universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

Preferences and behavioral decision-making

in operations-related settings

verfasst von / submitted by

Tamás Csermely

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2016 / Vienna 2016

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 094 151 403
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Logistics and Operations Management
Betreut von / Supervisor:	Univ.-Prof. Dr. Stefan Minner

Preface

This thesis was inspired by the discrepancies between normative theory and actually observed behavior in operations and economics-related settings. I am especially thankful for my supervisor, Prof. Dr. Stefan Minner, for drawing my attention to this interesting research agenda. In addition, I would like to thank him for supporting me among others in creating ideas, improving the text of my manuscripts and providing critical and constructive guidance to my work. Furthermore, I would also like to thank Prof. Dr. Richard Hartl for supporting my Ph.D. studies and for providing assistance in the administrative process.

Moreover, I am grateful to the behavioral and experimental economics community in Vienna. Prof. Dr. Rupert Sausgruber made it possible for me to work as a research and teaching associate at the Vienna University of Economics and Business in a supportive, challenging and highly professional environment for two and a half years. I acquired valuable knowledge in experimental methodology as well as professional insights from him, from Prof. Dr. Jean-Robert Tyran and Prof. Dr. Wieland Müller. I am also grateful to my supervisor and my co-authors, Assoc. Prof. Dr. Tony Haitao Cui and Alexander Rabas. This thesis is my own work, however, I will consistently use "we" upon presenting the findings out of gratitude, since this work would not have been possible without the help and the regular consultations.

I am honored that the Vienna Center for Experimental Economics allowed me to run my experiments in their laboratory. Both the University of Vienna and the Vienna University of Economics and Business had an enormous role in funding my experiments. The overall cost of the 4 experiments was slightly over €11,000 which came directly from Project S10307-G14 by the Austrian Science Fund (FWF) and from the Dean's Office at the Faculty of Economics, Business and Statistics at the University of Vienna. Furthermore, I received travelling allowance in a similar order of magnitude from the University of Vienna and the Vienna University of Economics and Business. Therefore I could disseminate my results on renowned conferences where I got to know several scholars in the behavioral operations management community. In particular Prof. Dr. Mirko Kremer, Prof. Dr. Enno Siemsen, Assoc. Prof. Dr. Kenneth Schultz and Assist. Prof. Dr. Brent Moritz helped me a lot in getting involved in this community.

Last but not least, I am very thankful to my parents and my sister for their ongoing help during the course of my Ph.D. studies. I received great support and encouragement during the ups and downs my path was paved with.

Abstract

This thesis studies behavioral factors in operations – primarily procurement-related – settings. It extends the current literature in providing knowledge on how people order products and make rationing decisions in different scenarios compared to the existing normative theory. The primary goal is to understand which preferences and biases are responsible for these decisions.

Experimental economics provides several ways to measure preferences in various domains, in most cases even multiple methodologies. On the pattern of risk preferences, we intend to shed light on how consistent and stable they are in a relatively short time frame. In the end, we provide a blueprint how to measure preferences. Furthermore, we consider several preferences jointly in the workhorse operations setting: the newsvendor problem, which is a repeated ordering game with perishable items. Besides quantifying the role of various individual preferences, we focus on how and why order quantities change over periods. The thesis also considers ordering behavior when the nature of supply is limited and products are rationed by a central warehouse. Furthermore, we extend our focus to settings where multiple supply modes exist and provide results on the usage pattern of a fast, expensive and a slow, cheap supplier.

The thesis concludes that preference elicitation mechanisms are very susceptible to the used methodology: Several of them deliver considerably different results across a 30-minute time frame. Our results show the connection between novel factors – most importantly income, extraversion, loss aversion – and ordering decisions. We also show that the existing history of past stochastic operations-related events' outcome shapes the behavior in the upcoming periods. Moreover, we provide multiple potential explanations why people do not tend to inflate their orders if they are aware that there is a bottleneck in the supply source. In such settings, warehouses ration their stock based on the retailers' order quantities and not their demand distribution under a wholesale price contract. Nevertheless, this changes somewhat when warehouses receive a profit share. Finally – and most surprisingly, we investigate the role of different supply modes. We document that an additional fast supply option can coordinate effectively and it leads to great cost savings – even if it is very expensive. All in all, we are convinced that this thesis provides relevant lessons stemming from human behavioral patterns when it comes to supply chain design, procurement practices and hiring decisions.

Zusammenfassung

Diese Dissertation befasst sich mit Verhaltenseinflüssen auf Operational Settings, primär im Zusammenhang mit Bestellverhalten. Die bestehende Literatur wird im Bezug auf die bisherige normative Theorie um das Wissen erweitert, wie Wirtschaftssubjekte Produkte bestellen und Zuteilungsentscheidungen unter verschiedenen Szenarien treffen. Das Hauptziel dieser Arbeit besteht in der Beantwortung der Frage, welche Präferenzen und Verhaltenstendenzen für diese Entscheidungen von Relevanz sind.

In der Verhaltensökonomie gibt es viele Möglichkeiten, Präferenzen in unterschiedlichsten Domänen zu messen, meistens mit Hilfe von mehreren Methoden. Unser Ziel ist es, eine Empfehlung abzugeben, wie jene Präferenzen konsistent und stabil zu messen sind. Daher betrachten wir mehrere unterschiedliche Arten von Präferenzen gleichzeitig im Newsvendor-Setting. Wir konzentrieren uns hierbei auf zwei Themen: Die Quantifizierung von individuellen Präferenzen, sowie wie und warum sich Bestellmengen im Laufe der Zeit verändern. Diese Dissertation betrachtet ebenso den Spezialfall des Bestellverhaltens wenn das Angebot beschränkt ist, oder wenn Produkte extern von einem zentral gesteuerten Warenhaus bereitgestellt werden. Wir erweitern unsere Analyse außerdem auf Szenarien mit mehreren Bestellmöglichkeiten für Händler, im Speziellen mit schnellen und teuren sowie langsamen und billigen Lieferanten.

