address the flat-maximum problem (or the flatness of the Newsvendor's expected profit function around the optimum) and its impediments for learning. For instance, some work has addressed the flat-maximum problem by sharpening payoff differentials (e.g.,

Bolton and Katok, 2008; Bostian *et al.*, 2008), whereas others have addressed it by modifying the frequency of decisions and feedback (e.g., Bolton and Katok, 2008; Bostian *et al.*, 2008; Lurie and Swaminathan, 2009). Results are not conclusive since they have shown no systematic positive effect by sharpening payoff differentials and mixed results by modifying the frequency of decisions and feedback.

In this paper, we run a Newsvendor experiment in which individuals make inventory repositioning decisions for emergency supplies and test a debiasing mechanism that departs from previously attempted debiasing efforts. In particular, we test the effectiveness of a joint decision framework that builds on cognitive dissonance theory as a possible debiasing mechanism for the pull-to-center effect. In the proposed joint decision framework, individuals make simultaneous inventory repositioning decisions for two items of different importance (high and low) each in one of two safety stock conditions (high and low). In our framework, item importance relates to its relevance to achieve the objectives of humanitarian action (e.g., save lives, alleviate suffering, and maintain human dignity), i.e., it is not safety stock related, whereas safety stock conditions refer to inventory levels that ensure that necessary supplies can be available as soon as possible at minimum cost.

Our experiment attempts to elicit a state of consonance or dissonance on individuals and explore how such states affect individuals' inventory repositioning decisions. We elicit a state of dissonance by asking individuals to simultaneously reposition a high-importance item in a low-safety stock condition and a low-importance item in a high-safety stock condition. Analogously, we elicit a state of consonance by asking individuals to simultaneously reposition a high-importance item in a high-safety stock condition and a low-importance item in a low-safety stock condition. Finally, we explore how the different states impact inventory repositioning decisions by comparing such decisions to those of a control in single inventory repositioning decision treatments. Our results show that dissonant and consonant states impact inventory repositioning decisions for high-importance items in Newsvendor settings, with dissonant states increasing the pull-to-center effect, and consonant states reducing it. Dissonant and consonant states do not seem to affect the decisions for low-importance items.

The rest of the paper is organized as follows. Section 3.2 frames prepositioning of emergency supplies as a Newsvendor problem, explores the pull-to-center effect literature, describes the joint decision framework, and develops hypotheses. Section 3.3 presents the lab design, treatments and experimental procedure. Section 3.4 presents the main results and hypothesis tests. Finally, section 3.5 summarizes the work, discusses the main findings, limitations, and opportunities for future research.

3.2. THEORY AND HYPOTHESES DEVELOPMENT

3.2.1. Prepositioning of emergency supplies as a Newsvendor problem

Several HOs preposition a number of emergency supplies or items (e.g., water, blankets, and vaccines) in preparation to humanitarian relief operations. Maintaining an adequate amount of prepositioned emergency items can have a significant impact on the success of humanitarian relief operations.

However, HOs must make prepositioning decisions without knowledge of beneficiary demand, since demand only materializes after a disaster strikes. In addition, a cost-effective use of funds is also an objective pursued by HOs given donors' pressure for effectiveness, transparency, and accountability (Thomas, 2003; Thomas and Kopczak, 2005; van der Laan *et al.*, 2009). By having uncertainty in beneficiary demand and cost-effectiveness metrics, the Newsvendor model (Arrow *et al.*, 1951) can inform inventory prepositioning decisions.

In an inventory prepositioning task, q prepositioned items are purchased at unit cost w . After a disaster strikes, beneficiary demand D is realized. If q exceeds D , then no items are expedited. However, each excess item $q - D$ incurs a disposal cost s . If, instead, D exceeds prepositioned amount q , then additional items $D - q$ must be expedited to meet beneficiary demand at an additional unit cost x . That is, unit overage cost c_o equals $w + s$ and unit underage cost c_u equals x . For simplicity, and following previous Newsvendor experiments, we assume no disposal cost when the amount prepositioned q exceeds demand D , i.e., $c_o = w$, and certainty in sales season (or emergency) occurrence (cf. Lodree Jr. and Taskin, 2008). If D is a random variable with cdf F , it is well-known that the optimal prepositioning quantity Q^* is characterized by the critical fractile $F(Q^*) = c_u / (c_u +$

c_o). Schweitzer and Cachon (2000) define a product as a *high-profit* (or *high-safety stock*) product when $F(Q^*) \geq 1/2$ and as a *low-profit* (or *low-safety stock*) product otherwise.⁵

3.2.2. Newsvendor's pull-to-center effect

Newsvendor experiments have consistently shown that individuals make biased inventory ordering decisions. In their seminal experimental study of the Newsvendor model, Schweitzer and Cachon (2000) showed that individuals' inventory ordering decisions are *lower* than the optimum in a *high-safety stock* condition, and *higher* than the optimum in a *low-safety stock* condition, i.e., individuals tend to order too few high-profit items and too many low-profit ones. This result is known as the *pull-to-center effect* (Bostian *et al.*, 2008).⁶

The pull-to-center effect has been replicated in several experiments since it was first documented by Schweitzer and Cachon (2000) (e.g., Benzion *et al.*, 2008, 2010; Bolton and Katok, 2008; Bostian *et al.*, 2008). Given the prevalence of this biased inventory ordering behavior and its incongruity with expected profit maximization, subsequent experimental research has proposed different mechanisms aimed at helping individuals to overcome it.

According to Bolton and Katok (2008), biased ordering behavior is consistent with the facts that “people are adaptive” and “have limited information processing capacity” (p. 521). In an attempt to overcome the bias, they experimentally tested modifications to experience and feedback “known to improve adaption or information processing” (p. 521). First, they provided individuals with extended experience by allowing them to make inventory ordering decisions during 100 rounds. Second, they sharpened payoff differentials by reducing the number of ordering options to potentially mitigate impediments to learning stemming from the flatness of the expected profit function around the maximum. Finally, they presented individuals with improved outcome feedback by providing them

⁵ The prepositioning task does not include any cost metric related to revenue, making thus the operational setting entirely cost-based yet structurally the same as profit-based settings as shown by $F(Q^*)$.

⁶ We emphasize the use of the expression pull-to-center effect over *anchoring and insufficient adjustment bias* (Schweitzer and Cachon, 2000) since the former does not make strong assumptions about what actually drives inventory ordering behavior, which remains an open question in Behavioral Operations Management research.

with payoffs of forgone options and by reducing decision frequency, imposing individuals to maintain the same order for 10 rounds.

A number of researchers have also provided individuals with extended experience by allowing them to make inventory ordering decisions for multiple rounds. Bostian *et al.* (2008) provided individuals with 30 decision rounds, whereas Bolton *et al.* (2012) and Benzion *et al.* (2008, 2010) provided individuals with 100 decision rounds. Others have also attempted to address the impediments to learning stemming from the flatness of the expected profit function around the maximum. Bostian *et al.* (2008) sharpened payoff differentials by making the economic consequences of under and over stocking twice as severe. They also presented individuals with improved outcome feedback by reducing decision frequency to once every 5 rounds. In a further manipulation, in addition to reducing decision frequency, outcome feedback itself was also shortened to once every 5 rounds. Analogously, Lurie and Swaminathan (2009) proposed a more systematic manipulation of decision frequency and outcome feedback frequency. They decoupled feedback frequency from decision frequency by fixing decision frequency and varying outcome feedback frequency in order to separate their effects.

Results from this line of research are inconclusive, since studies show mixed results. Some studies show improvements with trends in direction to optimal inventory ordering quantities (Benzion *et al.*, 2008; Bolton and Katok, 2008; Bostian *et al.*, 2008); others, however, show no improvement trends (Benzion *et al.*, 2010; Bolton *et al.*, 2012). Research results have shown no systematic positive effect of sharpening payoff differentials (Bolton and Katok, 2008; Bostian *et al.*, 2008), and mixed results by modifying the frequency of decisions and outcome feedback. For instance, Bolton and Katok (2008) find that reducing decision frequency matters; however, Bostian *et al.* (2008) and Lurie and Swaminathan (2009) find the opposite. Similarly, Lurie and Swaminathan (2009) find that reducing outcome feedback frequency matters; however, Bostian *et al.* (2008) find the contrary. Finally, Lurie and Swaminathan's (2009) research suggests that outcome feedback frequency may be more important than decision frequency.

This study departs from the previously described debiasing efforts by testing a novel joint decision framework that builds on cognitive dissonance theory as a potential debiasing mechanism for the pull-to-center effect. The framework and the corresponding hypotheses are described next.

3.2.3. Joint decision making as a debiasing mechanism

As the description above suggests, the pull-to-center effect is a robust finding in Newsvendor experiments. Still, while a number of experimental studies propose debiasing mechanisms in an attempt to move individuals' decisions closer to the optimum, the proposed mechanisms do not always work consistently across settings. This research proposes a debiasing mechanism that joins or bundles Newsvendor inventory ordering decisions for two items of different importance and different safety stock conditions. By joint decisions, we mean two simultaneous decisions (e.g., order quantity for two different items). Previous experimental studies on joint decisions have mainly considered choice partitioning (choice bracketing, outcome editing, and joint vs. separate evaluation of alternatives), which considers a joint evaluation of two mutually exclusive options (e.g., order *either* item A *or* item B). Hence, in such settings a person must choose a single option (e.g., ordering item A automatically rejects the alternative) (see Milkman *et al.* (2008) and Read *et al.* (1999) for reviews). The proposed joint decision framework is conceptually different from choice partitioning since people must make two decisions at the same time (e.g., order quantities for *both* item A *and* item B). Although the Newsstand model literature has studied simultaneous or multi-item Newsvendor decisions (e.g., Abdel-Malek and Montanari, 2005; Lau and Lau, 1995, 1996), the literature is normative in nature. Hence, little is known about how people actually make simultaneous inventory ordering decisions.⁷

By bundling decisions of high-importance items with low-importance ones in a Newsvendor experiment, we expect individuals to place large orders for high-importance items and small orders for low-importance items. Also, by pairing item importance and safety stock levels, we hope to further

⁷ An exception is Tong and Song's (2011) study of the effect of transaction utility (Thaler, 1980, 1985) on Newsvendor ordering decisions. However, their transaction utility framework is conceptually different from the proposed cognitive dissonant framework since they exploit safety stock condition comparisons, whereas we exploit both importance and safety stock condition comparisons, not to mention differences in transaction utility and cognitive dissonance arguments.

manipulate the traditional pull-to-center effect. In particular, we expect to *mitigate* biased inventory ordering behavior when we bundle orders for a high-importance item in a high-safety stock condition with those for a low-importance item in a low-safety stock condition. In contrast, we expect to *intensify* it when we bundle orders for a high-importance item in a low-safety stock condition with those for a low-importance item in a high-safety stock condition.

Baseline hypothesis

Our baseline hypotheses test the behavior of individuals when making separate inventory ordering decisions about items of difference importance. While we expect orders for items of different importance to change when those decisions are bundled (as we explain below), *a priori* we have no reason to expect them to change when those decisions are made separately. It is possible that individuals' separate decisions result in larger orders for a high-importance item in a high-safety stock condition than a low-importance item in the same condition. However, given the lack of a reference allowing individuals to compare their decisions in separate decision treatments, we do not expect decisions to be influenced by item importance. Therefore, we expect inventory prepositioning decisions in the same safety stock condition to be similar regardless of the importance of the item. Hence:

HYPOTHESIS 1A: In a separate decision treatment, the quantity of high-importance items prepositioned will be similar to the quantity of low-importance items prepositioned in a low-safety stock condition.

HYPOTHESIS 1B: In a separate decision treatment, the quantity of high-importance items prepositioned will be similar to the quantity of low-importance items prepositioned in a high-safety stock condition.

In other words, Hypothesis 1 tests that a potential alignment or misalignment among an item's importance and its safety stock condition by itself does not affect typical Newsvendor inventory ordering behavior. Decisions in the separate decision treatments serve thus as a baseline or controls for the joint decision treatments.

Cognitive dissonant hypothesis

Festinger's (1957) cognitive dissonance theory provides the underlying motivation for the proposed debiasing/bias-strengthening mechanism in the joint decision treatments. For Festinger (1957), individuals hold a multitude of *cognitions*, or bits of knowledge, simultaneously about different attributes (e.g., attitudes, emotions, and behaviors). Most of these cognitions have no relationship to each other and are said to be irrelevant. Some cognitions, however, are related to one another. Two related cognitions are said to be consonant if one cognition follows from, or fits with, the other. On the other hand, two cognitions are said to be dissonant if one cognition follows from the opposite of another. Instrumentally, Festinger's (1957) cognitive dissonance theory holds that individuals do not like to be in a state of *dissonance*, and are motivated to act to reduce the inconsistency. Actions that reduce dissonance can take place in three ways: (i) by changing one of the dissonant cognitions, with the action typically supporting the cognition most resistant to change; by (ii) adding consonant cognitions and/or subtracting dissonant cognitions to reduce the overall level of inconsistency; and by (iii) reaffirming, or increasing, the importance of consonant cognitions, or trivializing, or decreasing, the importance of dissonant cognitions (Harmon-Jones and Harmon-Jones, 2007; Simon *et al.*, 1995).

In our Newsvendor setting, a ***cognitive dissonant joint decision*** treatment has individuals placing orders both for a high-importance item in a low-safety stock condition and for a low-importance item in a high-safety stock condition. By bundling two decisions (cognitions) about two items of different importance, we remove by design dissonance reduction mechanisms (i) and (ii). Specifically, the importance and safety stock conditions of the items are fixed, and there are only two types of decisions, with no other cognitions added or subtracted. Hence, individuals in our experiment can only address the inherent dissonance by (iii) decreasing the importance of dissonant cognitions (Simon *et al.*, 1995).

Building on these arguments, we expect individuals in a cognitive dissonant joint decision treatment will reduce the importance of dissonant cognitions by placing large orders of the high-importance low-safety stock item and small orders of the low-importance high-safety stock item.

Moreover, a trivialization is often achieved by making an important cognition salient, directing attention towards it to the detriment of other cognitions (Simon *et al.*, 1995). In our experiment, the high-importance item is inherently more salient than the low-importance item, suggesting that individuals may direct attention toward inventory ordering decisions for the high-importance item to the detriment of those decisions for the low-importance one.

Taken together, we hypothesize that in a cognitive dissonant joint decision treatment individuals will trivialize orders of the high-importance low-safety stock item by ordering larger amounts relative to the corresponding low-safety stock baseline. Such result would strengthen the pull-to-center effect. In addition, trivialization suggests that individuals may place orders for the low-importance high-safety stock item that are smaller than the corresponding high-safety stock baseline. However, given the salience of the high-importance item, we hypothesize that individuals will not direct attention to decisions of the low-importance item, ordering a similar amount of the low-importance item than the baseline amount. Hence:

HYPOTHESIS 2A: In a cognitive dissonant joint decision treatment, the quantity of high-importance items prepositioned will be larger than the quantity of high-importance items prepositioned in a separate decision treatment in a low-safety stock condition.

HYPOTHESIS 2B: In a cognitive dissonant joint decision treatment, the quantity of low-importance items prepositioned will be similar to the quantity of low-importance items prepositioned in a separate decision treatment in a high-safety stock condition.

Since we cannot infer the magnitude of the dissonance effect, it is not possible to estimate how far orders for the high-importance item may be from the optimum. Hence, it is possible that orders may overshoot the mean in a low-safety stock condition, strengthening even further the pull-to-center effect.

Cognitive consonant hypothesis

While Festinger (1957) did not specifically theorize about consonance effects and the way individuals react to them, he asserted that consonant cognitions reaffirm each other (Harmon-Jones and Harmon-Jones, 2007; Simon *et al.*, 1995). Hence, we explore such reaffirmation and extend Festinger's (1957) and Simon *et al.*'s (1995) dissonance arguments to consonance. In our Newsvendor setting, a ***cognitive consonant joint decision*** treatment has individuals placing orders both for a high-importance item in a high-safety stock condition and for a low-importance item in a low-safety stock condition.

We conjecture that individuals in a cognitive consonant joint decision treatment will reaffirm the importance of consonant cognitions by placing large orders for the high-importance high-safety stock item. Hence, we first hypothesize that individuals will preposition a larger amount of the high-importance high-safety stock item relative to the corresponding high-safety stock baseline, potentially lessening the pull-to-center effect. Moreover, extending Simon *et al.*'s (1995) arguments on the salience of the high-importance item to the detriment of the low-importance one, we further hypothesize that individuals will preposition a similar amount of the low-importance item relative to the corresponding low-safety stock baseline. Hence:

HYPOTHESIS 3A: In a cognitive consonant joint decision treatment, the quantity of high-importance items prepositioned will be larger than the quantity of high-importance items prepositioned in a separate decision treatment in a high-safety stock condition.

HYPOTHESIS 3B: In a cognitive consonant joint decision treatment, the quantity of low-importance items prepositioned will be similar to the quantity of low-importance items prepositioned in a separate decision treatment in a low-safety stock condition.

Similarly, we cannot infer the magnitude of the consonance effect. It is possible that consonance may result in orders that overshoot the optimum.

In summary, our experiments treat decisions as cognitions and bundles consonant and dissonant decisions to explore their impact on Newsvendor inventory ordering behavior. Our hypotheses on cognitive dissonance build on Festinger's (1957) work, where those on dissonance reduction actions build on Simon *et al.* (1995). Our hypotheses on cognitive consonance parallel the arguments and theory available to cognitive dissonance. We conjecture that dissonant decisions may intensify the pull-to-center effect, whereas consonant decisions may mitigate it. The next section presents the lab experiment and the proposed treatments.