Wir kommen zu dem Schluss, dass Methoden, die Präferenzen messen, stark abhängig von der verwendeten Methode sind, und viele von ihnen wesentliche Unterschiede bei mehrmaligen Messungen innerhalb von 30 Minuten liefern. Unsere Resultate zeigen Verbindungen zwischen mehreren Arten von Präferenzen, im Besonderen von Einkommen, Extrovertiertheit, Verlustaversion und Bestellmengen im Newsvendor-Setting. Wir zeigen, dass die von Subjekten gemachten rein stochastischen Erfahrungen das zukünftige Verhalten beeinflussen. Des Weiteren geben wir mehrere Erklärungen ab, warum Subjekte nicht genügend auf Engpässe im Angebot reagieren. Dies geschieht, wenn Warenhäuser ihren Bestand nach den georderten Mengen der Händler und nicht nach der Nachfrageverteilung bestimmen. Am Überraschendsten war allerdings die Untersuchung unterschiedlicher Bestellmöglichkeiten für Händler: Trotz ihres hohen Preises führt eine zusätzliche schnelle aber teure Bestellmöglichkeit für Händler zu hohen Ersparnissen. Im Großen und Ganzen können wir wertvolle Lektionen über menschliche Verhaltensmuster bezüglich des Designs von Lieferketten oder des Bestellverhaltens lernen.

Contents

1. Introduction 11
2. Methodology 16
3. How Sensitive Are (Risk) Preferences to Elicitation Mechanisms? 20
   3.1. Introductory Remarks 20
   3.2. Multiple Price Lists Explained 22
   3.3. Literature Review and Hypotheses 25
      3.3.1. Standard Gamble Methods 26
      3.3.2. Paired Gamble Methods 27
      3.3.3. Questionnaire Methods 27
      3.3.4. Comparison Studies 28
      3.3.5. Further Considerations 29
   3.4. Experimental Design 31
   3.5. Results 35
      3.5.1. The Overall Distributions are Different 36
      3.5.2. Th Rank Correlations are Low 38
      3.5.3. Method Similarities 39
      3.5.4. Method Quality Indicators 40
      3.5.5. Results Conclusion 46
   3.6. Conclusion 46
4. Biased Probability Judgment and Other Bias in the Newsvendor Setting 48
   4.1. Introductory Remarks 48
   4.2. Literature Review and Hypotheses 50
   4.3. Experimental Design 55
      4.3.1. Part One 56
      4.3.2. Part Two 56
      4.3.3. Part Three 58
      4.3.4. Questionnaire 58
   4.4. Results 59
      4.4.1. Novel Factors 59
      4.4.2. Biased Probability Judgment and Demand Chasing 63
      4.4.3. Robustness Checks 67

4.5. Conclusions 68
5. Behavior in Inventory Rationing Games 71
5.1. Introductory Remarks 71
5.2. Literature Review 73
5.3. The Model and the Hypotheses 76
5.4. Experimental Design 79
5.5. Results 82
5.5.1. Size Effects 82
5.5.2. The Warehouses' Rationing Mechanism 83
5.5.3. Profit Implications 87
5.5.4. Shortage Gaming 89
5.6. Conclusion 92
6. Behavioral Decisions in Inventory Systems with Dual Sourcing Option 94
6.1. Introductory Remarks 94
6.2. Literature Review 96
6.3. The Dual Sourcing Setting and Research Hypotheses 97
6.4. The Theoretical Models 99
6.4.1. The Normative Benchmark 99
6.4.2. Alternative Behavioral Theories 100
6.5. The Experimental Design 101
6.6. Main Results 103
6.6.1. Cost Comparisons 103
6.6.2. Supplier Usage 106
6.6.3. Robustness Checks 108
6.7. The Ordering Policy 109
6.7.1. Supply Line Underweighting 109
6.7.2. Order Smoothing 111
6.7.3. The Random Error Model 112
6.7.4. Joint Effects 115
6.8. Conclusion 116
7. General Implications 118
8. Appendix 120
8.1. How Sensitive Are (Risk) Preferences to Elicitation Mechanisms? 120

8.1.1. Instructions 120
8.1.2. Control Questions 123
8.1.3. Decision and Feedback Screen 125
8.1.4. Risk Elicitation Methods 126
8.1.5. Robustness Checks 130
8.1.6. Tests Under Different Functional Forms 132
8.2. Biased Probability Judgment and Other Bias in the Newsvendor Setting 138
8.2.1. General Instructions 138
8.2.2. Elicitation Procedure for the Gambler's and Hot-hand Fallacies 139
8.2.3. Elicitation Procedure for Regret Aversion 140
8.2.4. Elicitation Procedure for Loss and Risk Aversion 141
8.2.5. Elicitation Procedure for Anchoring and Overconfidence 142
8.2.6. The Newsvendor Task 142
8.2.7. Elicitation Procedure for Personality Traits 144
8.2.8. Elicitation Procedure Cognitive Reflection 145
8.2.9. The Correlation Matrix of Elicited Preferences 145
8.3. Behavior in Inventory Rationing Games 146
8.3.1. Instructions 146
8.3.2. Decision Screens 152
8.3.3. Shortage Gaming 153
8.4. Behavioral Decisions in Inventory Systems with Dual Sourcing Option 155
8.4.1. Instructions 155
8.4.2. Control Questions 159
8.4.3. Decision and Feedback 159
8.4.4. Cognitive Reflection Task 160
8.4.5. Risk and Loss Attitude Elicitation Task 160
8.4.6. Regret Aversion Elicitation Task 162
8.4.7. Decision and Feedback Screenshots 162
8.4.8. Validity of the Psychological Cost Model 166
8.4.9. Further Robustness Checks 168
9. Bibliography 172

List of Tables

Table 3.1: Risk Parameter Intervals a la Holt and Laury (2002) 22
Table 3.2: Method Overview 26
Table 3.3: Link between MPL Representation and the Literature 31
Table 3.4: Pairwise Wilcoxon Tests for Equality of Distributions 37
Table 3.5: Spearman Rank Correlation Coefficients 38
Table 3.6: Similarities across All Methods 39
Table 3.7: Explanatory Power 41
Table 3.8: Deviations from Predictions 42
Table 3.9: Stability Measures 44
Table 4.1: Descriptive Statistics of Elicited Preferences 57
Table 4.2: Regressions Introducing Significant Elicited Factors Separately 60
Table 4.3: Regressions Introducing Elicited Factors Jointly – High Margin Setting 61
Table 4.4: Regressions Introducing Elicited Factors Jointly – Low Margin Setting 62
Table 5.1: Profit Levels and Statistical Significance 87
Table 6.1: Comparison of Normatively Achievable and Actually Realized Costs 103
Table 6.2: Single-sourcing Versus Dual Sourcing 105
Table 6.3: The Relationship between Cognitive Ability and the Actual Ordering Policy 106
Table 6.4: Actual and Model-predicted Supplier Usage 107
Table 6.5: Optimal and Actual Coefficients of Variations of Order Quantities 109
Table 8.1.1: SGp Method 126
Table 8.1.2: SGhigh Method 126
Table 8.1.3: SGlow Method 127
Table 8.1.4: SGsure Method 127
Table 8.1.5: SGall Method 127
Table 8.1.6: PGp Method 128
Table 8.1.7: PGhigh Method 128
Table 8.1.8: PGlow Method 128
Table 8.1.9: PGall Method 129
Table 8.1.10: PGhigh Method – Bottom-Up Appearance 129
Table 8.1.11: Correlation Coefficients between the Methods 130
Table 8.1.12: Testing for Order Effects 130
Table 8.1.13: Comparison of Results to Previous Studies 131

Table 8.1.14: Explanatory Power – CARA 132
Table 8.1.15: Explanatory Power – Prospect Theory 133
Table 8.1.16: Explanatory Power – DRRA 133
Table 8.1.17: Explanatory Power – IRRA 134
Table 8.1.18: Explanatory Power – DARA 134
Table 8.1.19: Explanatory Power – IARA 135
Table 8.1.20: Explanatory Power – CARA plus Probability Weighting 135
Table 8.1.21: Explanatory Power – Cumulative Prospect Theory 136
Table 8.1.22: Explanatory Power – DRRA plus Probability Weighting 136
Table 8.1.23: Explanatory Power – IRRA plus Probability Weighting 137
Table 8.1.24: Explanatory Power – DARA plus Probability Weighting 137
Table 8.1.25: Explanatory Power – IARA plus Probability Weighting 138
Table 8.2.1: Correlation Matrix of Elicited Individual Factors in the High Margin Setting . 145
Table 8.2.2: Correlation Matrix of Elicited Individual Factors in the Low Margin Setting .. 146
Table 8.4.1: Fit of Models Compared to Model 3 170

List of Figures

Figure 3.1: Distributions of Risk Preferences 36
Figure 3.2: Differences in Switching Points between the First and the Second Time 44
Figure 4.1a: Demand Chasing of Subjects Having Biased Probability Judgments 66
Figure 4.1b: Demand Chasing of Subjects Having Unbiased Probability Judgment 66
Figure 5.1: The Number of Periods When One Supplier Got Fully Satisfied 82
Figure 5.2: The Basis of Warehouses' Rationing Mechanism Across Treatments 84
Figure 5.3: Archetype of Warehouses Based on the Exhibited Rationing Mechanism 86
Figure 5.4: Actual Versus Optimal Profit Levels 88
Figure 5.5: Average Order Quantities Per Treatment 89
Figure 5.6: The Average Order Quantity Over Time in Treatment ID-RS 90
Figure 6.1a and 6.1b: Coefficients of Variation Across Different Groups 114
Figure 8.1.1: Sample Screen 120
Figure 8.1.2 Sample Screen with Color-coding 121
Figure 8.1.3: Decision-making Screen for Lotteries 125
Figure 8.1.4: Revision Screen for Lotteries 125
Figure 8.3.1: Decision Screen for Warehouses in ID Treatments 152
Figure 8.3.2: Decision Screen for Retailers in ID Treatments 152
Figure 8.3.3: Average Order Quantities Over Time in DIFF-WP 153
Figure 8.3.5: Average Order Quantities Over Time for the Small Retailer in DIFF-WP 153
Figure 8.3.6: Average Order Quantities Over Time for the Large Retailer in DIFF-WP 154
Figure 8.3.7: Average Order Quantities Over Time Without Rationing 154
Figure 8.4.1: Decision Screen for the Dual Sourcing Treatments 162
Figure 8.4.2: Feedback Screen for the Dual Sourcing Treatments 163
Figure 8.4.3: Decision Screen for the Single Sourcing Treatment 163
Figure 8.4.4: Feedback Screen for the Single Sourcing Treatment 164
Figure 8.4.5: Elicitation of Risk Aversion 164
Figure 8.4.6: Elicitation of Loss Aversion 165
Figure 8.4.7: Lottery Choice 165
Figure 8.4.8: Elicitation of Regret Aversion 166
Figure 8.4.9: Parametric Analysis of Psychological Costs for Treatment D7T 167
Figure 8.4.10: Parametric Analysis of Psychological Costs for Treatment D10T 168

Chapter 1

1. Introduction

Several decision-makers encounter the problem how to match supply with demand nowadays. On the one hand, customers want their needs satisfied in a short time frame, at a high quality and a reasonable cost. On the other hand, companies have to find a feasible trade-off among such requirements in order to achieve success. The academic literature has provided several models and normative solutions as benchmarks for such operations-related problems, however, it has neglected the human factor until the dawn of the millennium. The aim of this thesis is to consider the human factor and its motives behind these decisions. As a starting point, this thesis takes each company as an entity into account where decisions are formed and shaped by individuals. Most importantly, we seek to answer the following questions:

- How many products do people source from their supplier under stochastic demand?
- Which supplier do they prefer if there are multiple vendors?
- What are the underlying conscious preferences or subconscious behavioral biases that shape the observed behavior?
- How should these preferences be measured?
- How much do people order in supply chains with limited supplies and how do people ration in these settings?
- To what extent do behavioral models provide a better explanation for the observed behavior?
- What are the cost and profit implications of the observed behavior?

Thus, the objective of this dissertation is to extend the knowledge on behavioral factors in various operations-related settings by comparing actual results to the normative benchmark and behavioral models. The thesis is structured into eight parts. Chapters 3 to 6 present the main results, which have recently been accepted or are currently in different stages of review.