3.3. LABORATORY EXPERIMENT

3.3.1. Basic design

Typical Newsvendor experiments fix selling price p and manipulate purchasing cost w in order to create two basic treatments: one for high-profit products and another for low-profit products. Given our emphasis on cost, we fix the expediting cost at $x = 3$ and manipulate prepositioning cost w , setting it up at $w = 1$, implying a critical fractile of $3/4$ (high-safety stock items), and $w = 9$, implying a critical fractile of $1/4$ (low-safety stock items). While the experimental design guarantees that any unmet demand resulting from insufficient inventory prepositioning is eventually met through expediting, meeting demand takes place with a delay and at additional cost. Hence, one cannot simply assume that the dissonance associated with insufficient inventory prepositioning can be dismissed.

We consider an uniformly distributed beneficiary demand $D \sim U(0, 100)$ with mean quantity = 50 and integer values for both high- and low-safety stock items in all treatments, consistent with related literature (e.g., Bolton and Katok, 2008; Schweitzer and Cachon, 2000). Besides capturing uncertainty in beneficiary demand, a demand distribution also captures the implicit assumption that HOs will set the boundaries of uncertainty based on information about past emergencies. In each treatment, we use a different noise seed to ensure different realizations of demand in each period. During the instructions, we inform participants about the different realizations of demand. Results of previous Newsvendor experiments suggest the use of different noise seeds to avoid confusion since individuals tend to chase demand (e.g., Benzion *et al.*, 2008; Bolton and Katok, 2008; Kremer *et al.*, 2010).

The described parameterization implies an optimal inventory prepositioning quantity in the high-safety stock condition of 75 ($Q_H^* = 75$) and an optimal inventory prepositioning quantity in the low-safety stock condition of 25 ($Q_L^* = 25$).

3.3.2. Notation

Let X_i^j , with $X \in (C, N)$, denote the number of critical-to-life (C) items (high-importance items), or nice-to-have (N) items (low-importance items); where $i \in (H, L)$ refers to a high-safety stock (H) or a low-safety stock (L) condition; and $j \in (S, J)$ refers to a separate (S) or a joint decision (J) treatment. For instance, C_L^S refers to the number of critical-to-life items prepositioned in a separate decision treatment in a low-safety stock condition, whereas N_H^J refers to the number of nice-to-have items prepositioned in a joint decision treatment in a high-safety stock condition.⁸

3.3.3. Treatments

To explore the impact of dissonance theory on inventory prepositioning decisions, we create a 2x2 full factorial design. The factors are cognitive state, viz dissonant and consonant, and type of decision, viz separate and joint, i.e., there are two levels for each factor. The *cognitive dissonant separate decision* treatment corresponds to (i) a critical-to-life low-safety stock item and (ii) a nice-to-have high-safety stock item, which are run independently. The *cognitive consonant separate decision* treatment corresponds to (iii) a critical-to-life high-safety stock item and (iv) a nice-to-have low-safety stock item, which are also run independently. For clarity of exposition, the separate (or baseline) decision treatments are shown separately. Hence, we have 4 baseline treatments:

T1: critical-to-life low-safety stock items (C_L^S),

T2: nice-to-have high-safety stock items (N_H^S),

T3: critical-to-life high-safety stock items (C_H^S), and

⁸ Besides using generic names for the items, the context of the experiment is in general abstract (e.g., there is no mention to any emergency type or region). That way, we motivate items' importance only through the framing of the experiment and not through other factors such as participants' experience with any given emergency item, emergency type, and/or region, etc., avoiding thus leading participants (Katok, 2011).

T4: nice-to-have low-safety stock items (N_L^S).

In the joint decision treatments, we bundle participants' decisions about the quantity of critical-to-life and nice-to-have items to preposition in two joint treatments: a dissonant treatment ($T5$) and a consonant treatment ($T6$). In the *cognitive dissonant joint decision* treatment ($T5$), participants make joint inventory prepositioning decisions of a high-importance low-safety stock item (C_L^J) and a low-importance high-safety stock item (N_H^J), i.e., $T1$ and $T2$ are bundled. In the *cognitive consonant joint decision* treatment ($T6$), participants make joint inventory prepositioning decisions of a high-importance high-safety stock item (C_H^J) and a low-importance low-safety stock item (N_L^J), i.e., $T3$ and $T4$ are bundled. For simplicity, we assume no correlation between items' demand when they are prepositioned jointly, similar to typical normative Newsstand research (e.g., Abdel-Malek and Areeratchakul, 2007; Abdel-Malek and Montanari, 2005), and no resource constraints (cf. Lau and Lau, 1995, 1996). These assumptions avoid introducing confounding factors, providing thus a clean test of the effects of cognitive dissonance.

3.3.4. Experimental procedure

A total of 43 people participated in our experiment. The results include decisions from 42 participants. We excluded 1 participant because she did not complete one of the treatments. All participants are full-time humanitarian practitioners working in different areas (e.g., logistics, field operations, and program management) for national and international organizations (e.g., Oxfam, WVI, UNHCR, UNICEF, WFP, and IFRC). All participants were students enrolled in an executive Master program in humanitarian logistics and management in a Swiss University in 2012. Table 3.1 shows the treatments with their corresponding notations and number of participants and with the dissonance-related treatments highlighted to differentiate them from the consonance-related ones.

Table 3.1. Treatments, notation, and number of participants.

		<i>Item</i>		
		Critical-to-Life	Nice-to-Have	
SEPARATE	<i>Safety stock condition</i>	Low	$T1 (C_L^S)$ 11	$T4 (N_L^S)$ 10
		High	$T3 (C_H^S)$ 10	$T2 (N_H^S)$ 11
	<i>Safety stock condition</i>	Low	$T5 (C_L^J)$ 11	$T6 (N_L^J)$ 10
		High	$T6 (C_H^J)$ 10	$T5 (N_H^J)$ 11

All sessions were conducted in classroom as a class exercise; monetary rewards were not used. Protocols of experimental economics (Smith, 1976, 1982) call for performance-based monetary rewards. However, Camerer and Hogarth (1999) document a number of experiments showing that the improvement in performance expected by using monetary rewards is not observed. Similarly, Arkes (1991) shows that financially motivated individuals may perform suboptimal behaviors with more enthusiasm. Moreover, Osterloh and Frey (2002) and Perry *et al.* (2006) show that financial incentives may be ineffective in stimulating socially motivated actions, e.g., humanitarian actions. The effect of monetary rewards in a prepositioning task is left for future research.

Participants arrived, received instructions (see Appendix 3.1) and were seated so that they could not see decisions from other participants. They were informed that the purpose of the experiment was to understand how humanitarian practitioners make inventory prepositioning decisions. Participants had time to ask clarifying questions before initiating the simulation. After the instructions, they were directed to a web simulator with a randomly assigned treatment. The simulator was developed in Forio Business Simulations (www.forio.com). The simulator contained an introduction screen with information about the types of items and the need to preposition items in preparation for emergencies. The introduction screen also reminded participants about the basic characteristics of the decisions they were about to make. In all treatments, participants had to make 30 consecutive inventory

prepositioning decisions. After they entered their decisions, the system automatically revealed the demand realization(s) and cost(s). At any time, participants had access to information about prepositioning and expediting costs w and x , respectively. They also had access to all previous decisions and outcomes, including demand realizations, costs and total cumulative costs. This information was presented both in tables and graphs (Appendix 3.2 shows a snapshot of the game screen).

Given the difficulty in having access to a pool of humanitarian practitioners, we ran a within-subjects design along the dissonance-related ($T1$ - $T2$ and $T5$) and the consonance-related ($T3$ - $T4$ and $T6$) treatments, i.e., we ran two samples of participants through our treatments. Each sample of participants made separate decisions in the first experimental session, and joint decisions in a second session one day later. For example, in one of the first sessions, half of the participants in one sample made 30 separate inventory prepositioning decisions about a critical-to-life low-safety stock item C_L^S ($T1$), and then 30 separate decisions about a nice-to-have high-safety stock item N_H^S ($T2$). The other half of the sample first played $T2$ followed by $T1$, allowing us to control for order of presentation effects in separate decision treatments. In the second experimental session the following day, the same sample of participants played the cognitive dissonant joint decision treatment C_L^J, N_H^J ($T5$). A similar procedure was followed for the consonant-related treatments ($T3$ - $T4$ and $T6$). Notice that participants participated in the joint decision treatments always after separate decision ones. The choice to have separate treatments always before joint ones allows us to run a clean test for separate decisions without priming participants with a reference that could affect their separate decisions. Hence, we purposely did not control for order of presentation effects between separate and joint decision treatments.

3.4. RESULTS

3.4.1. Newsvendor biases in inventory prepositioning decisions

The main unit of analysis used to compare results across treatments is the population of *participants'* average prepositioned quantities *over time*, i.e., we compare populations of participants' averages over time since it is individual participants, not groups of participants, that exhibit behavioral patterns (Rudi

and Drake, 2011). For example, the average quantity of critical-to-life items (C) prepositioned in a low-safety stock condition (L) in the baseline separate treatment (S) is given by (C_L^S):

$$C_L^S = \frac{\sum_{i \in C_L^S} C_{L,i}^S}{\langle C_L^S \rangle} \quad (3.1)$$

where $C_{L,i}^S$ is the average quantity of critical-to-life items prepositioned by participant i across all rounds and $\langle C_L^S \rangle$ is the number of participants in the treatment.

We tested for the pull-to-center behavior described in traditional Newsvendor experiments. Table 3.2 provides the mean, standard error and 95% bootstrap confidence intervals (Appendix 3.3 provides an explanation) of the average quantities prepositioned in the four baseline treatments. The results for the baseline treatments are consistent with typical Newsvendor biased inventory ordering behavior. In particular, none of the confidence intervals around the average inventory prepositioning behavior contains the optimal inventory prepositioning quantity. Moreover, in all the baseline treatments there is a violation of the pull-to-center region since all the intervals include the mean, a robust result within the literature (e.g., Rudi and Drake, 2011; Thomas *et al.*, 2007).

Table 3.2. 95% bootstrap confidence intervals of average inventory prepositioning behaviors.

	\bar{x}	SE	CI	<i>Pull-to-center region</i>
<i>Separate decision treatments</i>				
C_L^S	44.42	5.59	[32.89 54.46]	[25 50]
N_H^S	44.17	5.20	[33.08 53.52]	[50 75]
C_H^S	51.84	4.00	[43.84 59.56]	[50 75]
N_L^S	45.72	4.43	[36.39 53.74]	[25 50]
<i>Joint decision treatments</i>				
C_L^J	51.19	5.96	[37.98 61.50]	[25 50]
N_H^J	41.28	5.59	[30.83 51.97]	[50 75]
C_H^J	59.77	4.09	[51.47 67.33]	[50 75]
N_L^J	43.30	4.25	[33.56 50.74]	[25 50]

Table 3.2 also provides the mean, standard error and 95% bootstrap confidence intervals for the average quantities prepositioned in the two joint decision treatments. The results for the joint decision treatments are consistent with typical Newsvendor biased ordering behavior since none of the confidence intervals around the average inventory prepositioning behavior contains the optimal inventory prepositioning quantity. Compared to separate decision treatments, however, the results seem to be consistent with the different hypotheses since it appears that dissonant and consonant decisions strengthen and lessen the pull-to-center effect for high-importance items, respectively, whereas it appears that biased inventory ordering behavior for low-importance items is not influenced by dissonance and consonance.

3.4.2. Baseline results: separate decision treatments

As a first step, we tested for order of presentation effects in the baseline treatments using Wilcoxon Rank-Sum tests (Appendix 3.3 provides an explanation). The results are consistent with no order of presentation effects since Wilcoxon Rank-Sum tests show that we cannot statistically rule out the possibility that there is a similar average inventory prepositioning behavior regardless of the order of presentation for C_L^S ($W = 26, z = -0.73, p\text{-value}_{2\text{ tails}} = 0.4652, r = -0.22$), N_H^S ($W = 26, z = -0.73, p\text{-}$

$value_{2\ tails} = 0.4652, r = -0.22$), $C_H^S (W = 22, z = -1.15, p\text{-value}_{2\ tails} = 0.2506, r = -0.36)$, and $N_L^S (W = 27, z = -0.10, p\text{-value}_{2\ tails} = 0.9168, r = -0.03)$. Because the no order of presentation effects hypotheses are stated as the null, the best we can do is to fail to reject it. This result does not automatically allow us to accept no order of presentation effects since we could be making a type II error. However, observing the effect size —denoted as r — provides us with an objective measure of the importance of the effect (Field, 2009). Three out of four effect sizes (in absolute value) have a low (N_L^S) or a low-to-middle importance (C_L^S and N_H^S) since they are below the cut-off values of 0.10 and 0.30, respectively (Field, 2009). Hence, the results are fairly consistent with no order of presentation effects.

Here, we investigate whether participants in separate decision treatments make similar inventory prepositioning decisions for critical-to-life and nice-to-have items in low- and high-safety stock conditions (Hypotheses 1A and 1B, respectively). Visual inspection of Figure 3.1 shows no pronounced differences in inventory prepositioning decisions for separate decisions in the low-safety stock condition, while a slightly more noticeable difference in the high-safety stock condition. A Wilcoxon Rank-Sum test shows that we cannot statistically rule out the possibility that participants make similar inventory prepositioning decisions for critical-to-life and nice-to-have items both in a low- safety stock (H1A: C_L^S and N_L^S) and in a high-safety stock (H1B: N_H^S and C_H^S) condition (H1A: $W = 112, z = 0.14, p\text{-value}_{2\ tails} = 0.8880, r = 0.03$ — H1B: $W = 123, z = 0.92, p\text{-value}_{2\ tails} = 0.3600, r = 0.20$). In addition, the effect size for the low-safety stock condition has a low importance since it is below the cut-off value of 0.10, whereas the effect size for the high-safety stock condition has a low-to-middle importance since it is below the cut-off value of 0.30 (Field, 2009). Hence, our results are fairly consistent with hypotheses H1A and H1B. Together, they suggest that in absence of a reference provided by a joint decision, participants treat inventory prepositioning decisions for critical-to-life and nice-to-have items (in the same safety stock condition) similarly when making separate decisions.

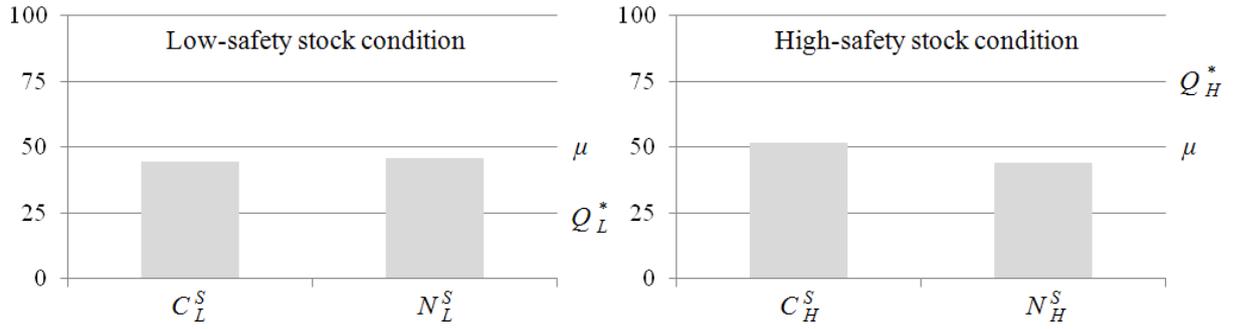


Figure 3.1. Average inventory prepositioning behaviors in separate decision treatments.

3.4.3. Cognitive dissonant joint treatment's results

Although we did not control for order of presentation effects between the separate and the joint decision treatments in the design to avoid priming participants in separate decision treatments with a prior experimental reference, we tested for learning. Evidence of learning in joint decision treatments, which are performed after the separate ones, would suggest that the separate treatments influence decisions in our within-subjects design. To test for learning, we ran the following fixed-effects regression in each treatment:

$$|C_{L,i,t+1}^J - Q_L^*| = \beta_0 + \beta_1 t + \beta_2 \text{Over}_{i,t} + \beta_3 \text{Under}_{i,t} + v_i + \varepsilon_{i,t}, \quad t = 1, \dots, 29 \quad (3.2)$$

where the dependent variable captures participants' tendency to get closer to the optimal inventory prepositioning quantity over time, t refers to round, $\text{Over}_{i,t}$ and $\text{Under}_{i,t}$ refer to the amount of excess and short of demand items of participant i in round t , respectively, and serve as a control for demand chasing effects, v_i is the participants' effect, and $\varepsilon_{i,t}$ is the error term.. Evidence of learning, provided by a significant and negative round coefficient, is observed for C_H^J only. However, if the order of presentation had a learning effect, it would be evident in the first rounds through differences between C_H^S and C_H^J . We tested for such differences with a Wilcoxon Signed-Rank test (Appendix 3.3 provides an explanation) instead of a Wilcoxon Rank-Sum test given the within-subjects design along these treatments. We could not observe differences in the first ten rounds. Specifically, our results show that

we cannot statistically rule out the possibility that there is a similar average inventory prepositioning behavior between C_H^S and C_H^J in the first ten rounds ($T = 25, z = -0.25, p\text{-value}_{2\text{ tails}} = 0.7989, r = -0.06$). In addition, the effect size (in absolute value) has a low importance since it is below the cut-off value of 0.10 (Field, 2009). Hence, the results are fairly consistent with no learning from having the separate decision treatments before the joint decision ones.