Chapter 2 provides a short introduction on experimental methodology and explains why it is necessary to conduct empirical and behavioral research to answer our research questions. In addition, it outlines the advantages and drawbacks that experiments have in operations and economics-related settings.

Chapter 3 deals with the problem of stability and consistency of preferences which we investigate on the pattern of risk preferences. The problem is motivated by the extensive literature on the connection between elicited preferences and several settings with economic relevance. A complete assessment of multiple price list preference elicitation methods is conducted. We conjecture that methods which vary the high payment option and which present a choice between two gambles have the highest predictive power. We solidify these results with various measures of within-method consistency and elaborate on the connection between risk aversion – eventually its consistency – and the measurement methodology. We relax our assumption on expected utility theory and conclude that results do not change if we take probability weighting, prospect theory or rank-dependent utility theory into account. This chapter provides insight on how risk preference elicitation methods should be designed. Consequently, we extend our findings to other domains such as time preferences, loss aversion, willingness to pay or any measurement method that uses lists. The research paper (Csermely and Rabas, 2016) that is based on this chapter has been published in the Journal of Risk and Uncertainty (VHB Jourqual 3.0: B).

Our motivation for Chapter 4 originates in the fact that the vast majority of past behavioral operations studies focuses on specific single factor explanations and remain inconclusive on the competing, complementary and collectively exhaustive nature of behavioral biases, preferences and their profit implications. In order to achieve this goal, we link individual preferences to the behavior in the workhorse operations-related problem: The newsvendor setting. Thus, this section considers a single-echelon with repeated ordering decisions and a previously known demand distribution. Previous investigations show fluctuations of order quantities over time. We ask what exactly drives this widely documented within-subject variation of order quantities across periods. In order to resolve these issues, we present a laboratory experiment that incorporates several new factors and which controls for the previously documented ones. Intuitively, our findings show that people with a higher self-reported income, less loss aversion and an extroverted personality outperform their peers in the newsvendor setting. Most interestingly, we observe that biased probability judgment that stems from the previous realization of stochastic events is responsible for demand chasing. Subjects who believe that a recent event makes it more (less) likely that a similar event will occur in the next period are biased in line with the hot hand (gambler's) fallacy. Consequently, they tend to adjust their newsvendor order quantity towards (away from) the previous demand realization. In addition, the extent of within-subject order quantity adjustments is explained well by the

corresponding out-of-context measure for biased probability judgment. Our findings have implications for the focus of de-biasing efforts. The corresponding paper (Csermely and Minner, 2016) that is based on this section of the thesis received a revise and resubmit (major revision) for the second time from Decision Sciences (VHB Jourqual 3.0: B) and the paper converges well in the process. In line with the reviewers', the senior and the associate editors' recommendation, we will rewrite the paper and include additional analyses. The resubmission will occur in the beginning of 2017.

Chapter 5 is the author's personal favorite section, since it explores a less isolated and unexplored setting. It complements the previous chapter on perishable products and considers supply chains with limited supplies. We extend our focus to a two-echelon setting with two retailers posting newsvendor ordering decisions to the same warehouse. The retailers serve different markets – e. g. regional distribution network – and the warehouse rations the available products among them based on their demand distribution, the amount the retailers ordered and the profit they achieved in the previous period. If there is a wholesale price contract between the warehouse and the retailers, the warehouse allocates based on the retailers' orders, not their demand distribution. Furthermore, the warehouses do not make adjustments based on the retailers' previously achieved profit. Consequently, no consistent fairness preferences are documented even if there are no costs attached to such a behavioral pattern. If the warehouse receives a share of the retailers' revenue, the rationing occurs mostly based on the demand distribution, and to a moderate extent based on the actually ordered quantities similarly to the case of the wholesale price contract. In parallel, the exhibited behavior does not suggest any prevailing fairness preferences. The potential explanation is that the retailers cannot reject the warehouse's allocation proposal and that the retailers' demand distributions or their actual orders are clear reference points for the warehouses. These findings are key distinctions from the previous operations-related studies that documented fairness. Counterintuitively, even though the warehouses allocate their products to a large extent based on the retailers' actual orders, we do not document shortage gaming from the retailers' side. This implies that the retailers do not exaggerate their orders even in the presence of limited supplies when there is competition for the same resource opposed to previous studies. This might be driven by the differences in our design compared to previous studies: The retailers' stochastic demand, the human role of the warehouse, the lack of knowledge on the rationing mechanism or that we have repeated interaction across parties. Our insight for managers is that, in an environment with limited supplies, decision-makers tend to neglect the ex ante identical nature of

downstream supply chain members and ration using the retailers' order ratio as anchors. In experimental research, the actually exhibited behavior is never obvious in advance – especially in novel settings. Consequently, some of our results seem surprising in this section. As we did not expect them, our experimental design is not able to shed light on the origin of a few results in this section: It can only to provide a list of potential behavioral drivers. Therefore, we decided to design and run further treatments in the future that will answer the absence of fairness preferences and the lack of shortage gaming. We plan to run these treatments together with Prof. Dr. Stefan Minner and Assoc. Prof. Dr. Tony Haitao Cui.

Chapter 6 explores a setting that is the closest to what the author does during his every day procurement job. Thus, this setting has pronounced real world implications, similarly to Chapter 5. This part is motivated by the fact that companies tend to source the same products from multiple supply modes, yet we lack any evidence on how exactly people behave in such settings. Therefore, this section covers an ordering game similar to the newsvendor setting that was considered in Chapter 4, but with slightly different specifications. The most pronounced differences are that we have non-perishable products across periods and two supply modes. We examine individuals' ordering decisions: In particular, we are interested in how many products they order from an expensive supplier which can deliver instantly and from a different and less expensive vendor, which has a lead time of 1 period. We find that the fast supply mode is highly overused compared to what normative theory predicts, however, the difference between the normatively achievable and the actually observed cost decreases considerably compared to the setting when only the slow supply mode was present. This manifests in large cost savings in the end. We also find that the overusage of the fast supplier is driven by reference-dependent preferences and is explained by psychological costs that subjects attach to inventory shortages and leftovers. In general, subjects tend to smooth their order quantities across periods. They also tend to choose a preferred supply mode – which can be different across periods – instead of the optimal strategy, which would be combining orders from the two modes in most cases. Furthermore, the individual variation in cognitive reflection drives cost differences: Subjects with higher cognitive reflection have less random errors, less order smoothing and tend to use the fast option less frequently in their ordering pattern. Besides these interesting findings, the main driver of cost differences is that subjects who have higher cognitive ability take a larger fraction of products that are en route into account. These products have already been shipped but not yet received by the decision-maker. Our findings highlight the importance of having procurement professionals with high cognitive reflection, since they underweight the supply

line to a smaller extent. We also show that the benevolent nature of the additional supply mode stems primarily from its lower lead time. The corresponding practical implication is that the supply base should be broadened with vendors that have different characteristics, since decision-makers can utilize these options relatively well. The research article (Csermely and Minner, 2015) based on this section of the thesis was rejected from Management Science (VHB Jourqual 3.0: A+) with an encouragement to rewrite the paper and resubmit to the journal. We were asked to alter the focus of the paper and to establish the robustness of our results under various different parameter values. The corresponding treatments have already been designed and programmed. We plan to run the new sessions in the near future. This project was also presented at the 2015 MSOM Conference where the selection quota was 50%. This paper also received favorable comments on the 2015 Best Paper Award of the INFORMS Behavioral Operations Section.

The main conclusion and the summary of our results is provided in Chapter 7. Chapter 8 is the Appendix which provides further analyses and the text of experimental instructions for each main section. We also note that the notation of the variables is introduced in each chapter separately. This was necessary due to the relatively large number of variables we elicited and used in our analyses. This approach makes sure that the meaning of variables has been provided recently to the reader. In addition, it helps to avoid the potentially conflicting or not self-explanatory notation of variables or coefficients.

Chapter 2

2. Methodology

This chapter discusses the history of experimental methodology, its most important objectives, advantages and drawbacks. For a broader methodological introduction on experimental economics, refer to Friedman et al. (2004), whereas Bardsley (2010) provides an overview for advanced readers. In specific for behavioral operations management, several other reviews (Bendoly et al., 2006; Loch and Wu, 2007; Bendoly et al., 2010; Katok, 2011) give further insights on the methodology and the respective findings.

The empirical and experimental approach has become a popular and widely accepted method in natural sciences already during the ancient times. Well-known anecdotes document the fact that scientists – e. g. Archimedes, Isaac Newton, Michael Faraday, Albert Einstein etc. – verified or rejected their hypotheses based on real-world observations or experiments since then. Economics used to be an artifact in the sense that the first experimental work was published only after the Second World War: Chamberlain (1948) showed that market prices do not necessarily reach the equilibrium. Building upon this work, Smith (1962) provided evidence on competitive market behavior in the lab. In the following decades, experimental work has grown at a slow but steady pace until the millennium. Then a renaissance came which still lasts. Several Nobel Laurates received their Nobel Prize – at least partially – based on their experimental investigations: Gary Becker (1992), James Heckman (2000), George Akerlof (2001), Daniel Kahneman (2002), Vernon L. Smith (2002) and Alvin E. Roth (2012). The aim of experiments is widespread and it includes testing theories, eliciting preferences, providing political advice, exploring novel settings, documenting empirical regularities and even teaching.

Operations management overtook the experimental methodology even later than economics had done so. The first experimental paper was published on the beer distribution game by Sterman (1989), which investigated inventory and order fluctuations in a serial supply chain. The initial skepticism regarding this field can be characterized by an anecdote that Kenneth L. Schultz – Associate Professor at the Air Force Institute of Technology at Dayton, Ohio, USA – told me at one of the conferences we attended together:

"Back then when I was a Ph. D. student, I received a manuscript from my advisor. It was about a laboratory experiment testing some operations problem and I had to express my opinion on the article. Honestly, I thought that the entire approach was a mere Hocus Pocus."

Interestingly and ironically, Schultz and his co-authors were the ones who made the second contribution (Schultz et al., 1998) to this field by outlining the connection between worker productivity and the amount of inventory in a serial production system. Behavioral operations has received more attention since the seminal paper by Schweitzer and Cachon (2000): This paper investigated ordering behavior for perishable products under stochastic demand – which is the workhorse operations context – it is often referred to as the newsvendor setting. Dozens of newsvendor papers were published afterwards and the field extended its coverage to various different topics. The corresponding behavioral operations sections and departments were founded in renowned professional organizations, conferences and journals (INFORMS and POMS).

In a nutshell, the above mentioned snapshot of history makes it clear that behavioral operations management uses the methodology of experimental economics. Smith (1976) formalized the main rules for it. Firstly, subjects' homegrown preferences should be neutralized while new preferences are induced by the experimenter. This assumes that subjects' money (m) composes of their initial money (m_0) and the payments they receive during the experiment (Δm). Assuming that an individual's preference (I) depends on m and all other motives (o), the following three principles have to be satisfied:

A) Monotonicity – More reward medium is preferred to less and no saturation is possible:

$$I_m = \frac{\partial I}{\partial m} \tag{2.1}$$

should exist and be positive for every (m,o).

B) Salience – The reward is dependent on subjects' actions (a):

$$\frac{\partial m}{\partial a} \neq 0 \tag{2.2}$$

C) Dominance – The influence of the other motives is negligible, the individuals' utility stems primarily from the payment they receive during the experiment:

$$\left|\frac{\partial m}{\partial a}\right| \gg \left|\frac{\partial o}{\partial a}\right| \tag{2.3}$$

Although surveys and questionnaires provided – and still provide – crucial insights to understand behavioral motives, they do not meet the above mentioned criteria – most importantly equation (2.2). Furthermore, there is evidence that individuals behave differently in actual real-life situations from what they claim they would do in hypothetical settings (Fershtman and Gneezy, 2001). Therefore, it is crucial to deliver evidence on the actual behavior which meets the criteria of external and internal validity.