Here, the separate decision treatments set benchmarks for expected behavior without the joint manipulation treatments. The first joint treatment hypotheses explore the potential impact of dissonance as well as salience. Hypothesis 2A tests whether in a low-safety stock condition the prepositioned quantity of high-importance items in a *cognitive dissonant joint decision* treatment will be larger than that of high-importance items in a *separate decision* treatment. That is, the hypothesis explores in a low-safety stock condition how participants change their inventory prepositioning decisions for a high-importance item due to the availability of a *dissonant* reference (the low-importance high-safety stock item) in a *joint decision* treatment. A different prepositioned amount suggests that the availability of the reference influences participants' decisions. Higher prepositioned amounts in the *cognitive dissonant joint decision* treatment imply a *larger* distance to the optimum, which indicates that *cognitive dissonant joint decisions* for a high-importance item in a low-safety stock condition strengthen the pull-to-center effect. Visual inspection of Figure 3.2 seems to support this conclusion. A Wilcoxon Signed-Rank test shows that the strengthening holds statistically since the prepositioning quantity for the critical-to-life (high-importance) item in the cognitive dissonant joint decision treatment (C_L^J) is significantly above the quantity for the critical-to-life item in its separate decision counterpart (C_L^S) ($T = 8, z = -2.22, p\text{-value}_{1\text{ tail}} = 0.0131, r = -0.47$). Hence, the test supports Hypothesis 2A and suggests a strengthening of the pull-to-center effect in the cognitive dissonant joint decision treatment for the high-importance item.

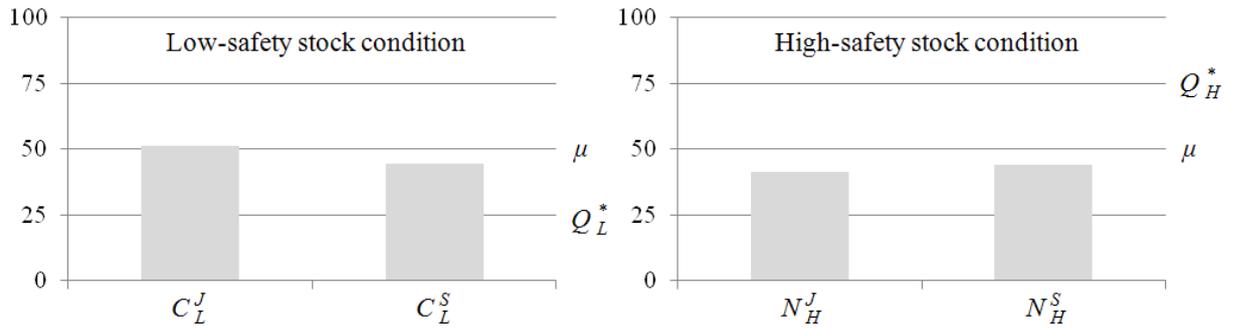


Figure 3.2. Average inventory prepositioning behaviors in dissonance-related treatments.

Hypothesis 2B tests whether the salience of the high-importance item leads to less attention to the low-importance item. In particular, it tests whether in a high-safety stock condition the prepositioned quantity of low-importance items in a *cognitive dissonant joint decision* treatment will be similar to that of low-importance items in a *separate decision* treatment. That is, the hypothesis explores in a high-safety stock condition how participants change their inventory prepositioning decisions for a low-importance item due to the availability of a *salient* and *dissonant* reference (the high-importance low-safety stock item) in a *joint decision* treatment. Hypothesis 2B states that the quantity of low-importance high-safety stock items prepositioned in a *cognitive dissonant joint decision* treatment will be similar to the quantity of low-importance high-safety stock items prepositioned in a *separate decision* treatment. Visual inspection of Figure 3.2 does not suggest a pronounced difference. A Wilcoxon Signed-Rank test shows that we cannot statistically rule out the possibility that there is a similar prepositioned quantity between nice-to-have (low-importance) items in the cognitive dissonant joint decision treatment (N_H^J) and the separate decision counterpart (N_H^S) ($T = 24$, $z = -0.80$, $p\text{-value}_{2\text{ tails}} = 0.4236$, $r = -0.17$). In addition, the effect size (in absolute value) has a low-to-middle importance since it is below the cut-off value of 0.30 (Field, 2009).

Hence, our results are fairly consistent with hypotheses H2A and H2B. Together, they suggest that participants (a) in a low-safety stock condition prepositioning high-importance items *strengthen* the pull-to-center effect when making *cognitive dissonant joint decisions* and (b) in a high-safety stock condition prepositioning low-importance items present a similar pull-to-center effect in *cognitive*

dissonant joint and *separate decision* treatments. Our results suggest that the availability of a *dissonant* reference in the *joint decision* treatment (a) influences participants' inventory repositioning behavior for high-importance items and (b) does not for low-importance ones. We conjecture that participants pay more attention to the salient decisions (high-importance item) to the detriment of other ones (low-importance item) when making cognitive dissonant joint decisions.

3.4.4. Cognitive consonant joint treatment's results

Here, we explore the potential impact of consonance and, again, salience. Hypothesis 3A tests whether in a high-safety stock condition the repositioned quantity of high-importance items in a *cognitive consonant joint decision* treatment will be larger than that of high-importance items in a *separate decision* treatment. That is, the hypothesis explores in a high-safety stock condition how participants change their inventory repositioning decisions for a high-importance item due to the availability of a *consonant* reference (the low-importance low-safety stock item) in a *joint decision* treatment. Again, a different repositioned amount suggests that the availability of the reference influences participants' decisions. Higher repositioned amounts in the *cognitive consonant joint decision* treatment imply a *shorter* distance to the optimum, which indicates that *cognitive consonant joint decisions* for a high-importance item in a high-safety stock condition lessen the pull-to-center effect. Visual inspection of Figure 3.3 seems to support this conclusion. A Wilcoxon Signed-Rank test shows that the statistical result holds at the 10% significance level. The repositioning quantity for the critical-to-life (high-importance) item in the cognitive consonant joint decision treatment (C_H^J) is marginally significantly above the quantity for the critical-to-life item in its separate decision counterpart (C_H^S) ($T = 13$, $z = -1.48$, $p\text{-value}_{1\text{ tail}} = 0.0697$, $r = -0.33$). Hence, the test marginally supports Hypothesis 3A and suggests a lessening of the pull-to-center effect in the cognitive consonant joint decision treatment for the high-importance item.

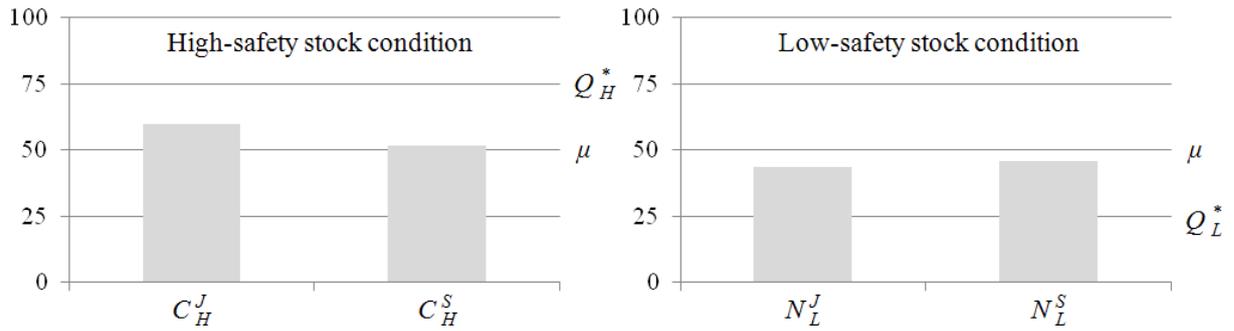


Figure 3.3. Average inventory positioning behaviors in consonance-related treatments.

Hypothesis 3B again tests whether the salience of the high-importance item leads to less attention to the low-importance item. In particular, it tests whether in a low-safety stock condition the prepositioned quantity of low-importance items in a *cognitive consonant joint decision* treatment will be similar to that of low-importance items in a *separate decision* treatment. That is, the hypothesis explores in a low-safety stock condition how participants change their inventory positioning decisions for a low-importance item due to the availability of a *salient* and *consonant* reference (the high-importance high-safety stock item) in a *joint decision* treatment. Hypothesis 3B states that the quantity of low-importance low-safety stock items prepositioned in a *cognitive consonant joint decision* treatment will be similar to the quantity of low-importance low-safety stock items prepositioned in a *separate decision* treatment. Visual inspection of Figure 3.3 does not suggest a pronounced difference. A Wilcoxon Signed-Rank test shows that we cannot statistically rule out the possibility that there is a similar prepositioned quantity between nice-to-have (low-importance) items in the cognitive consonant joint decision treatment (N_L^J) and the separate decision counterpart (N_L^S) ($T = 15$, $z = -1.27$, $p\text{-value}_{2\text{ tails}} = 0.2026$, $r = -0.28$). In addition, the effect size (in absolute value) has a low-to-middle importance since it is below the cut-off value of 0.30 (Field, 2009).

Hence, our results are fairly consistent with hypotheses H3A and H3B. Together, they suggest that participants (a) in a high-safety stock condition prepositioning high-importance items *lessen* the pull-to-center effect when making *cognitive consonant joint decisions* and (b) in a low-safety stock condition prepositioning low-importance items present a similar pull-to-center effect in *cognitive*

consonant joint and *separate decision* treatments. Our results suggest that the availability of a *consonant* reference in the *joint decision* treatment once more (*a*) influences participants' inventory prepositioning behavior for high-importance items and (*b*) does not for low-importance ones. The results for participants making cognitive consonant joint decisions also provide support to our conjecture that they pay more attention to the salient decisions (high-importance item) to the detriment of other ones (low-importance item).

3.4.5. Summary of results

Table 3.3 presents a summary of the overall results for our hypotheses (H1 – H3). The results suggest that participants treat inventory prepositioning decisions for high- and low-importance items in the same safety stock condition similarly in the separate decision treatments. Moreover, our results suggest that the availability of a *dissonant* and *consonant* reference in the *cognitive dissonant joint* and *cognitive consonant joint decision* treatments, respectively, (*a*) influences participants' inventory prepositioning behavior for high-importance items and (*b*) does not for low-importance ones. Our results show that inventory prepositioning decisions are influenced by bundled decision making. In particular, we find that for high-importance items (*a*) *cognitive consonant joint* decisions marginally *reduce* the pull-to-center effect; and (*b*) *cognitive dissonant joint* decisions *increase* the pull-to-center effect. We also find that low-importance items are not influenced by our dissonant or consonant treatments. These results are consistent with our hypotheses.

Table 3.3. Summary of hypothesis tests.

	H_0	<i>Statistic</i>	z	<i>p-value</i>	r	<i>Support?</i>
1A	$C_L^S - N_L^S = 0$	$W = 112$	0.14	0.8880 ^a	0.03	Consistent
1B	$C_H^S - N_H^S = 0$	$W = 123$	0.92	0.3600 ^a	0.20	Consistent
2A	$C_L^J - C_L^S \leq 0$	$T = 8$	-2.22	0.0131 ^b	-0.47	High
2B	$N_H^J - N_H^S = 0$	$T = 24$	-0.80	0.4236 ^a	-0.17	Consistent
3A	$C_H^J - C_H^S \leq 0$	$T = 13$	-1.48	0.0697 ^b	-0.33	Marginal
3B	$N_L^J - N_L^S = 0$	$T = 15$	-1.27	0.2026 ^a	-0.28	Consistent

^a 2-tailed p-value, ^b 1-tailed p-value.

3.5. DISCUSSION

This research presents the results of a Newsvendor experiment in which individuals make joint inventory repositioning decisions for two items of different importance. In such setting, individuals' decisions are prone to the pull-to-center effect commonly observed in traditional profit-based Newsvendor experiments. To influence their decisions, we design a framework to create either a cognitive dissonant treatment, by bundling a high-importance low-safety stock item with a low-importance high-safety stock one; or a cognitive consonant one, by bundling a high-importance high-safety stock item with a low-importance low-safety stock one. Our results show that inventory repositioning decisions are influenced by bundled decision making. In particular, we find that for high-importance items (a) *cognitive consonant joint* decisions *reduce* the pull-to-center effect; and (b) *cognitive dissonant joint* decisions *increase* the pull-to-center effect. We also find that low-importance items are not influenced by our dissonant or consonant treatments. We conjecture that individuals pay attention to high-importance items to the detriment of low-importance ones.

Our results have implications to the way Newsvendor inventory ordering decisions are potentially framed. In humanitarian settings, a cognitive consonant joint decision mechanism for perishable items (e.g., vaccines, pharmaceutical drugs, ready-to-use therapeutic foods) may be implemented combining decisions for critical items with non-critical ones. Both decisions may be bundled in order to reduce

the pull-to-center effect obtained with high-importance items, increasing not only decisions' cost-effectiveness but also items' availability. In industrial settings, a cognitive consonant joint decision framework may be implemented anytime a high-profit product is perceived as more important than a low-profit product. Both decisions may be bundled in order to increase the expected profits achieved on the high-profit product. In addition, a cognitive dissonant joint decision framework may be implemented when a strategic item is costly but critical (e.g., a low-profit part with long replenishment delays) in order to increase its availability and customer service satisfaction. Also, bundling inventory ordering decisions of different importance items in different safety stock conditions is presumably easier and less costly to implement than strategies that seek to align multiple partners and coordinate their decisions (e.g., a buyer and a supplier, or coordinating relief response across multiple organizations).

The proposed Newsvendor framework is aligned with the more practical case of the Newsstand problem (multi-product Newsvendor problem) and is closer to reality than previous experiments. Hence, it represents an improvement with respect to the external validity of previous experiments (Smith, 1982) and extends Newsvendor research according to Khouja's (1999) recommendations. The context of the experiment —inventory prepositioning in preparation to emergency response— is also novel and shows the application of the Newsvendor model in a non-traditional (non-profit based) operational setting, extending Newsvendor research as well.

Methodologically, we believe that the proposed framework can be generalized to traditional manufacturing and profit-based Newsvendor settings. The conditions for the framework to apply are established given a difference between the importance of items and their safety stock levels. Nevertheless, the novelty of the experimental setting imposes additional hurdles that could be addressed with future research. First, future research could explore whether individuals making inventory ordering decisions of different importance items stocked at the same safety stock level would also be influenced by the framework. Second, to generalize our findings, future research could explore how the proposed framework influences decisions in more traditional manufacturing and profit-based settings. Third, the treatments were conducted with a small number of humanitarian

practitioners. This is a difficult pool of participants to have access to. Future research could test more traditional pool of individuals such as undergraduate, graduate, or MBA students. In addition, one could compare the behavior of humanitarian practitioners to that of students to analyze whether previous experience with inventory repositioning decisions or humanitarian settings affects overall behavior. Fourth, we did not control for order of presentation effects between the separate and the joint decision treatments. The separate treatments were presented first to avoid priming individuals with a prior experimental reference. Future work could explore how a reference obtained from a prior joint decision treatment may affect separate decisions. Finally, to run a clean test of the effects of cognitive dissonance, the decision task assumed, e.g., certainty in disaster occurrence (cf. Lodree Jr. and Taskin, 2008), no correlation between items' demand (cf. Anupindi and Akella, 1993), and no resource constraints (cf. Lau and Lau, 1995, 1996). Future work could relax these assumptions one by one and in combination to explore how they affect the proposed effects.

We also acknowledge the novelty of the application of cognitive dissonance and the challenges that it imposes. First, the conjectures about consonance effects are exploratory. Future work could systematically explore consonance effects. Second, future research could systematically explore possible magnitude effects of reactions to dissonant or consonant states. *A priori* one may argue that unpleasant dissonant states may trigger stronger reactions than consonant ones; however, there are no theoretical grounds to make such claims. Further exploration of consonance effects and the strength of reactions to dissonant or consonant states could contribute to Festinger's (1957) cognitive dissonance theory.

Furthermore, our hypotheses testing the similarity in quantity amounts in (i) the separate decision treatments (H1A and H1B) and (ii) the low-importance items in the joint decision treatments (H2B and H3B) are stated in the traditional way, i.e., as the null (H_0). Our inability to reject them does not immediately allow us to accept them since we could be making a type II error—accepting the null when it is in fact false—. Power Analysis (Cohen, 1988) could potentially be used to ensure that the probability of a type II error would be negligible; however, the likely number of observations required

would be large (Cohen, 1990). Future work could use Power Analysis in the design of an improved experiment with a larger sample size.

In summary, our results suggest that a joint decision framework stressing cognitive dissonance may influence Newsvendor inventory ordering decisions. In particular, bundling Newsvendor inventory ordering decisions that differ in importance may help decision makers lessen (or strengthen) the traditional pull-to-center effect. We hope that this work can prompt further research and interest on possible mechanisms that can debias inventory ordering decisions in Newsvendor experiments.

Appendix 3.1. Sample of written instructions (joint decision treatment)

INSTRUCTIONS

DESCRIPTION OF THE GAME

You are a senior supply manager in a large Humanitarian Organization (HO) that provides relief during emergencies. When preparing for emergencies, you must preposition two kinds of emergency items:

- Critical-to-Life items: These items are critical to life; they must be available to prevent the loss of lives.
- Nice-to-Have items: These items are nice to have as they alleviate adverse conditions.

In this game, you must decide the quantity of Critical-to-Life and Nice-to-Have items to preposition (purchase) to prepare for regional emergencies. Your preposition decisions influences the time beneficiaries get the items they need and the cost that the HO must pay. For simplicity, for each kind of item, assume that each beneficiary demands exactly one item. Moreover, the quantities demanded for both kinds of items are independent of each other (i.e., they are not necessarily the same). If beneficiary demand is lower than the amount prepositioned, they get the items promptly (and the HO meets beneficiary demand at purchase cost). If beneficiary demand is higher than the amount

prepositioned, the excess demand will get the items with a delay (and the HO will pay an additional cost for expediting the excess items).