Internal validity means that all the rules of the methodology are followed during the design of the experiment, the data gathering and its analysis, whereas external validity implies whether and to what extent the results can be generalized to the real world. For this purpose, we have to differentiate between laboratory and field experiments. Field experiments are conducted on site under natural circumstances where the participants are not aware of the fact that they are observed. Due to the nature of these type of investigations, the challenge is that researchers cannot control for several factors. On the contrary, laboratory experiments occur in the laboratory, but researchers can obtain all the desired control variables they want to investigate. The main advantages of this approach is this controlled environment, since the circumstances can be modified through exogenous interventions by the researchers. This could mean varying parameters ceteris paribus, changing the market design, providing different information conditions, introducing new institutions such as communication, using different framing, etc.

This thesis uses laboratory experiments to provide insights on behavior primarily related to operations management. We used unframed language, and we dealt with relatively general settings. Nonetheless, in order to provide a comprehensive picture, we need to address potential concerns for the methodology as well. A popular reservation towards the methodology targets the subject pool, namely whether students are appropriate means to an end for providing externally valid conclusions. There is evidence (Fréchette, 2016) that real-world practitioners have the same degree of biases and do not behave in a systematically different way compared to student subject pools in the overwhelming majority of settings. In particular for operations-related settings, Bolton et al. (2012) documented that there are at most negligible differences among students, operations employees and executives. Another critical issue is whether the

stake of approx. €20 is sufficient to draw far-reaching conclusions. Notice that alternatives exist by increasing the stake level (Holt and Laury, 2002) or by replicating the results with a relatively moderate amount of money in third world countries (Binswanger, 1980). Finally, some scientists object to the low number of participants and to the inexperienced subjects. A solution to these issues could be recruiting more subjects (Bostian et al., 2008), increasing the number of participants (Holt and Laury, 2005) or using experienced subjects (Kagel and Levin, 1986) for high-context decisions. As there is a feasible workaround to these issues, we are convinced that our investigation can draw valid conclusions and that it has relevant implications.

Chapter 3

3. How Sensitive Are (Risk) Preferences to Elicitation Mechanisms?

3.1. Introductory Remarks

Risk is a fundamental concept that affects human behavior and decisions in many real-life situations. Whether a person wants to invest in the stock market, tries to select the best health insurance or just wants to cross the street, he/she will face risky decisions every day. Therefore, risk attitudes are of high importance for decisions in many economics-related contexts. A multitude of studies elicit risk preferences in order to control for risk attitudes, as it is clear that they might play a relevant role: De Véricourt et al. (2013) in operations-related settings, Murnighan et al. (1988) in bargaining, Müller (2001) in dynamic environments or Tanaka et al. (2010) in linking experimental data to household income, to name just a few. Moreover, several papers try to shed light on the causes of risk-seeking and risk-averse behavior in the general population with laboratory (Harrison and Rutström, 2008), internet (Von Gaudecker et al., 2011) and field experiments (Harrison et al., 2007). Furthermore, recent evidence supports the presence of risk preferences in further operations-related settings (Becker-Peth et al., 2016; Gavirneni and Robinson, 2015).

Since the seminal paper by Holt and Laury (2002), approximately 20 methods have been published which provide alternatives to elicit risk preferences. They differ from each other in terms of the varied parameters, representation and framing. Many of these different risk elicitation methods have the same theoretical foundation and therefore claim to measure the same parameter – a subject's "true" risk preference. However, there are significant differences in results depending on the method used, as an increasing number of evidence suggests. It follows that if someone's revealed preference is dependent on the measurement method used, scientific results and real-world conclusions might be biased and misleading.

As far as existing comparison studies are concerned, they usually compare two methods with each other and often use different stakes, parameters, framing, representation, etc., which makes their results hardly comparable. This chapter complements existing experimental literature by making the following contribution: Taking the method by Holt and Laury (2002) as a basis, we conduct a comprehensive comparison of the multiple price list (MPL) versions

of risk elicitation methods by classifying all methods into nine categories. To the best of our knowledge, no investigation – including various measures of between- and within-method consistency – has ever been conducted in the literature that incorporates such a high number of methods. To isolate the effect of different methods, we consistently use the MPL representation and calibrate the risk intervals to be the same for each method assuming expected utility theory (EUT) and constant relative risk aversion (CRRA), while also keeping the risk-neutral expected payoff of each method constant and employing a within-subject design. Moreover, our design allows us to investigate whether differences across methods can be reconciled by assuming different functional forms documented in the literature such as constant absolute risk aversion (CARA), decreasing relative risk aversion (DRRA), decreasing absolute risk aversion (DARA), increasing relative risk aversion (IRRA) and increasing absolute risk aversion (IARA). Additionally, we extend our analysis to incorporate EUT with probability weighting and also to prospect theory (PT) and cumulative prospect theory (CPT).

We investigate the within-method consistency of each risk elicitation procedure by comparing the differences in subjects' initial and repeated decisions within the same MPL method. Moreover, we assess the methods' self-perceived complexity and we also shed light on the differences and similarities between methods. In the end, we provide suggestions which specific MPL representation to use by comparing our results to decisions in two benchmark games that resemble real-life settings: investments in capital markets and auctions. Therefore, we analyze the methods along two dimensions: robustness and predictive power, and determine which property of particular methods drives risk attitude and its consistency.

We find that a particular modification of the method by Holt and Laury (2002) derived by Drichoutis and Lusk (2012) has the highest predictive power in investment settings both according to the OLS regression and Spearman rank correlation. In addition, specific methods devised by Bruner (2009) perform also relatively well in these analyses. However, the method by Drichoutis and Lusk (2012) clearly outperforms the other methods in terms of within-method consistency. It is perceived as relatively simple – in the end, Chapter 3 provides the recommendation for researchers to implement this method when measuring risk attitudes using the MPL framework. Moreover, our results remain qualitatively the same if we relax our assumption on the risk aversion function, or if we take probability weighting or alternative theories such as prospect theory or cumulative prospect theory into account.