For each emergency, your preposition decisions must be made before you know for certain what the beneficiary demands are. Based on past emergencies, however, you know that the required number of items for each kind of item is uniformly distributed between 1 and 100 items. That is, there is a 1/100 chance that beneficiary demand will be 1, a 1/100 chance that beneficiary demand will be 2, and so on. Moreover, emergencies are independent of each other. That is, a small or large beneficiary demand in one emergency has no influence on whether beneficiary demand is small or large in future emergencies.

GOAL

Your goal is to minimize the total costs accumulated during the span of the game (60 decisions, 30 emergencies).

PLAYING THE GAME

To access the game, follow the link [*game link*].

DECISIONS

Please write down your preposition decisions (as you are making them) in the table provided (see reverse of the sheet).

[Decision table]

After completing all your preposition decisions, send the electronic data by e-mail by following the next steps:

- **Right click on the table with your preposition decisions** (the table on the upper right of the screen).
- **Select Copy Data.**
- **Open your preferred e-mail application and start writing a new e-mail.**
- **Right click on the body of the e-mail.**
- **Select Paste.**
- **Copy and Paste also the link of the game** (web address that appears in your Internet browser)
- **Send the e-mail to:** [e-mail address 1] and [e-mail address 2].

Appendix 3.2. Sample of game screen (T6)

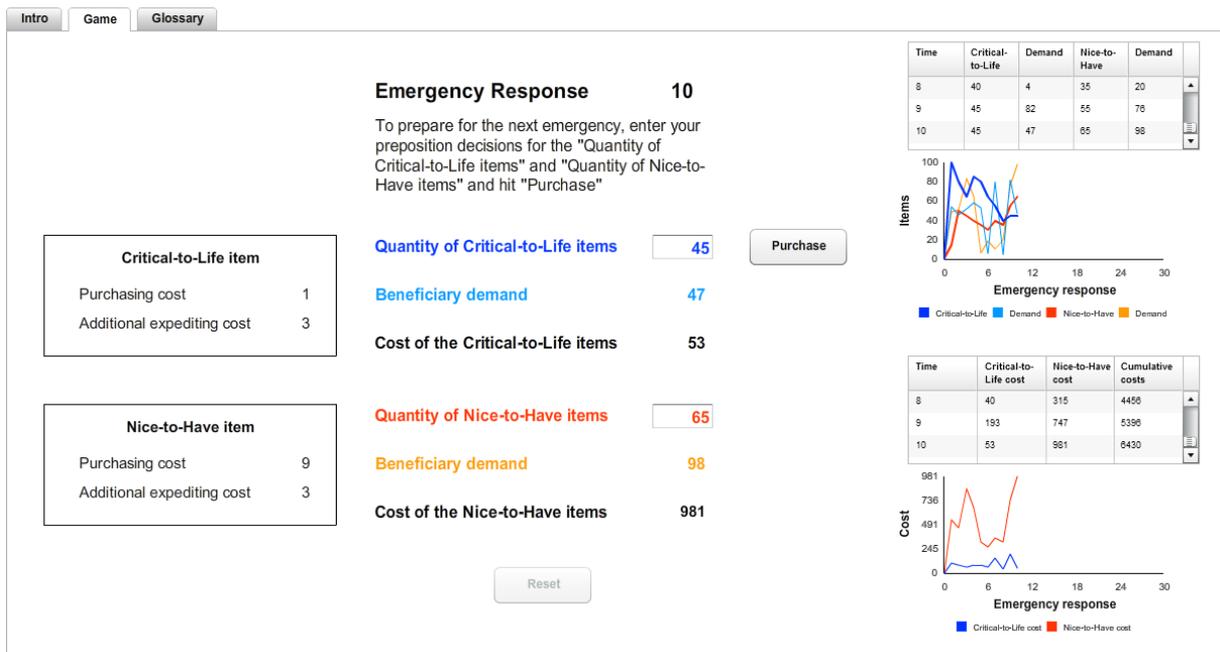


Figure A3.2.1. Sample of game screen.

Appendix 3.3. Statistical tests

Bootstrap confidence intervals

We rely on bootstrap confidence intervals since we have small sample sizes and hence cannot

guarantee that the samples conform to the assumptions needed to compute standard confidence intervals. Bootstrap methods estimate the properties of the sampling distribution from the sample data by treating the sample as a population from which smaller samples (bootstrap samples) are taken (putting the data back before a new sample is drawn). The statistic of interest is calculated in each sample, and by taking many samples the sampling distribution can be estimated. The standard error of the statistic is estimated from the standard deviation of this sampling distribution. From this standard error, confidence intervals and significance tests can be computed (Field, 2009). The reported bootstrap confidence intervals are bias-corrected bootstrap confidence intervals, for which 1'000 bootstrap replications were used as suggested when computing bias-corrected intervals (Efron and Tibshirani, 1986).

Wilcoxon Rank-Sum test

We rely on the non-parametric Wilcoxon Rank-Sum test since we have small sample sizes and hence cannot guarantee that the samples conform to the assumptions needed to run the independent (or unpaired data) *t*-test. The Wilcoxon Rank-Sum test is the non-parametric equivalent of the independent *t*-test. The test rests on the calculation of a statistic that compares the ranked data from the two samples of interest. When the data from both samples are ranked from lowest to highest ignoring the samples or groups to which the data belonged, then one would expect the higher ranks to be in one group and the lower ranks to be in the other group in case there was a difference between the groups; specifically, if one added up the ranks, then one would expect the summed total of ranks in each group to be different. When the groups have unequal number of observations in them, the test statistic *W* is the sum of ranks in the group that contains the fewer observations; it is the smaller summed rank otherwise (Field, 2009). To determine if the statistic is significant, the statistic can be converted to a *z*-score; $W \sim N(n_1(n_1+n_2+1)/2, \sqrt{n_1n_2(n_1+n_2+1)/12})$. It has been shown that the normal approximation appears appropriate very quickly (Bellera *et al.*, 2010).

Wilcoxon Signed-Rank test

The Wilcoxon Signed-Rank test is used since, besides having small sample sizes, $T1-T2$ are not independent from $T5$ given the within-subjects design along them (the same applies for $T3-T4$ and $T6$). The Wilcoxon Signed-Rank test is the non-parametric equivalent of the dependent (or paired data) t -test. The test rests on the calculation of a statistic that compares the ranked differences between the observations of interest. Once the differences have been calculated, they are ranked the same way as with the Wilcoxon Rank-Sum test, but the sign of the difference is assigned to the rank. Then, the ranks that came from a positive difference between the groups are collected and added up to get the sum of positive ranks T_+ . The same is done with the negative differences to get T_- . The test statistic T is the smaller of the two values (Field, 2009). To determine if the statistic is significant, the statistic can be converted to a z-score; $T \sim N(n(n+1)/4, \text{sqrt}(n(n+1)(2n+1)/24))$. It has been shown that the normal approximation appears appropriate very quickly (Bellera *et al.*, 2010).

CHAPTER 4

INVENTORY ORDERING DECISIONS IN A SINGLE ECHELON: THE EFFECT OF BACKORDERS

ABSTRACT

This paper studies supplier inventory ordering behavior in a Newsvendor extension to the case of backorders in order to assess (i) the effect of backorders on inventory ordering behavior in a simple inventory system and (ii) whether suppliers realize the benefits of an inventory system with backorders compared to one with lost sales. The paper presents results from a laboratory (lab) experiment that compares individuals' inventory ordering behavior in the Newsvendor extension to the case of backorders to that observed in the traditional —lost sales— Newsvendor inventory system in both low- and high-safety stock conditions. Consistent with a theoretical comparison of both inventory models, results show that backorders drive individuals' inventory ordering quantities upwards compared to lost sales. In addition, consistent with behavioral arguments based on reference dependence and misperceptions of feedback, results show that individuals react to shortages in a stronger manner when unmet demand is backlogged than when is lost and underweight backorders when making inventory ordering decisions, respectively. These findings suggest that suppliers may benefit in terms of profits and/or customer service satisfaction by backlogging rather than losing unmet demand.

Keywords: Backorders, Behavioral Operations Management, Laboratory Experiments, Lost Sales, Newsvendor Model.

4.1. INTRODUCTION

Inventory shortages occur when the amount of a given product in stock falls short of a customer's order. They are often an indicator of suboptimal supply chain performance (Lee and Lodree Jr., 2010)

and are usually classified as lost sales, backorders, or partial backorders—a fraction of the shortage is lost and the remaining fraction is backlogged—. Backorders incur increased administrative costs, potential emergency transportation costs, and cost of delayed revenue, among others. Lost sales costs are sometimes even more expensive than backorders costs given the opportunity cost of lost revenue and the loss of customer goodwill or loyalty associated with the former (Lodree Jr., 2007). Hence, suppliers frequently offer economic incentives to customers to place a backorder rather than risk losing sales (DeCroix and Arreola-Risa, 1998).

How suppliers should make inventory ordering decisions when unmet demand is backlogged is hence a relevant issue for business success. Accordingly, it has been largely addressed from a normative point of view. A number of studies have modeled optimal supplier inventory ordering behavior along with the option of emergency replenishments to fill backorders (e.g., Gallego and Moon, 1993; Khouja, 1996; Lodree Jr. *et al.*, 2008). Others have modeled optimal supplier inventory ordering behavior in non-competitive environments or those in which a customer either places backorders or leaves without making a purchase (e.g., Lee and Lodree Jr., 2010; Lodree Jr., 2007), whereas others have added to the analysis the option of offering incentives to customers to place backorders (e.g., Cheung, 1998; DeCroix and Arreola-Risa, 1998). Also, I am aware of one study that has addressed both issues—supplier inventory ordering behavior and customer incentives—in competitive environments or those in which a customer can switch to competing suppliers (Netessine *et al.*, 2006)⁹.

To the best of my knowledge, no previous work has tested behaviorally any of the previous models. However, behavioral research in Operations Management studying both the Newsvendor pull-to-center effect and the Beer Game bullwhip effect offers some insights about how individuals place inventory orders when unmet demand is backlogged. On the Newsvendor side, Bloomfield and Kulp

⁹ Although there are other normative studies addressing optimal supplier inventory ordering behavior in competitive environments (e.g., Gaur and Park, 2007; Liu *et al.*, 2007), their main interest is on analyzing customer switching behaviors. Hence, for the most part, they assume lost sales and do not consider the option of offering incentives to customers to place backorders.

(2013) studied how individuals react to product durability in a single-echelon inventory ordering lab experiment. They showed that just as Newsvendors tend to adjust orders insufficiently to over and under stocking costs when the product is perishable and unmet demand is lost (e.g., Bolton and Katok, 2008; Bostian *et al.*, 2008; Kremer *et al.*, 2010; Schweitzer and Cachon, 2000), they also tend to adjust orders insufficiently to inventory and backorders when the product is non-perishable and unmet demand is backlogged. On the Beer Game side, lab experiments have consistently shown that individuals tend to underweight orders in transit, ordering too much when orders in transit call for smaller orders and too little when orders in transit call for larger orders (e.g., Croson and Donohue, 2006; Steckel *et al.*, 2004; Sterman, 1989; Wu and Katok, 2006).

Both Newsvendor and Beer Game research have identified what appears to be a robust underweighting of backorders. Nevertheless, I add to this body of research by running a Newsvendor experiment in a simple yet informative experimental design that varies whether unmet demand is backlogged and product safety stock (or profitability) level, addressing two gaps. On the one hand, previous experiments (e.g., Croson and Donohue, 2006; Steckel *et al.*, 2004; Sterman, 1989; Wu and Katok, 2006) portray too complex inventory systems in which cross-echelon coordination, gaming, or some other unspecified dynamics driven by individuals' interactions are not accounted for and, hence, the causes of biased inventory ordering behavior cannot be clearly ascribed to them or to particular product and/or environmental characteristics (Bloomfield and Kulp, 2013). In order to run a clean test of the effect of backorders on inventory ordering behavior, I assume a single-echelon inventory system with no inventory accumulation, isolating further the effect of backorders on inventory ordering behavior.

On the other hand, Bloomfield and Kulp (2013) focused more on how individuals react to product durability and not to safety stock levels and, hence, they controlled for safety stock level by equating under and over stocking costs. In addition, to the best of my knowledge, Beer Game research has exclusively studied inventory ordering behavior when under stocking costs are greater than over stocking costs. However, inventory ordering patterns may differ across different safety stock levels as suggested in Newsvendor research (e.g., Bolton and Katok, 2008; Chen *et al.*, 2013; Ho *et al.*, 2010;

Schweitzer and Cachon, 2000). Hence, I assume two different safety stock conditions in order to assess the effect of an inventory system with backorders on inventory ordering behavior across different safety stock levels.

I thus run a Newsvendor experiment in a 2x2 between-subjects design with lost sales vs. backorders and low- vs. high-safety stock condition to assess the effect of an inventory system with backorders on inventory ordering behavior more accurately. I provide normative arguments based on Bulinskaya's (1964) Newsvendor problem extension to the case of backorders and behavioral arguments based on reference dependence- (Ho *et al.*, 2010), loss aversion- (Harinck *et al.*, 2007; Smith *et al.*, 2009) and misperceptions of feedback- (e.g., Bloomfield and Kulp, 2013; Croson and Donohue, 2006; Sterman, 1989) related behaviors to explain the effect of an inventory system with backorders on inventory ordering behavior. In doing so, I also provide a behavioral test of the Newsvendor problem with backorders (Bulinskaya, 1964), which has not been previously tested behaviorally to the best of my knowledge.

In addition to contribute to the Behavioral Operations Management literature by accounting more accurately for the effect of an inventory system with backorders on inventory ordering behavior, I also contribute to the literature on incentives to backorders. Although an inventory system with backorders may be more beneficial to business compared to an inventory system with lost sales (DeCroix and Arreola-Risa, 1998; Lodree Jr., 2007), research is lacking on whether suppliers actually realize the benefits of the former compared to the latter by modifying their inventory ordering behavior accordingly. Hence, I also add to the literature on incentive to backorders (e.g., DeCroix and Arreola-Risa, 1998; Lodree Jr., 2007; Netessine *et al.*, 2006) by comparing inventory ordering behavior in an inventory system with backorders to that of an inventory system with lost sales across different safety stock conditions to assess whether suppliers realize the benefits of an inventory system with backorders.

The rest of the paper is organized as follows. Section 4.2 describes the Newsvendor problem with backorders and develops hypotheses based on normative and behavioral arguments. Section 4.3 presents the lab experiment, describing its design and the experimental procedure. Section 4.4 presents

the main results and hypothesis tests. Finally, section 4.5 summarizes the work and discusses the main findings, implications, limitations, and opportunities for future research.

4.2. NEWSVENDOR PROBLEM WITH BACKORDERS

4.2.1. Normative implications

In the Newsvendor problem (Arrow *et al.*, 1951), a manager places an order quantity q at unit cost c facing an uncertain demand D in a single selling season. Once D is realized, the manager faces either an over stock or an under stock. If q exceeds D , then D units are sold and $q - D$ units incur a disposal cost h . That is, unit over stock cost h . For simplicity, and following previous Newsvendor experiments, I assume no disposal cost. That is, the only cost associated with an over stock is the purchasing cost c associated with the units in excess of demand. If D exceeds q , then q units are sold and $D - q$ units incur a shortage cost p . That is, unit under stock cost equals p . If D is a random variable with cdf F , it is well-known that the optimal inventory ordering quantity Q^* is a base-stock policy characterized by the critical fractile:

$$F_L(Q^*) = \frac{p - c}{p} \tag{4.1}$$

Schweitzer and Cachon (2000) define a product as a high-safety stock (or high-profit) product when $F_L(Q^*) \geq 1/2$ and as a low-safety stock (or low-profit) product otherwise.

In Bulinskaya's (1964) Newsvendor problem extension to the case of backorders, periods cannot be longer separated since a shortage in a given period carries over to the following period. However, when the shortage (backorder) cost p is charged per unit per unit time —backorder costs are assessed based on both the amount and length of backorders—, a base-stock policy is still optimal as shown by Bulinskaya (1964). In particular, the optimal inventory ordering quantity Q^* is characterized by the critical fractile:

$$F_B(Q^*) = \frac{p}{p + c} \quad (4.2)$$

Notice that both F_B and F_L change non-linearly to changes in the shortage cost p — $F_B'(p) = c/(p + c)^2$ and $F_L'(p) = c/p^2$ —. F_B also changes non-linearly to changes in the purchasing cost c — $F_B'(c) = -p/(p + c)^2$ —. F_L , on the other hand, changes linearly to changes in c — $F_L'(c) = -1/p$ —, serving hence as a simple yet informative reference to compare both inventory systems. Holding p constant, Figure 4.1 shows how increasing values of c affects both F_L and F_B .¹⁰

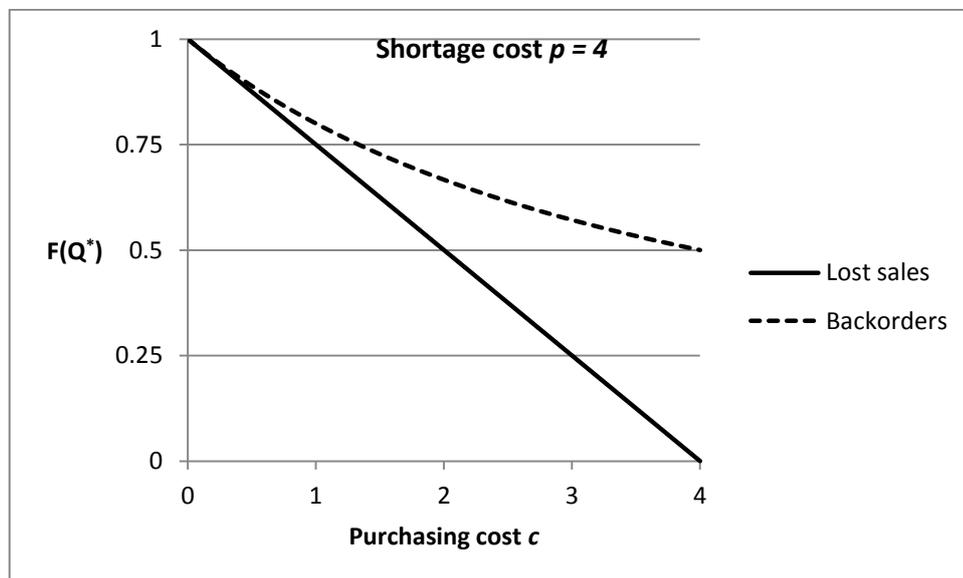


Figure 4.1. Optimality behavior to increasing values of the purchasing cost c .

Figure 4.1 reveals that Bulinskaya's (1964) Newsvendor problem extension to the case of backorders leads to larger optimal inventory orders than the traditional Newsvendor problem. Penalties due to backorders endures more than the ones due to lost sales since a shortage in a given period carries over to following periods until it is filled. Accordingly, the prospect of backorders

¹⁰ The value of c at which a given difference between both inventory systems is observed changes proportionally to changes in p . For instance, doubling p implies that the difference now observed for $c = 1$ will be observed for $c = 2$. In other words, the difference between both inventory systems is qualitatively invariant to p .

induces larger inventory orders to buffer the inventory system against the endurance of shortage penalties. Hence, it is reasonable to expect that the amount actually ordered in the backorders case will be larger than the one ordered in the lost sales case. In other words, it is reasonable to expect an *inventory system effect* in the same safety stock condition. This leads to the first hypothesis:

HYPOTHESIS 1: In the same safety stock condition, inventory ordering quantities in the backorders case will be larger than inventory ordering quantities in the lost sales case.

Figure 4.1 also reveals that reactions to the way unmet demand is handled are different in low- and high-safety stock conditions. First, as mentioned previously, penalties due to shortages endure more in the backorders than in the lost sales case. And second, not only does the supplier have to pay a shortage cost in the backorders case for every unit short of demand, but she also has to pay a purchasing cost associated with the units short of demand to fill the backlog, cost that is larger in low- than in high-safety stock conditions. Hence, overall, the cost of the backlog is larger in low- than in high-safety stock conditions due to the larger backlog filling cost in low-safety stock conditions. Taken together, the prospect of backorders induces larger differences in inventory orders with respect to lost sales in low- than in high-safety conditions to buffer the inventory system against the high purchasing costs of backorders. Hence, it is reasonable to expect that the amounts actually ordered will lead to a larger difference in inventory orders between backorders and lost sales in low- than in high-safety conditions. In other words, it is reasonable to expect a *larger inventory system effect* in low- than in high-safety stock conditions. This leads to the second hypothesis:

HYPOTHESIS 2: In the low-safety stock condition, differences in inventory ordering quantities between the backorders and lost sales cases will be larger than differences in inventory ordering quantities between the backorders and lost sales cases in the high-safety stock condition.

4.2.2. Behavioral implications

From a normative point of view, Bulinskaya's (1964) Newsvendor problem extension to the case of backorders should lead to larger inventory orders compared to the traditional Newsvendor problem. From a behavioral point of view, Ho *et al.*'s (2013) behavioral study of reference dependence in a multilocation Newsvendor problem offers some insights about the effects that backorders could have on inventory ordering behavior compared to lost sales. They showed that their proposed reference dependence model, which adds disutilities of over stocking and under stocking to traditional inventory multilocation models, explained their experimental data better than both standard inventory multilocation models and Schweitzer and Cachon's (2000) preference model for minimizing ex post inventory error. In a validation experiment, Ho and colleagues manipulated the relative salience of the disutilities of over stocks in a low-safety stock condition to reduce the pull-to-center effect by either asking individuals to compute and write down the amount of leftovers and their associated profit loss or imposing a cash penalty for leftovers. Similarly, they manipulated the relative salience of the disutilities of under stocks in a high-safety stock condition to reduce the pull-to-center effect by either asking individuals to compute and write down the amount of shortage and their associated forgone profits or awarding a cash bonus for having no shortage. Ho and colleagues showed that salient leftovers induced smaller inventory orders in low-safety stock conditions, whereas salient shortages induced larger inventory orders in high-safety stock conditions, proving that their salient disutility manipulations were effective in reducing the pull-to-center effect.

Compared to lost sales, backorders make shortages arguably more salient since they make shortages and their associated penalties to carry over to following periods until they are filled, making their detrimental effects to endure more in time. Building on Ho *et al.*'s (2013) salient disutility results, it is then reasonable to expect that shortages will lead to larger order adjustments when they are backlogged than when they are lost due to the salience that backorders arguably provide to shortages. In other words, it is reasonable to expect an *inventory system shortage effect* in the same safety stock condition. This leads to the third hypothesis:

HYPOTHESIS 3: In the same safety stock condition, order adjustments after a shortage in the backorders case will be larger than order adjustments after a shortage in the lost sales case.

In addition, loss aversion offers some insights about potential differences in reactions to the way unmet demand is handled in low- and high-safety stock conditions. Loss aversion refers to the phenomenon that the disutility of losses exceeds the utility of commensurate gains, i.e., losses loom larger than corresponding gains (Kahneman and Tversky, 1979). In the domain of losses, larger losses are presumably more important than smaller losses (Harinck *et al.*, 2007; Wilson and Gilbert, 2005) and consequently more likely to affect behavior (Smith *et al.*, 2009). Analyzing poker player behavior, Smith *et al.* (2009) observed that the majority of the analyzed players played more aggressively after a large loss than after a large win and that the fraction playing more aggressively consistently increased as the size of the large loss increased. A somewhat related finding is Harinck *et al.*'s (2007) study on reversed loss aversion for small amounts of money, in which they observed that, compared to large outcomes, loss aversion is reversed for small outcomes, i.e., gains loom larger than losses.

As mentioned previously, the cost associated with the backlog is larger in low- than in high-safety stock conditions due to the larger backlog filling cost in low-safety stock conditions. Building on loss aversion arguments (Harinck *et al.*, 2007; Smith *et al.*, 2009), it is then reasonable to expect that shortages will lead to larger order adjustments for backorders than for lost sales in low- than in high-safety stock conditions due to the larger cost associated with backlogged shortages in low-safety stock conditions. In other words, it is reasonable to expect a *larger inventory system shortage effect* in low- than in high-safety stock conditions. This leads to the fourth hypothesis:

HYPOTHESIS 4: In the low-safety stock condition, differences in order adjustments after a shortage between the backorders and lost sales cases will be larger than differences in order adjustments after a shortage between the backorders and lost sales cases in the high-safety stock condition.

Although an inventory system with backorders should induce large inventory orders, it is not clear whether individuals will be closer to the optimum compared to an inventory system with lost sales.

This is further supported by Beer Game research studying misperceptions of feedback, which has consistently shown that individuals tend to underweight orders in transit, ordering too much when orders in transit call for smaller orders and too little when orders in transit call for larger orders (e.g., Croson and Donohue, 2006; Steckel *et al.*, 2004; Sterman, 1989; Wu and Katok, 2006). In order words, individuals tend to underweight inventory and order too much and to underweight backorders and order too little. The underweighting of inventory and backorders has also been observed, although to a lesser extent, when POS data is available to all echelons (Croson and Donohue, 2003; Steckel *et al.*, 2004), when communication prior the game is allowed or inventory information is shared among echelons (Croson and Donohue, 2005, 2006; Wu and Katok, 2006), when transit lags are reduced and/or the number of echelons is less than four (Cantor and Katok, 2012; Steckel *et al.*, 2004), among others. That is, the underweighting of inventory and backorders appears robust (Croson *et al.*, 2013).

The underweighting of inventory and backorders has also been observed in simpler —single-echelon— inventory systems. In particular, Bloomfield and Kulp (2013) showed that just as Newsvendors tend to adjust orders insufficiently to over and under stocking when the product is perishable and unmet demand is lost (e.g., Bolton and Katok, 2008; Bostian *et al.*, 2008; Kremer *et al.*, 2010; Schweitzer and Cachon, 2000), they also tend to adjust orders insufficiently to inventory and backorders when the product is non-perishable and unmet demand is backlogged. Given that the underweighting of backorders is a robust component of biased inventory ordering behavior, it is then reasonable to expect an *underweighting of backorders* when the product is perishable and unmet demand is backlogged. This leads to the final hypothesis:

HYPOTHESIS 5: In the backorders case, inventory ordering quantities will not be fully adjusted to backorders.

4.3. LABORATORY EXPERIMENT

I follow Bloomfield and Kulp (2013) and run a single-echelon inventory ordering lab experiment to control for potential cross-echelon coordination, gaming, or some other unspecified dynamics driven by individuals' interactions not accounted for in Beer Game experiments. I further simplify

Bloomfield and Kulp's (2013) setup by assuming no inventory accumulation, providing thus a clean test for the effect of backorders on inventory ordering behavior. Unlike Bloomfield and Kulp (2013), I also include high- and low-safety stock levels since Newsvendor research suggests that behavioral effects may differ across different safety stock levels (e.g., Bolton and Katok, 2008; Chen *et al.*, 2013; Ho *et al.*, 2010; Schweitzer and Cachon, 2000), not to mention the different reactions of optimality to the way unmet demand is handled in low- and high-safety stock conditions observed in Figure 1.

4.3.1. Experimental design

I set unit shortage cost at $p = 4$ and manipulate unit purchasing cost c . In particular, I set unit purchasing cost for high-safety stock items at $c = 1$, and for low-safety stock items at $c = 3$. I consider an approximately normally distributed customer demand with a mean of $\mu = 50$ units and a standard deviation of $\sigma = 20$ units. Following Ho *et al.* (2010), I restrict the demand to positive integer values and use the term “approximately normal” instead of “truncated normal” to avoid confusing individuals. Information about the demand process was available in the instructions and was explained using the *empirical rule*. In addition, a graph of the demand process was also shown when reading the instructions.

The use of a non-uniform demand follows Su's (2008) recommendation of studying Newsvendor behavioral biases under non-uniform demand distributions (e.g., Ho *et al.*, 2010; Moritz *et al.*, 2013). In addition, $\mu = 50$ units and $\sigma = 20$ units assures that the coefficient of variation is large enough to make an impact and small enough for a normal distribution to be reasonable (Rudi and Drake, 2011). All individuals experienced realizations from the same set of demand values, controlling for the impact of demand realizations on inventory ordering decisions.

For the Newsvendor problem (Arrow *et al.*, 1951), the described parameterization implies optimal inventory ordering quantities of 64 units ($q_{L,H}^* = 64$) in a high-safety stock condition and 37 units ($q_{L,L}^* = 37$) in a low-safety stock condition. For the Newsvendor problem extension to the case of backorders (Bulinskaya, 1964), the described parameterization implies optimal inventory ordering quantities of 67 units ($q_{B,H}^* = 67$) in a high-safety stock condition and 54 units ($q_{B,L}^* = 54$) in a low-

safety stock condition. Notice that $q_{B,L}^*$ is not strictly in the domain of a low-safety stock condition according to Switcher and Cachon's (2000) definition since it is larger than μ . However, it serves as a reference to explore whether an inventory system with backorders leads to larger inventory ordering quantities than an inventory system with lost sales as stated previously. In addition, it is referred as a low-safety stock condition for ease of exposition.

To explore the impact of an inventory system with backorders on inventory ordering behavior, the experiment hence considers a 2x2 full factorial between-subjects design. The factors are inventory system, viz Newsvendor (lost sales hereafter) and Newsvendor extension to the case of backorders (backorders hereafter), and safety stock condition, viz low and high. Notation-wise, X_i , with $X \in (L, B)$, refers to lost sales (L) or backorders (B), where $i \in (L, H)$ refers to a low-safety stock (L) or a high-safety stock (H) condition. For example, L_L refers to lost sales low-safety stock items, whereas B_H to backorders high-safety stock items. Thus, the experiment considers four treatments:

- T1: lost sales low-safety stock item (L_L)
- T2: lost sales high-safety stock item (L_H)
- T3: backorders low-safety stock item (B_L)
- T4: backorder high-safety stock item (B_H)

4.3.2. Experimental procedure

A total of 96 individuals participated in the experiment. The analysis includes results from 89 participants. Seven participants were removed from the analysis —5 from B_L and 2 from B_H — since their inventory ordering behaviors suggest that they were not particularly responding to shortages, resulting in unusual large backlogs during most of the game¹¹. A robustness check to outliers at the end of the results section includes results from all participants. All participants were students attending a graduate Operations Management course in a Swiss university. The experiment was programmed

¹¹ A non-reported box plot analysis identifies 5 out of 7 of these participants as outliers using the 1.5IQR rule of thumb. A subsequent box plot analysis identifies the remaining 2 participants as outliers.

and run with the software z-Tree (Fischbacher, 2007). Table 4.1 shows the treatments with their corresponding notation and number of participants.

Table 4.1. Treatments, notation, and number of participants.

		<i>Safety stock condition</i>	
		Low	High
<i>Inventory system</i>	Lost sales	<i>T1 (L_L)</i> 22	<i>T2 (L_H)</i> 22
	Backorders	<i>T3 (B_L)</i> 24	<i>T4 (B_H)</i> 21

The experiment was run in two back-to-back sessions. First, lost sales treatments were run in two computer rooms, one for L_L and the other for L_H . Then, backorders treatments were run immediately after in the same computer rooms, one for B_L and the other for B_H . Participants arrived and were given the instructions (see Appendix 4.1), which were read aloud by an assistant. Participants had time to ask clarifying questions before initiating the experiment. After having read the instructions and answered any clarifying questions, the assistant initiated the experiment. Initially, participants were asked, through the experiment software, to answer a series of control questions to check they understood the instructions. Participants then played five practice rounds to get familiarized with the interface and the task. Following, they played the assigned treatment for 30 rounds aiming at minimizing cumulative costs. After participants entered their decisions, the system automatically revealed the demand realization, the corresponding lost sales (or backorders) or leftovers, and the corresponding costs. At any time, participants had access to information about unit purchasing and shortage costs c and p , respectively. Participants had also access to all previous decisions and outcomes, including demand realizations, lost sales (or backorders), leftovers, costs, and total cumulative costs (Appendix 4.2 shows a snapshot of the game screen). After having played the 30 rounds, participants were asked, through the experiment software, to answer a series of questions about

the information cues they were more inclined to use to inform their decisions. Monetary rewards were not used to incentivize participants.

4.4. RESULTS

4.4.1. Normative hypotheses

Before showing the formal hypothesis tests, I first show an overview of the average inventory ordering behavior for all treatments. Average inventory ordering behavior for a treatment is given by averaging *average inventory ordering quantities across rounds for each participant* across the number of participants in the treatment. Figure 4.2 provides 95% bootstrap confidence intervals of the average quantities ordered in all treatments¹². Lost sales results show the typical pull-to-center effect —average inventory ordering behavior falls between the expected demand and the optimum— in L_H (e.g., Bolton *et al.*, 2012; Kremer *et al.*, 2010; Schweitzer and Cachon, 2000) and a strong asymmetry of the pull-to-center effect —average inventory ordering behavior is above the expected demand— in L_L (e.g., Bolton and Katok, 2008; Bostian *et al.*, 2008; Schweitzer and Cachon, 2000). Backorders results show that average inventory ordering behavior in B_L is above optimum behavior, whereas it falls between the expected demand and the optimum in B_H . Comparing both inventory systems in the same safety stock condition, backorders appear to induce larger inventory orders than lost sales in both safety stock conditions. Following, I present the formal hypothesis tests.

¹² Non-reported normal Q-Q plots and Shapiro-Wilk tests suggest the four samples follow normal distributions. However, the small sample sizes cannot warrant that the samples conform to the assumptions needed to compute standard confidence intervals. Hence, I report bootstrap confidence intervals. Non-reported standard confidence intervals show qualitatively the same results.

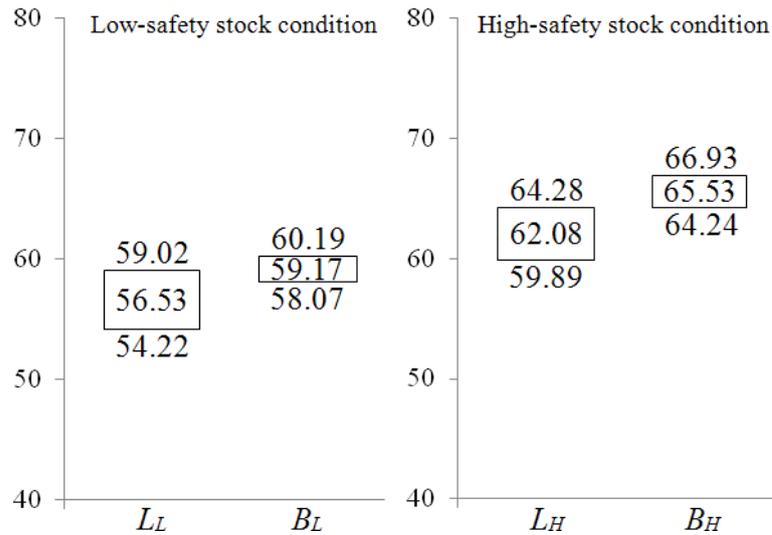


Figure 4.2. 95% bootstrap confidence intervals of average inventory ordering behaviors.