3.2. Multiple Price Lists Explained

Incentivized risk preference elicitation methods aim to quantify subjects' risk perceptions based on their revealed preferences. We present nine methods in a unified structure – the commonly used MPL format – to our subjects, taking one of the most cited methods as a basis: Holt and Laury (2002). The MPL table structure is as follows: Each table has multiple rows, and in each row, all subjects face a lottery (two columns) on one side of the table, and a lottery or a certain payoff (one or two columns) on the other side, depending on the particular method. Then, from row to row, one or more of the parameters change. The methods differ from each other by the parameter which is changing. As the options on the right side become strictly more attractive from row to row, the subject indicates the row where he/she wants to switch from the left option to the right option. This switching point then gives us an interval for a subject's risk preference parameter according to Table 3.1, assuming EUT and CRRA:

$$U(C) = \frac{C^{1-\rho}}{1-\rho} \tag{3.1}$$

Where U denotes the individual's utility of endowment C and ρ is the elicited risk aversion coefficient. In order to ease comparison to the existing studies, we used exactly the same coefficient intervals as Holt and Laury (2002).

Table 3.1: Risk Parameter Intervals a la Holt and Laury (2002)

Interpretation by Holt and Laury (2002)	Switching point	Risk parameter interval
Highly risk loving	1	$\rho \leq$ -0.95
Very risk loving	2	-0.95 $< \rho \leq$ -0.49
Risk loving	3	-0.49 $< \rho \leq$ -0.15
Risk neutral	4	-0.15 $< \rho \leq$ 0.15
Slightly risk averse	5	0.15 $< \rho \leq$ 0.41
Risk averse	6	0.41 $< \rho \leq$ 0.68
Very risk averse	7	0.68 $< \rho \leq$ 0.97
Highly risk averse	8	0.97 $< \rho \leq$ 1.37
Stay in bed	Never	1.37 $< \rho$

Notes: This table indicates the mapping from a subject's chosen switching point into the resulting risk parameter intervals in each method; the leftmost column contains the interpretation of the risk intervals; "Never" means that a subject prefers the option "Left" in each row.

Note that several other representations of risk elicitation methods exist besides the MPL such as the bisection method (Andersen et al., 2006), the trade-off method (Wakker and Deneffe, 1996), questionnaire-based methods (Weber et al., 2002), willingness-to-pay (Hey et al., 2009), etc., but the MPL is favored because of its common usage. Andersen et al. (2006)

consider that the main advantages of the MPL format are that it is transparent to subjects and that it provides simple incentives for truthful preference revelation. They additionally list its simplicity and the little time it takes as further benefits. As far as the specific risk elicitation method in the MPL framework designed by Holt and Laury (2002) is concerned, it has proven itself numerous times in providing explanations for several phenomena such as behavior in 2 by 2 games (Goeree et al., 2003), market settings (Fellner and Maciejovsky, 2007), smoking, heavy drinking, being overweight or obese (Anderson and Mellor, 2008), consumption practices (Lusk and Coble, 2005) and many others.

Several studies document violations of EUT under risky decision making (Huck and Müller, 2012) and provide suggestions how these differences can be reconciled (Bleichrodt et al., 2001). In addition, recent studies (Tanaka et al., 2010; Bocqueho et al., 2014) document potential empirical support for prospect theory – PT, Kahneman and Tversky (1979) – when it comes to risk attitudes: Harrison et al. (2010) found that PT describes behavior of half of their sample best where:

$$U(C) = \begin{cases} C^{\alpha} & if\ C \geq 0 \\ -\lambda \cdot (-C)^{\delta} & if\ C < 0 \end{cases} \tag{3.2}$$

We denote the loss aversion parameter with λ; α and δ determine the curvature of the loss aversion function in the gains and the losses domain. There is also evidence that subjective probability weighting – PW (Quiggin, 1982) – should be taken into account:

$$w(p) = \frac{p^{\gamma}}{[p^{\gamma}+(1-p)^{\gamma}]^{\frac{1}{\gamma}}} \tag{3.3}$$

where *w(p)* is the probability weight and the technical γ parameter enhances that the preferences are as they are supposed to be. In addition, cumulative prospect theory (CPT) – PT combined with PW (Tversky and Kahneman, 1992) – might also be a candidate that can explain the documented anomalies under EUT. Wakker (2010) provides an extensive review on risk under PT.

In addition to its computational ease, Wakker (2008) claims that CRRA is the most commonly postulated assumption among economists in experimental work. Most recently, Chiappori and Paiella (2011) provide evidence on the validity of this assumption in economic-

financial decisions. Nevertheless, alternative functional forms have been introduced, such as CARA (Pratt, 1964):

$$U(C) = \frac{-e^{\rho C}}{\rho} \tag{3.4}$$

It was also questioned whether social status – and mostly the role of wealth or income – might shape risk attitude, which would lead to functions which are increasing or decreasing in these factors such as IRRA and DRRA (Andersen et al., 2012):

$$U(C,W) = \frac{(\omega W^r + C^r)^{\frac{1-\rho}{r}}}{1-\rho} \tag{3.5}$$

where W is the wealth, ω is the distributive share parameter and r is the elasticity of substitution between wealth and the payment received in our experiment. Similarly, we introduce the functions of IARA and DARA (Saha, 1993):

$$U(C,W) = \frac{-e^{-\rho r(C+W\omega)}}{\rho} \tag{3.6}$$

A review on these functions is provided by Levy (1994). In our robustness analysis, we relax our original assumptions on EUT and CRRA and incorporate all of the above mentioned alternative theories and functional forms. Note that even though we calibrated our parameters to accomodate EUT and CRRA, one is still able to calculate the risk parameter ρ using the aforementioned alternative specifications. This implies that the same switching point in two methods does not yield the same risk parameter estimate under different specifications but these are still directly comparable according to theory, as they claim to measure a subject's underlying risk attitude ceteris paribus.

We group our aforementioned nine risk elicitation methods into two categories:

A) The standard gamble methods (SG methods), where on one side of the table there is always a 100% chance of getting a particular certain payoff and on the other side there is a lottery.
B) The paired gamble methods (PG methods), with lotteries on both sides.