Building on Bulinskaya's (1964) Newsvendor problem extension to the case of backorders, I first test whether backorders lead to larger inventory ordering quantities than lost sales in the same safety stock condition (Hypothesis 1) by comparing population of participants' average inventory ordering quantities. Wilcoxon Rank-Sum tests show that average inventory ordering behavior in backorders is significantly larger than average inventory ordering behavior in lost sales in the low-safety stock condition ($W = 426$, $z = -2.00$, $p\text{-value}_{1\text{ tail}} = 0.0227$, $r = -0.30$), whereas highly significantly larger in the high-safety stock condition ($W = 563.5$, $z = 2.47$, $p\text{-value}_{1\text{ tail}} = 0.0068$, $r = 0.38$)¹³. Hence, there is a significant inventory system effect in both safety stock conditions, providing support for Hypothesis 1.

I next test whether there is a larger difference in inventory orders between backorders and lost sales in low- than in high-safety conditions (Hypothesis 2). This implies testing whether there is a difference between the difference between average inventory ordering behavior in B_L and L_L and the difference between average inventory ordering behavior in B_H and L_H —a difference between the two differences examined in Hypothesis 1—, which is essentially an interaction effect (Clogg *et al.*, 1995; Paternoster

¹³ Non-reported unpaired *t*-tests show qualitatively the same results for both comparisons.

et al., 1998) between inventory system and safety stock condition. I hence estimate the following regression model:

$$\bar{q}_i = \beta_0 + \beta_1 \text{InvSyst}_i + \beta_2 \text{SafStockCond}_i + \beta_3 \text{InvSyst}_i * \text{SafStockCond}_i + \varepsilon_i \quad (4.3)$$

where the dependent variable \bar{q}_i refers to the average inventory ordering quantity across rounds of participant i , InvSyst_i is a dummy for inventory system ($0 = \text{lost sales}$ and $1 = \text{backorders}$), SafStockCond_i is a dummy for safety stock condition ($0 = \text{low-}$ and $1 = \text{high-safety stock condition}$), $\text{InvSyst}_i * \text{SafStockCond}_i$ captures the interaction between inventory system and safety stock condition, and ε_i is the error term. A significantly negative coefficient for the interaction term $\text{InvSyst}_i * \text{SafStockCond}_i$ will indicate a larger inventory system effect in low- than in high-safety stock conditions. Table 4.2 shows the regression results.

Table 4.2. Regression of *larger inventory system effect* in low-safety stock conditions.^{a, b}

β_0 (Constant)	β_1 (InvSyst)	β_2 (SafStockCond)	β_3 (InvSyst*SafStockCond)	R^2	F	$p\text{-value}$
56.53*	2.6391 [†]	5.5495*	0.8112	0.3636	16.19	0.0000
(0.9530)	(1.3193)	(1.3477)	(1.8974)			

^a Standard errors in parentheses.

^b Significance for coefficients other than the constant is based on 1-tailed p-values.

* Highly significant, [†] Significant, [‡] Marginally significant.

Results of Hypothesis 1 suggest there could be a reversed effect from the predicted by Hypothesis 2—a larger inventory system effect in high- than in low-safety stock conditions—. Consistent to some extent with the results of Hypothesis 1, the coefficient for the interaction term $\text{InvSyst} * \text{SafStockCond}$ is directionally consistent; however, it is a non-significant effect ($\beta_3 = 0.8112$, $t(85) = 0.43$, $p\text{-value}_{1\text{ tail}} = 0.3350$). Hence, there is no a larger inventory system effect in low- than in high-safety stock conditions, providing no support for Hypothesis 2.

4.4.2. Behavioral hypotheses

Before showing the formal hypothesis tests, I first show an overview of the average order adjustment behavior after a shortage for all treatments. Average order adjustment behavior after a shortage for a treatment is given by averaging *average order adjustments after a shortage across shortage cases for each participant* across the number of participants in the treatment. To compute order adjustments after a shortage I check in round t whether there was a shortage case in round $t - 1$. If there was, I then compute and order adjustment as the difference between the order quantities in round t and round $t - 1$. For backorders, this could be an imprecise metric to the extent that it does not account for backlogged unmet demand in round $t - 2$ when assessing whether there was a shortage case in round $t - 1$. Hence, for backorders, I account for backlogged unmet demand in round $t - 2$ when computing order adjustments after a shortage.

Figure 4.3 provides 95% bootstrap confidence intervals of the average order adjustments after a shortage in all treatments¹⁴. Order adjustments after a shortage appear larger in backorders than in lost sales regardless of the safety stock condition. Comparing both inventory systems in the same safety stock condition, backorders appear to induce larger order adjustments after a shortage than lost sales in both safety stock conditions. Following, I present the formal hypothesis tests.

¹⁴ Non-reported normal Q-Q plots and Shapiro-Wilk tests suggest 2 out of 4 samples do not follow normal distributions. In addition, the small sample sizes cannot warrant that the remaining samples conform to the assumptions needed to compute standard confidence intervals. Hence, I report bootstrap confidence intervals. Non-reported standard confidence intervals show qualitatively the same results.

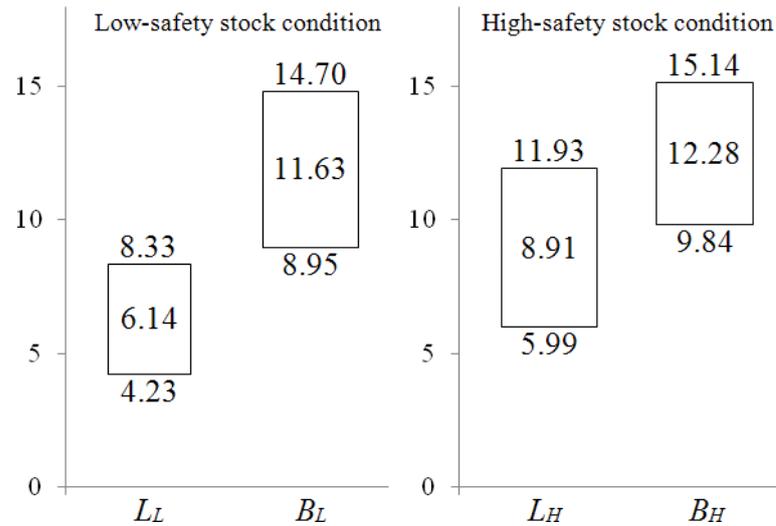


Figure 4.3. 95% bootstrap confidence intervals of average order adjustment behaviors after a shortage.

Building on reference dependence-related behavior (Ho *et al.*, 2010), I test whether backorders lead to larger order adjustments after a shortage than lost sales in the same safety stock condition (Hypothesis 3) by comparing population of participants' average order adjustments after a shortage. Wilcoxon Rank-Sum tests show that order adjustments after a shortage are highly significantly larger in backorders than in lost sales in the low-safety stock condition ($W = 386$, $z = -2.88$, $p\text{-value}_{1\text{ tail}} = 0.0020$, $r = -0.42$), whereas marginally significantly larger in the high-safety stock condition ($W = 527$, $z = 1.58$, $p\text{-value}_{1\text{ tail}} = 0.0571$, $r = 0.24$)¹⁵. Hence, there is a significant inventory system shortage effect in the low-safety stock condition only, providing partial support for Hypothesis 3.

Building on loss aversion-related behavior (Harinck *et al.*, 2007; Smith *et al.*, 2009), I next test whether there is a larger difference in order adjustments after a shortage between backorders and lost sales in low- than in high-safety conditions (Hypothesis 4). Analogous to Hypothesis 2, this implies testing whether there is a difference between two differences, which in this case corresponds to an interaction effect (Clogg *et al.*, 1995; Paternoster *et al.*, 1998) between inventory system shortage and safety stock condition. I hence estimate a regression model similar to (4.3); the only difference is the dependent variable, which now refers to the average order adjustment after a shortage across shortage

¹⁵ Non-reported unpaired *t*-tests show qualitatively the same results for both comparisons.

cases of participant i . A significantly negative coefficient for the interaction term

$InvSyst_i * SafStockCond_i$ will indicate a larger inventory system shortage effect in low- than in high-safety stock conditions. Table 4.3 shows the regression results.

Table 4.3. Regression of *larger inventory system shortage effect* in low-safety stock conditions.^{a, b}

β_0 (Constant)	β_1 (InvSyst)	β_2 (SafStockCond)	β_3 (InvSyst*SafStockCond)	R^2	F	p -value
6.14* (1.4077)	5.4864* (1.9489)	2.7673 [‡] (1.9908)	-2.1108 (2.8029)	0.1234	3.99	0.0104

^a Standard errors in parentheses.

^b Significance for coefficients other than the constant is based on 1-tailed p-values.

* Highly significant, [†] Significant, [‡] Marginally significant.

Results of Hypothesis 3 suggest there could be an effect consistent with the one predicted by Hypothesis 4. Consistent to some extent with the results of Hypothesis 3, the coefficient for the interaction term $InvSyst * SafStockCond$ is directionally consistent; however, it is a non-significant effect ($\beta_3 = -2.1108$, $t(85) = -0.75$, p -value_{1 tail} = 0.2267). Hence, there is no a significantly larger inventory system shortage effect in low- than in high-safety stock conditions, providing no support for Hypothesis 4.

Finally, following misperceptions of feedback (e.g., Croson and Donohue, 2006; Steckel *et al.*, 2004; Sterman, 1989; Wu and Katok, 2006), I test whether individuals adjust inventory ordering quantities insufficiently to backorders (Hypothesis 5). Building on Bloomfield and Kulp's (2013) misperceptions of feedback test, I estimate the following fix-effects panel regression model for each treatment:

$$q_{i,t} = \beta_0 + \beta_1 Shortage_{i,t-1} + \beta_2 Leftovers_{i,t-1} + v_i + \varepsilon_{i,t}, t = 2, \dots, 50 \quad (4.4)$$

where $q_{i,t}$ refers to the inventory ordering decision of participant i in period t , $Shortage_{i,t-1}$ captures the amount of lost sales (or backorders) of participant i in period $t - 1$, $Leftovers_{i,t-1}$ captures

the amount of leftovers of participant i in period $t - I$, v_i is the participants' effect, and $\varepsilon_{i,t}$ is the error term. The model allows identifying whether participants are insufficiently adjusting to lost sales (or backorders) and leftovers. Notice that optimal inventory ordering behavior would result in (i) β_0 having values of 37 and 64 for L_L and L_H , respectively, whereas values of 54 and 67 for B_L and B_H , respectively; (ii) β_1 having values of 0 and 1 in lost sales and backorders, respectively; and (iii) β_2 being 0 in both lost sales and backorders (see highlighted values in Table 4.4). That is, in lost sales participants should not adjust inventory ordering quantities neither to shortages nor to leftovers, whereas in backorders they should adjust inventory ordering quantities to shortages only. Table 4.4 shows the regression results.

Table 4.4. Fixed-effects panel regression of misperceptions of feedback.^a

	L_L		B_L		L_H		B_H	
β_0 (Constant)	56.42* (0.6487)	37	56.01* (0.9729)	54	60.99* (0.7811)	64	63.31* (0.7565)	67
β_1 (Shortage)	0.1254* (0.0362)	0	0.3758* (0.0333)	1	0.2326* (0.0539)	0	0.5368* (0.0486)	1
β_2 (Leftovers)	-0.0645* (0.0243)	0	-0.2616* (0.0440)	0	-0.0014 (0.0263)	0	-0.0618 [†] (0.0269)	0
R^2	0.3217		0.3101		0.2191		0.2945	
F	19.56		137.88		12.41		95.44	
p -value	0.0000		0.0000		0.0000		0.0000	

^a Standard errors in parentheses.

* Highly significant, [†] Significant, [‡] Marginally significant.

Consistent with misperceptions of feedback, participants under react to shortages in B_L . Specifically, each unit in the backlog changes inventory ordering quantities by 0.3758 units. The effect, although highly significant ($t(670) = 11.27$, $p\text{-value}_{2\text{ tails}} = 0.0000$), is below the optimal reaction ($t(670) = -18.72$, $p\text{-value}_{1\text{ tail}} = 0.0000$). A qualitatively similar result is observed in B_H . Specifically, each unit in the backlog changes inventory ordering quantities by 0.5368 units. The effect, although highly significant ($t(586) = 11.04$, $p\text{-value}_{2\text{ tails}} = 0.0000$), is also below the optimal reaction ($t(586) =$

-9.53 , $p\text{-value}_{1\text{ tail}} = 0.0000$). In addition, results show that participants react to shortages in both L_L ($\beta_1 = 0.1254$, $t(614) = 3.47$, $p\text{-value}_{2\text{ tails}} = 0.0006$) and L_H ($\beta_1 = 0.2326$, $t(614) = 4.31$, $p\text{-value}_{2\text{ tails}} = 0.0000$) despite the fact that shortages should not affect inventory ordering behavior when unmet demand is lost, which is consistent with results reported in Bloomfield and Kulp (2013). Hence, there is a significant underweighting of backorders in both backorders cases, providing support for Hypothesis 5.

4.4.3. Robustness check to outliers

As mentioned previously, 7 participants were removed from the analysis since their inventory ordering behaviors resulted in unusual large backlogs during most of the game, suggesting that they did not fully understand the implications of backlogging unmet demand. Here I repeat the hypothesis tests including the removed participants as a robustness check.

For Hypothesis 1 (*inventory system effect* in the same safety stock condition), results go from significant to non-significant between B_L and L_L , whereas from highly significant to significant between B_H and L_H . By including outliers, average inventory ordering behavior is driven downwards in both backorders cases, reducing the average difference between lost sales and backorders in the same safety stock condition. For Hypothesis 2 (*larger inventory system effect* in low- than in high-safety stock conditions), the magnitude of the effect remains reversed and increases from 0.8112 to 2.1201. By including outliers, average inventory ordering behavior seems to be more affected in B_L than in B_H , leading to the observed increased reversed effect. However, the effect remains non-significant.

For Hypothesis 3 (*inventory system shortage effect* in the same safety stock condition), results go from highly significant to significant between B_L and L_L , whereas from marginally significant to non-significant between B_H and L_H . By including outliers, average order adjustments after a shortage are driven downwards in both backorders cases, reducing the average difference between lost sales and backorders in the same safety stock condition. For Hypothesis 4 (*larger inventory system shortage effect* in low- than in high-safety stock conditions), the magnitude of the effect reduces from -2.1108

to -1.1496 . By including outliers, order adjustments after a shortage seem to be more affected in B_L than in B_H , leading to the observed reduced effect. In addition, the effect remains non-significant.

Finally, for Hypothesis 5 (*underweighting of backorders*), an underweighting of backorders remains highly significant in both safety stock conditions, but the magnitude of the effect is reduced from 0.5368 to 0.2570 in B_H , whereas from 0.3758 to 0.1251 in B_L . By including outliers, reaction to shortages is affected in both backorders cases, reducing the impact of prior shortages on inventory ordering behavior.

Summarizing, a robustness check to outliers suggests that the hypotheses most affected by outliers in terms of whether they are supported are Hypotheses 1 and 3 since one comparison in the former downgrades from significance to non-significance, whereas one in the latter downgrades from marginally significance to non-significance. Thus, Hypothesis 1 downgrades to partially supported, whereas Hypothesis 3 remains partially supported. Hence, the reported results are fairly robust to outliers.

4.5. DISCUSSION

I presented an experimental comparison of the traditional Newsvendor problem to its extension to the case of backorders in order to assess the effect of backorders on inventory ordering behavior more accurately and assess whether suppliers realize the benefits of an inventory system with backorders. Specifically, I compared differences in inventory ordering behaviors and in order adjustments after a shortage, and analyzed the extent to which participants underweight backorders.

Comparisons of inventory ordering behaviors reveals there is an inventory system effect in both safety stock conditions (H1), showing consistency with normative arguments developed based on a comparison of the optimality structures of both inventory systems (Arrow *et al.*, 1951; Bulinskaya, 1964). However, the analysis reveals there is no a larger different inventory system effect in low- than in high-safety stock conditions (H2), showing no consistency with additional normative arguments developed based on the referred inventory systems comparison. To assess more clearly the impact of backorders with respect to lost sales, I further compare the average distance to the optimal inventory

ordering quantity between B_L and L_L and between B_H and L_H . In low-safety stock conditions, participants are highly significantly closer to the optimum in backorders than in lost sales ($W = 663$, $z = 3.21$, $p\text{-value}_{1\text{ tail}} = 0.0007$, $r = 0.47$), whereas we cannot statistically rule out the possibility that the average distance to the optimal inventory ordering quantity between B_H and L_H is similar ($W = 482.5$, $z = 0.50$, $p\text{-value}_{1\text{ tail}} = 0.3092$, $r = 0.08$). In addition, following the same approach of Hypotheses 2 and 4, the effect is found to be significantly larger in low- than in high-safety stock conditions ($\beta_3 = 6.0741$, $t(85) = 1.97$, $p\text{-value}_{1\text{ tail}} = 0.0261$).

These results show that an inventory system with backorders is more beneficial than an inventory system with lost sales for low-safety stock items, suggesting that suppliers should prefer backorders over lost sales for costly or low-profit products. Although results do not show that an inventory system with backorders is more beneficial than an inventory system with lost sales for high-safety stock items, the former did lead to larger inventory ordering quantities than the latter. This result suggests that suppliers should prefer backorders over lost sales for cheap or high-profit products with high customer service expectations. Overall, the previous results suggest that suppliers may run higher performing businesses in terms of profits and/or customer service satisfaction by backlogging unmet demand instead of losing sales.

Comparison of order adjustments after a shortage reveals there is an inventory system shortage effect in low-safety stock conditions, whereas the effect is marginal in high-safety stock conditions (H3). Such results are consistent with reference dependence-related arguments, more specifically with salient disutilities (Ho *et al.*, 2010). Although there is a larger different inventory system shortage effect in low- than in high-safety stock conditions consistent with loss aversion-related arguments (Harinck *et al.*, 2007; Smith *et al.*, 2009), the analysis reveals the effect is not significant (H4). The previous results suggest that participants do adjust inventory ordering quantities to backorders. However, their inventory ordering quantities do not fully account for backorders as suggested by the underweighting of backorders observed in both B_L and B_H (H5). Such results are consistent with misperceptions of feedback-related arguments (e.g., Bloomfield and Kulp, 2013; Croson and Donohue, 2006; Sterman, 1989).

These results show that more salient shortages (backorders) influence inventory ordering behavior to a greater extent than less salient shortages (lost sales). Nevertheless, less salient shortages do influence inventory ordering behavior as shown by the misperceptions of feedback test. Although from a normative perspective backorders can be thought as the opposite of lost sales regarding the way unmet demand is handled, the previous results show that from a behavioral perspective they are related. In particular, and following the same line of reasoning of Ho *et al.* (2013), backorders do not eliminate the behavioral effect of lost sales but increase it in order to influence inventory ordering behavior in an intended direction. Although backorders do influence behavior in the intended direction, misperceptions of feedback show there is room for improvement even in the relatively simple inventory setting portrayed in this study.

Notwithstanding its contributions, the study has a number of limitations that future research could address. First, the experiment assumed full backlogging. In reality, customers may choose whether to place backorders (e.g., Lee and Lodree Jr., 2010; Lodree Jr., 2007). Hence, future work could study partial backlogging and assess its effect on inventory ordering behavior. Second, the experiment assumed no customer incentives to place backorders. Although this study shows that suppliers may realize the benefits of an inventory system with backorders, customers may not be willing to place backorders unless incentives are offered to them (e.g., Cheung, 1998; DeCroix and Arreola-Risa, 1998; Netessine *et al.*, 2006). Hence, future work could study the supplier's option of offering incentives to customers to place backorders and assess behaviorally its effect on inventory ordering behavior. Third, the backorder cost was assumed to be charged per unit per unit time following Bulinskaya's (1964) model. Backorder costs can also be time-independent, a fixed penalty, or a mix of them (e.g., Çetinkaya and Parlar, 1998; Ray *et al.*, 2010). Hence, future work could manipulate the way backorders costs are charged in order to assess behaviorally which cost structure is more beneficial for suppliers. Fourth, the experiment had no revenue metric and hence asked participants to minimize costs. Hence, future research could explore how the proposed framework influences inventory ordering decisions in a more traditional profit-based Newsvendor experiment. Finally, monetary rewards were not used to incentivize participants. On the one hand, it could be argued that

monetary rewards would improve results. On the other hand, the fact that backorders effects were observed may cast doubt on this observation. Nevertheless, future research could use monetary rewards for the sake of experimental rigor and analyze whether the use of incentives affects the observed effects.

In summary, in this study I tested behaviorally the Newsvendor problem extension to the case of backorders (Bulinskaya, 1964). I offered normative as well as behavioral arguments to explain the observed behavior and showed the benefits that suppliers may realize by backloging unmet demand, contributing thus to the literatures on Behavioral Operations Management and incentives to backorders. I hope this study can prompt further interest in behavioral tests of different backloging mechanisms and the response of both suppliers and customers to them.

Appendix 4.1. Sample of written instructions (B_H)

INSTRUCTIONS

The purpose of this session is to study how people make decisions in inventory management. From now until the end of the session, you are not allowed to talk with one another.

OBJECTIVE

Your goal is to **minimize** the **costs** you accumulate over 30 rounds of play.

TASK

You are a retailer who orders a single product or item from a supplier. In each decision round, you have to place **Orders** for items to satisfy an uncertain customer **Demand**.

COSTS

You order items from a supplier at a *purchasing cost* of **1** experimental francs (e\$) per item.

If **Order** < **Demand**, you incur a *backorder cost* of e\$ **4** per unit short of demand. Units short of demand carry over or accumulate as **Backorders** (equivalent to having negative inventory). Hence, they do affect following rounds. That is, if **Backorders** > **0**, your next **Order** will:

1. Be directed automatically to meet **Backorders**.
2. The remaining items will meet **Demand** for the round. If the remaining items are not enough to meet **Demand**, a (updated) backorder situation remains and you incur *backorder costs* accordingly.

For example, if **Backorders** = **10**, **Order** = **40**, and **Demand** = **60**, then **Order** meets **Backorders**, but the remaining items do not meet **Demand**, resulting in 30 **Backorders**.

$$\text{Purchasing cost} = \mathbf{1} \times 40 = \text{e\$ } 40,$$

$$\text{Backorder cost} = \mathbf{4} \times 30 = \text{e\$ } 120,$$

$$\text{Total cost} = 40 + 120 = \text{e\$ } 160.$$

Alternatively, if **Backorders** = **10**, **Order** = **60**, and **Demand** = **40**, then **Order** meets **Backorders**, and the remaining items do meet **Demand**, resulting in an excess of 10 units that are discarded at no cost.

$$\text{Purchasing cost} = \mathbf{1} \times 60 = \text{e\$ } 60,$$

$$\text{Backorder cost} = \mathbf{4} \times 0 = \text{e\$ } 0,$$

$$\text{Total cost} = 60 + 0 = \text{e\$ } 60.$$

If **Order** ≥ **Demand**, you discard the excess units at no cost. However, you will have incurred unnecessary *purchasing costs* for the excess units, yet excess units do not carry over or accumulate as inventory. Hence, they do not affect following rounds.

DEMAND

In each round, your ordering decision is made before you know with certainty what quantity of items your customers will demand. However, you know that demand is approximately normally distributed with mean 50 and standard deviation 20, i.e. 68% of the values lie between $50 - 20$ and $50 + 20$ (30, 70), 95% of the values lie between $50 - 40$ and $50 + 40$ (10, 90) and approximately 100% of the values lie between $50 - 50$ and $50 + 60$ (0, 110). In addition, demand is independent in each round, i.e. a small or large demand in earlier rounds has no influence on whether demand is small or large in later rounds.

Once you place your order, the computer selects the customer demand following the described demand distribution. You will receive demand and performance results for the round and history of play to date. The computer does not advance to the next decision round until all players are done with the current one.

[Page change]

Name: _____

Please write down the **Orders** you place in each round in the following table:

[Decision table]

Appendix 4.2. Sample of game screen (B_H)

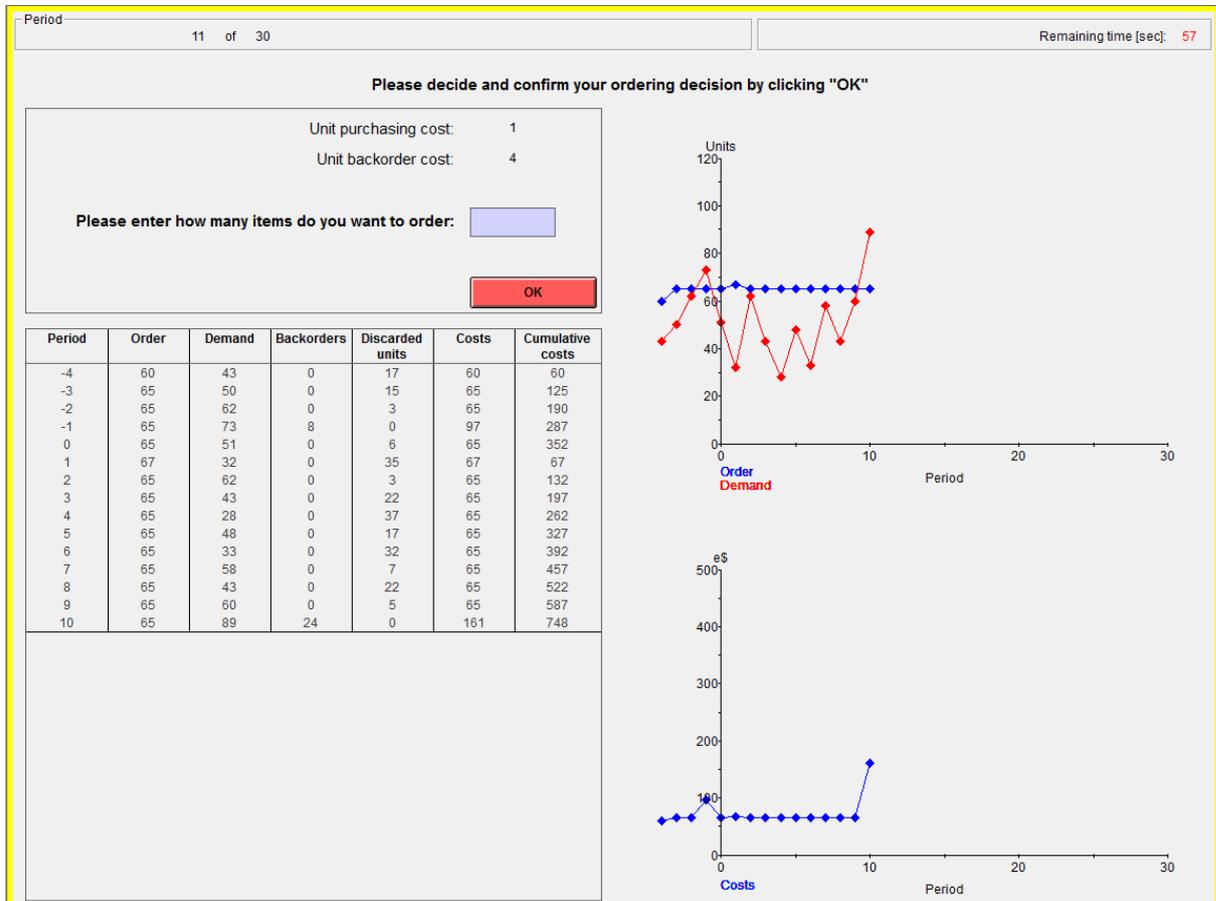


Figure A4.2.1. Sample of game screen.

CHAPTER 5

CONCLUSIONS

This dissertation presented the application of laboratory experiments to study biased Newsvendor ordering behavior. It contributes to the literature on Behavioral Operations Management (Bendoly *et al.*, 2006; Bendoly *et al.*, 2010; Gino and Pisano, 2008; Loch and Wu, 2007) by exploring the application of the Newsvendor model to a structurally similar decision making context, applying a well-known psychological theory as a debiasing mechanism for biased Newsvendor ordering behavior, and testing behaviorally a Newsvendor extension to the case of backorders. A discussion of each of these points is presented below, followed by a discussion of the limitations of the dissertation and opportunities for future research.

5.1. CONTRIBUTIONS

The first essay showed the application of the Newsvendor model to complexity level and resource allocation decisions in NPD projects under innovation uncertainty. It contributes to the analytical literature on NPD and innovation by presenting the Innovator model, a simple yet informative decision making model that can inform complexity level and resource allocation decisions in a simplified innovation pipeline, addressing thus the complex nature of previous analytical models (e.g., Loch and Terwiesch, 2005; Pich *et al.*, 2002). Moreover, it contributes to the Behavioral Operations Management literature by shedding light on the potential decision making biases NPD managers may be prone to. Building on insights from previous Newsvendor experiments and the results observed in the behavioral test of the Innovator model, this essay suggests that NPD managers may poorly understand decision making under innovation uncertainty. Consequently, they may underperform in demanding markets and over perform in less challenging ones, reducing their chances of bringing successful products to the market place.

Building on Festinger's (1957) cognitive dissonance theory, the second essay showed how dissonant and consonant states may be elicited in a Newsvendor setting to influence inventory ordering decisions in intended directions. It contributes to the Behavioral Operations Management

literature by applying Festinger's (1957) and Simon *et al.*'s (1995) cognitive dissonance and dissonance reduction arguments, respectively, as the foundation of a debiasing mechanism that pairs items' importance and safety stock condition in joint Newsvendor ordering decisions to debias or strengthen biased inventory ordering behavior for critical items. Building on the behavioral test of the debiasing mechanism, this essay suggests that consonant Newsvendor settings may help Newsvendors achieving higher profits and product availability for critical items, whereas dissonant Newsvendor settings may help Newsvendors achieving higher product availability for critical items.

Finally, the third essay tested behaviorally Bulinskaya's (1964) Newsvendor extension to the case of backorders, comparing it to the traditional lost sales Newsvendor model. It contributes to the Behavioral Operations Management literature by comparing both models theoretically and also behaviorally building on reference dependence (Ho *et al.*, 2010), loss aversion (Harinck *et al.*, 2007; Smith *et al.*, 2009), and misperceptions of feedback (e.g., Bloomfield and Kulp, 2013; Croson and Donohue, 2006; Sterman, 1989) arguments. Moreover, it also contributes to the literature on incentives to backorders by showing that backorders induce better inventory ordering behavior compared to lost sales. Building on the experimental comparison of both models, this essay suggests that suppliers may benefit in terms of both profits and product availability by backloging unmet demand instead of losing sales for costly items, whereas in terms of product availability for cheap items.

5.2. LIMITATIONS AND FUTURE WORK

Lab experiments offer the advantage of a tight control of confounding factors, allowing establishing cause-and-effect relationships. However, this comes at the expense of external validity. Accordingly, limitations of this dissertation go hand in hand with external validity concerns.

In the first essay we assumed a single stage-gate innovation pipeline under a single uncertainty source and built the Innovator model under this simplifying assumption. The Innovator model is intended to be a foundational model to inform decision making in NPD projects. However, we recognize that innovation pipelines are more complex since they typically comprise more than one

development stage and have several uncertainty sources affecting each screening process simultaneously (Cooper *et al.*, 1998). Hence, future work could add more than one development stage, more than one uncertainty source, and more than one decision, separately and jointly, to assess how increasing complexity levels of the innovation pipeline affect managerial performance. This will certainly complicate the tractability of the model, but has also the potential of offering more managerial insights to NPD managers.

In the second essay we assumed an unconstrained Newsstand (multi-item Newsvendor) setting to run a clean test of the proposed debiasing mechanism. Newsstand settings recognize that many times managers have to conduct a balancing act between how much to order from several competing products and their available resources (Abdel-Malek and Montanari, 2005). Accordingly, the Newsstand formulation comprises resource constraints. Hence, future work could add resource constraints (e.g., budget, storage capacity) and assess how these affect inventory ordering behavior in both consonant and dissonant decision frameworks.

We also assumed in the second essay that cognitive dissonance is the underlying psychological mechanism through which individuals change their inventory ordering decisions in joint decision frameworks. However, we lack an out-of-task measure of cognitive dissonance. Despite the control of confounding factors, it is difficult to determine if cognitive dissonance is the only psychological mechanism at work in our joint decision framework. Hence, future work could use an out-of-task measure of cognitive dissonance adapted to the particular case of our joint decision framework and test whether the measure explains the observed behavior (e.g., Moritz, 2010).

In the third essay we showed how backlogging unmet demand instead of losing sales is beneficial for suppliers. However, we did not consider the case of offering customer incentives to place backorders, which is usually how a backorders system is implemented in practice (e.g., Cheung, 1998; DeCroix and Arreola-Risa, 1998; Netessine *et al.*, 2006). Hence, future work could study how the option of offering customer incentives to place backorders affect inventory ordering behavior. In addition, in this essay as well as in the second one we assumed a cost-based inventory framework, i.e.,

there was no revenue metric. Hence, future work could test both frameworks in more traditional profit-based Newsvendor experiments.

Finally, monetary rewards were not used to incentivize participants. Although the protocols of Experimental Economics call for monetary rewards to incentivize participants (Smith, 1976, 1982), there is no systematic evidence showing that offering monetary rewards to incentivize participants leads to better performance (Arkes, 1991; Camerer and Hogarth, 1999). Given the social nature of the decision making task in the second essay —inventory prepositioning in preparation to emergency response— and the pool of participants —humanitarian practitioners—, it is unclear whether monetary rewards could have led to qualitatively different results. What it is clear is that results were consistent with the hypothesized effects, especially in the dissonant decision making framework. Although in the first essay we applied the Newsvendor framework to a new context, the context is structurally equivalent to the one portrayed in typical Newsvendor experiments. Moreover, results resembled those observed in previous Newsvendor experiments. The same holds in the third essay, in which lost sales results resembled those observed in previous Newsvendor experiments, whereas backorders results were qualitatively consistent with previous inventory management experiments that have found that backorders are underweighted. Hence, it does not seem safe to argue that monetary rewards would have led to qualitatively different results. However, in the spirit of experimental rigor, monetary rewards to incentivize participants should be used in extensions of this dissertation, especially in cases in which the decision framework has not been explored behaviorally before.

BIBLIOGRAPHY

- Abdel-Malek, L. L., Areeratchakul, N. 2007. A Quadratic Programming Approach to the Multi-Product Newsvendor Problem with Side Constraints. *European Journal of Operational Research*, 176(3): 1607-1619.
- Abdel-Malek, L. L., Montanari, R. 2005. On the Multi-Product Newsboy Problem with Two Constraints. *Computers & Operations Research*, 32(8): 2095-2116.
- Anupindi, R., Akella, R. 1993. Diversification under Supply Uncertainty. *Management Science*, 39(8): 944-963.
- Arkes, H. 1991. Cost and Benefits of Judgment Errors: Implications for Debiasing. *Psychological Bulletin*, 110(3): 486-498.
- Arrow, K. J., Harris, T., Marschak, J. 1951. Optimal Inventory Policy. *Econometrica*, 19(3): 250-272.
- Balcik, B., Beamon, B. M. 2008. Facility Location in Humanitarian Relief. *International Journal of Logistics: Research and Applications*, 11(2): 101-121.
- Beamon, B. M., Kotleba, S. A. 2006. Inventory Modelling for Complex Emergencies in Humanitarian Relief Operations. *International Journal of Logistics: Research and Applications*, 9(1): 1-18.
- Beinhocker, E. D. 1999. Robust Adaptive Strategies. *Sloan Management Review*, 40(3): 95-106.
- Bellera, C. A., Julien, M., Hanley, J. A. 2010. Normal Approximations to the Distributions of the Wilcoxon Statistics: Accurate to What N ? Graphical Insights. *Journal of Statistics Education*, 18(2): 17 pp.
- Bendoly, E., Croson, R., Gonçalves, P., Schultz, K. L. 2010. Bodies of Knowledge for Research in Behavioral Operations. *Production and Operations Management*, 19(4): 434-452.
- 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.
- Benzion, U., Cohen, Y., Peled, R., Shavit, T. 2008. Decision-Making and the Newsvendor Problem: An Experimental Study. *Journal of the Operational Research Society*, 59(9): 1281-1287.

- Benzion, U., Cohen, Y., Shavit, T. 2010. The Newsvendor Problem with Unknown Distribution. *Journal of the Operational Research Society*, 61(6): 1022-1031.
- Bloomfield, R. J., Kulp, S. L. 2013. Durability, Transit Lags and Optimality of Inventory Management Decisions. *Production and Operations Management*, in press.
- Bolton, G. E., Katok, E. 2008. Learning by Doing in the Newsvendor Problem: A Laboratory Investigation of the Role of Experience and Feedback. *Manufacturing & Service Operations Management*, 10(3): 519-538.
- Bolton, G. E., Ockenfels, A., Thonemann, U. W. 2012. Managers and Students as Newsvendors. *Management Science*, 58(12): 2225-2233.
- Bostian, A. A., Holt, C. A., Smith, A. M. 2008. Newsvendor “Pull-to-Center” Effect: Adaptive Learning in a Laboratory Experiment. *Manufacturing & Service Operations Management*, 10(4): 590-608.
- Bulinskaya, E. V. 1964. Some Results Concerning Optimum Inventory Policies. *Theory of Probability and Its Applications*, 9(3): 389-403.
- Cachon, G. P., Terwiesch, C. 2009. *Matching Supply with Demand: An Introduction to Operations Management*. New York, NY: McGraw-Hill/Irwin.
- Camerer, C. F., Hogarth, R. M. 1999. The Effects of Financial Incentives in Experiments: A Review and Capital-Labor Production Framework. *Journal of Risk and Uncertainty*, 19(1-3): 7-42.
- Cantor, D. E., Katok, E. 2012. Production Smoothing in a Serial Supply Chain: A Laboratory Investigation. *Transportation Research Part E*, 48(4): 781-794.
- Çetinkaya, S., Parlur, M. 1998. Optimal Myopic Policy for a Stochastic Inventory Problem with Fixed and Proportional Backorder Costs. *European Journal of Operational Research*, 110(1): 20-41.
- Chen, L., K k, A. G., Tong, J. D. 2013. The Effect of Payment Schemes on Inventory Decisions: The Role of Mental Accounting. *Management Science*, 59(2): 436-451.
- Cheung, K. L. 1998. A Continuous Review Inventory Model with a Time Discount. *IIE Transactions*, 30(8): 747-757.

- Clogg, C. C., Petkova, E., Haritou, A. 1995. Statistical Methods for Comparing Regression Coefficients between Models. *American Journal of Sociology*, 100(5): 1261-1293.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 1990. Things I Have Learned (So Far). *American Psychologist*, 45(12): 1304-1312.
- Cooper, L. P. 2003. A Research Agenda to Reduce Risk in New Product Development through Knowledge Management: A Practitioner Perspective. *Journal of Engineering and Technology Management*, 20(1-2): 117-140.
- Cooper, R. G., Edgett, S. J., Kleinschmidt, E. J. 1998. *Portfolio Management for New Products*. Reading, MA: Addison-Wesley.
- Croson, R., Donohue, K. L. 2003. Impact of POS Data Sharing on Supply Chain Management: An Experimental Study. *Production and Operations Management*, 12(1): 1-11.
- 2005. Upstream versus Downstream Information and Its Impacts on the Bullwhip Effect. *System Dynamics Review*, 21(3): 249-260.
- 2006. Behavioral Causes of the Bullwhip Effect and the Observed Value of Inventory Information. *Management Science*, 52(3): 323-336.
- Croson, R., Donohue, K. L., Katok, E., Serman, J. D. 2013. Order Stability in Supply Chains: Coordination Risk and the Role of Coordination Stock. *Production and Operations Management*, in press.
- Cui, Y., Chen, L. G., Chen, J., Gavirneni, S., Wang, Q. 2013. Chinese Perspective on Newsvendor Bias: An Exploratory Note. *Journal of Operations Management*, 31(1-2): 93-97.
- de Véricourt, F., Jain, K., Bearden, J. N., Filipowicz, A. 2013. Sex, Risk and the Newsvendor. *Journal of Operations Management*, 31(1-2): 86-92.
- DeCroix, G. A., Arreola-Risa, A. 1998. On Offering Economic Incentives to Backorder. *IIE Transactions*, 30(8): 715-721.
- Edgeworth, F. Y. 1888. The Mathematical Theory of Banking. *Journal of the Royal Statistical Society*, 51(1): 113-127.

- Efron, B., Tibshirani, R. 1986. Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, 1(1): 54-75.
- 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*. Stanford, CA: Stanford University Press.
- Field, A. 2009. *Discovering Statistics Using SPSS*. London: Sage.
- Fischbacher, U. 2007. z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics*, 10(2): 171-178.
- Gallego, G., Moon, I. 1993. The Distribution Free Newsboy Problem: Review and Extensions. *Journal of the Operational Research Society*, 44(8): 825-834.
- Gaur, V., Park, Y.-H. 2007. Asymmetric Consumer Learning and Inventory Competition. *Management Science*, 53(2): 227-240.
- Gavirneni, S., Isen, A. M. 2010. Anatomy of a Newsvendor Decision: Observations from a Verbal Protocol Analysis. *Production and Operations Management*, 19(4): 453-462.
- Gino, F., Pisano, G. 2008. Toward a Theory of Behavioral Operations. *Manufacturing & Service Operations Management*, 10(4): 676-691.
- Griffin, A. 1997. The Effect of Project and Process Characteristics on Product Development Cycle Time. *Journal of Marketing Research*, 34(1): 24-35.
- Harinck, F., Van Dijk, E., Van Beest, I., Mersmann, P. 2007. When Gains Loom Larger than Losses: Reversed Loss Aversion for Small Amounts of Money. *Psychological Science*, 18(12): 1099-1105.
- Harmon-Jones, E., Harmon-Jones, C. 2007. Cognitive Dissonance Theory After 50 Years of Development. *Zeitschrift für Sozialpsychologie*, 38(1): 7-16.
- Ho, T.-H., Lim, N., Cui, T.-H. 2010. Reference Dependence in Multilocation Newsvendor Models: A Structural Analysis. *Management Science*, 56(11): 1891-1910.
- Hobday, M. 1998. Product Complexity, Innovation and Industrial Organisation. *Research Policy*, 26(6): 689-710.

- Jauch, L. R., Kraft, K. L. 1986. Strategic Management of Uncertainty. *Academy of Management Review*, 11(4): 777-790.
- Kahneman, D. 2003. A Perspective on Judgment and Choice: Mapping Bounded Rationality. *American Psychologist*, 58(9): 697-720.
- Kahneman, D., Tversky, A. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2): 263-291.
- Katok, E. 2011. Laboratory Experiments in Operations Management. In Geunes, J. P. (Ed.), *Tutorials in Operations Research*. Hanover, MD: INFORMS, pp. 15-35.
- Khouja, M. 1996. A Note on the Newsboy Problem with an Emergency Supply Option. *Journal of the Operational Research Society*, 47(12): 1530-1534.
- 1999. The Single-Period (News-Vendor) Problem: Literature Review and Suggestions for Future Research. *Omega*, 27(5): 537-553.
- Kim, J., Wilemon, D. 2003. Sources and Assessment of Complexity in NPD Projects. *R&D Management*, 33(1): 15-30.
- Kremer, M., Minner, S., van Wassenhove, L. N. 2010. Do Random Errors Explain Newsvendor Behavior? *Manufacturing & Service Operations Management*, 12(4): 673-681.
- Krishnan, V., Ulrich, K. T. 2001. Product Development Decisions: A Review of the Literature. *Management Science*, 47(1): 1-21.
- Lau, H.-S., Lau, A. H.-L. 1995. The Multi-Product Multi-Constraint Newsboy Problem: Applications, Formulation and Solution. *Journal of Operations Management*, 13(2): 153-162.
- 1996. The Newsstand Problem: A Capacitated Multiple-Product Single-Period Inventory Problem. *European Journal of Operational Research*, 94(1): 29-42.
- Lee, H., Lodree Jr., E. J. 2010. Modeling Customer Impatience in a Newsboy Problem with Time-Sensitive Shortages. *European Journal of Operational Research*, 205(3): 595-603.
- Liu, L., Shang, W., Wu, S. 2007. Dynamic Competitive Newsvendors with Service-Sensitive Demands. *Manufacturing & Service Operations Management*, 9(1): 84-93.

- Loch, C. H., Terwiesch, C. 2005. Rush and Be Wrong or Wait and Be Late? A Model of Information in Collaborative Processes. *Production and Operations Management*, 14(3): 331-343.
- Loch, C. H., Wu, Y. 2007. Behavioral Operations Management. *Foundations and Trends in Technology, Information and Operations Management*, 1(3): 121-232.
- Lodree Jr., E. J. 2007. Advanced Supply Chain Planning with Mixtures of Backorders, Lost Sales, and Lost Contract. *European Journal of Operational Research*, 181(1): 168-183.
- Lodree Jr., E. J., Kim, Y., Jang, W. 2008. Time and Quantity Dependent Waiting Costs in a Newsvendor Problem with Backlogged Shortages. *Mathematical and Computer Modelling*, 47(1-2): 60-71.
- Lodree Jr., E. J., Taskin, S. 2008. An Insurance Risk Management Framework for Disaster Relief and Supply Chain Disruption Inventory Planning. *Journal of the Operational Research Society*, 59(5): 674-684.
- 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.
- Lynn, G. S., Morone, J. G., Paulson, A. S. 1996. Marketing and Discontinuous Innovation: The Probe-and-Learn Process. *California Management Review*, 38(3): 8-36.
- McCarthy, I. P., Tsinopoulos, C., Allen, P., Rose-Andersen, C. 2006. New Product Development as a Complex Adaptive System of Decisions. *Journal of Product Innovation Management*, 23(5): 437-456.
- Milkman, K. L., Rogers, T., Bazerman, M. H. 2008. Harnessing Our Inner Angels and Demons: What We Have Learned About Want/Should Conflicts and How That Knowledge Can Help Us Reduce Short-Sighted Decision Making. *Perspectives on Psychological Science*, 3(4): 324-338.
- Moenaert, R. K., Souder, W. E. 1990. An Information Transfer Model for Integrating Marketing and R&D Personnel in New Product Development Projects. *Journal of Product Innovation Management*, 7(2): 91-107.

- Moritz, B. B. 2010. *Cognition and Heterogeneity in Supply Chain Planning: A Study of Inventory Decision Making*. Ph.D. Dissertation, The Graduate School, University of Minnesota, Minneapolis, MN.
- Moritz, B. B., Hill, A. V., Donohue, K. L. 2013. Individual Differences in the Newsvendor Problem: Behavior and Cognitive Reflection. *Journal of Operations Management*, 31(1-2): 72-85.
- Mullins, J. W., Sutherland, D. J. 1998. New Product Development in Rapidly Changing Markets: An Exploratory Study. *Journal of Product Innovation Management*, 15(3): 224-236.
- Murmann, P. A. 1994. Expected Development Time Reductions in the German Mechanical Engineering Industry. *Journal of Product Innovation Management*, 11(3): 236-252.
- Netessine, S., Rudi, N., Wang, Y. 2006. Inventory competition and incentives to back-order. *IIE Transactions*, 38(11): 883-902.
- Osterloh, M., Frey, B. S. 2002. Does Pay for Performance Really Motivate Employees? In Neely, A. (Ed.), *Business Performance Measurement: Theory and Practice*. Cambridge: Cambridge University Press, pp. 107-122.
- Paternoster, R., Brame, R., Mazerolle, P., Piquero, A. 1998. Using the Correct Statistical Test for the Equality of Regression Coefficients. *Criminology*, 36(4): 859-866.
- Perry, J. L., Mesch, D., Paalberg, L. 2006. Motivating Employees in a New Governance Era: The Performance Paradigm Revisited. *Public Administration Review*, 66(4): 505-514.
- Pich, M. T., Loch, C. H., De Meyer, A. 2002. On Uncertainty, Ambiguity, and Complexity in Project Management. *Management Science*, 48(8): 1008-1023.
- Rawls, C. G., Turnquist, M. A. 2010. Pre-Positioning of Emergency Supplies for Disaster Response. *Transportation Research Part B*, 44(4): 521-534.
- Ray, S., Song, Y., Verma, M. 2010. Comparison of Two Periodic Review Models for Stochastic and Price-Sensitive Demand Environment. *International Journal of Production Economics*, 128(1): 209-222.
- Read, D., Loewenstein, G., Rabin, M. 1999. Choice Bracketing. *Journal of Risk and Uncertainty*, 19(1-3): 171-197.

- Rudi, N., Drake, D. 2011. Observation Bias: The Impact of Demand Censoring on Newsvendor Level and Adjustment Behavior. Working Paper No. 12-042, Harvard Business School, Boston, MA.
- Schrader, S., Riggs, W. M., Smith, R. P. 1993. Choice over Uncertainty and Ambiguity in Technical Problem Solving. *Journal of Engineering and Technology Management*, 10(1-2): 73-99.
- 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.
- Simon, L., Greenberg, J., Brehm, J. 1995. Trivialization: The Forgotten Mode of Dissonance Reduction. *Journal of Personality and Social Psychology*, 68(2): 247-260.
- Smith, G., Levere, M., Kurtzman, R. 2009. Poker Player Behavior After Big Wins and Big Losses. *Management Science*, 55(9): 1547-1555.
- Smith, V. L. 1976. Experimental Economics: Induced Value Theory. *American Economic Review*, 66(2): 274-279.
- 1982. Microeconomic Systems as an Experimental Science. *American Economic Review*, 72(5): 923-955.
- Sobek II, D. K., Ward, A. C., Liker, J. K. 1999. Toyota's Principles of Set-Based Concurrent Engineering. *Sloan Management Review*, 40(2): 67-83.
- Steckel, J. H., Gupta, S., Banerji, A. 2004. Supply Chain Decision Making: Will Shorter Cycle Times and Shared Point-of-Sale Information Necessarily Help? *Management Science*, 50(4): 458-464.
- Sterman, J. D. 1989. Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment. *Management Science*, 35(3): 321-339.
- Su, X. 2008. Bounded Rationality in Newsvendor Models. *Manufacturing & Service Operations Management*, 10(4): 566-589.
- Taskin, S., Lodree Jr., E. J. 2010. Inventory Decisions for Emergency Supplies Based on Hurricane Count Predictions. *International Journal of Production Economics*, 126(1): 66-75.
- Tatikonda, M. V., Rosenthal, S. R. 2000. Technology Novelty, Project Complexity, and Product Development Project Execution Success: A Deeper Look at Task Uncertainty in Product Innovation. *IEEE Transactions on Engineering Management*, 47(1): 74-87.

- Thaler, R. 1980. Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior & Organization*, 1(1): 39-60.
- 1985. Mental Accounting and Consumer Choice. *Marketing Science*, 4(3): 199-214.
- Thomas, A. S. 2003. Humanitarian Logistics: Enabling Disaster Response. White Paper, Fritz Institute, San Francisco, CA.
- Thomas, A. S., Kopczak, L. R. 2005. From Logistics to Supply Chain Management: The Path Forward in the Humanitarian Sector. White Paper, Fritz Institute, San Francisco, CA.
- Thomas, L. J., McClain, J. O., Robinson, L. W., Schultz, K. L. 2007. The Use of Framing in Inventory Decisions. Research Paper No. 02-07, Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, NY.
- Thomke, S. H. 1998. Managing Experimentation in the Design of New Products. *Management Science*, 44(6): 743-762.
- 2008. Learning by Experimentation: Prototyping and Testing. In Loch, C. H., Kavadias, S. (Eds.), *Handbook of New Product Development Management*. Oxford: Butterworth-Heinemann, pp. 401-420.
- Tong, J., Song, J.-S. 2011. Reference Prices and Transaction Utility in Inventory Decisions. Working paper, The Fuqua School of Business, Duke University, Durham, NC.
- Tversky, A., Kahneman, D. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157): 1124-1131.
- Wilson, T. D., Gilbert, D. T. 2005. Affective Forecasting: Knowing What to Want. *Current Directions in Psychological Science*, 14(3): 131-134.
- Wu, D. Y., Katok, E. 2006. Learning, Communication, and the Bullwhip Effect. *Journal of Operations Management*, 24(6): 839-850.
- van der Laan, E. A., de Brito, M. P., Vergunst, D. A. 2009. Performance Measurement in Humanitarian Supply Chains. *International Journal of Risk Assessment and Management*, 13(1): 22-45.