We therefore primarily conduct a comparison of different MPL risk elicitation methods. What we do not claim, however, is that the method devised by Holt and Laury (2002) – or MPL

in general – is the most fitting to measure people's risk preferences. We strive for giving a recommendation to researchers who already intend to use the method by Holt and Laury (2002) in their studies, and provide a better alternative that shares its attributes with the original MPL design. It should be mentioned that there is an alternative interpretation of our investigation: The different MPL methods can also be conceived as a mapping of existing risk elicitation methods (from other frameworks) to the MPL space. Several methods exist where the risk elicitation task is provided in a framed context – such as pumping a balloon until it blows (Lejuez et al., 2002) or avoiding a bomb explosion (Crosetto and Filippin, 2013). Similarly, some methods differ due to the representation of probabilities with percentages (Holt and Laury, 2002) or charts (Andreoni and Harbaugh, 2010). All these methods can be displayed with different MPLs by showing the probabilities and the corresponding payoffs in a table format. We provide a complete classification of these methods in section 3.3.

Up to now, different risk elicitation methods were compared by keeping the original designs, but this approach comes at a price: As the methods differ in many dimensions, any documented difference can be attributed to those particular characteristics. Our approach can be understood as a way to make all risk elicitation methods as similar as possible, with the drawback of losing the direct connection to the original representation. Chapter 3 should therefore primarily be seen as a comparison of different MPL risk elicitation methods. The resulting comparison of existing risk elicitation methods – by mapping them into the same space – is only reported for the sake of completeness.

3.3. Literature Review and Hypotheses

We will now discuss the different methods in greater detail and how they are embedded in the literature. Table 3.2 provides a summary of the exact parameter that is changing across methods. For a complete list of all methods with the corresponding parameter values (as presented to subjects), refer to the Appendix 8.1.4.

Table 3.2: Method Overview

	What is Changing?			
Method	Probability	Highest Payoff	Lowest Payoff	Certainty Equivalent
SGp	**Yes**	No	No	No
SGhigh	No	**Yes**	No	No
SGlow	No	No	**Yes**	No
SGsure	No	No	No	**Yes**
SGall	No	**Yes**	**Yes**	**Yes**
PGp	**Yes**	No	No	Not Available
PGhigh	No	**Yes**	No	Not Available
PGlow	No	No	**Yes**	Not Available
PGall	**Yes**	**Yes**	**Yes**	Not Available

Notes: This table indicates which parameters change from row to row in each method.

3.3.1. Standard Gamble Methods

Among the SG methods, there are four parameters that can be changed: The sure payoff (*sure*), the high payoff of the lottery (*high*), the low payoff of the lottery (*low*) or the probability of getting the high payoff (p) – eventually the probability of getting the low payoff ($1 - p$) respectively. The parameters must of course be chosen in such a way that $p_{high} > p_{sure} > p_{low}$ always holds. For instance, we denote the SG method where the low payoff is changing by "SGlow", the SG method with the varying certainty equivalent by "SGsure" or the standard gamble method where the probabilities are changing as "SGp".

Binswanger (1980) introduced a method (SGall) where only one of the options has a certainty equivalent. The other options consist of lotteries where the probabilities are fixed at 50-50, but both the high and the low payoff are changing. Cohen et al. (1987) used risk elicitation tasks in which probabilities and lottery outcomes were held constant and only the certainty equivalent was varied. These methods have later been redesigned and presented in a more sophisticated format as a single choice task by Eckel and Grossman (2002)

A recent investigation by Abdellaoui et al. (2011) presents a similar method (SGsure method) in an MPL format with 50-50 probabilities. Bruner (2009) presents a particular certainty equivalent method, where the certainty equivalent and the lottery outcomes are held constant, but the corresponding probabilities of the lotteries are changing (SGp method). Additionally, Bruner (2009) introduces another method where only the potential high outcomes of lotteries vary (SGhigh method). Although not present in the literature, we chose to include a

method where the potential low outcome varies for reasons of completeness (SGlow method). Fort the exact examples, see Tables 8.1.1 – 8.1.5 in the Appendix 8.1.4, which correspond to the SG methods.

3.3.2. Paired Gamble Methods

Holt and Laury (2002) introduced one of the most-cited elicitation methods under EUT up to now, where potential outcomes are held constant and the respective probabilities change (PGp). Drichoutis and Lusk (2012) suggest a similar framework where the outcomes of different lotteries change while the probabilities are held constant. We separate these methods further into PGhigh and PGlow depending on whether the high or the low outcome is varied in the MPL. Additionally, the PGall method varies both the probabilities and the potential earnings at the same time.

Many framed risk elicitation tasks fit into the framework of choosing between different lotteries. Sabater-Grande and Georgantzis (2002) provide ten discrete options with different probabilities and outcomes to choose from. Lejuez et al. (2002) introduce the Balloon Analogue Risk Task where subjects could pump up a balloon, and their earnings depend on the final size of the balloon. The larger the balloon gets, the more likely it will explode, in which case the subject earns nothing. Visschers et al. (2009) and Andreoni and Harbaugh (2010) use a pie chart for probabilities and a slider for outcomes to visualize a similar trade-off effect in their experiment. Crosetto and Filippin (2013) present their Bomb Risk Elicitation Task with an interesting framing: This quantifies the aforementioned tradeoff between probability and potential earnings with the help of a bomb explosion. For concise examples, see Tables 8.1.6 – 8.1.9 in the Appendix 8.1.4, which correspond to the PG method.

3.3.3. Questionnaire Methods

In addition to the MPL methods, we chose to incorporate questionnaire risk elicitation methods into our investigation. Several methods have been introduced that evaluate risk preferences with non-incentivized survey-based methods. The questions and methodology they use are very similar. The most recently published ones include the question from the German Socio-Economic Panel Study by Dohmen et al. (2011) or the Domain-Specific Risk-Taking Scale (DOSPERT) by Blais and Weber (2006). For a more detailed description, see the last paragraph of Section 3.4.

3.3.4. Comparison Studies

The question arises of which method to use if there is such a large variety of risk elicitation methods and whether they lead to the same results. Comparison studies exist, but the majority compares only two methods with each other, thus their scope is limited. The question of within-method consistency has been addressed by a few papers: Harrison et al. (2005) document high re-test stability of the method introduced by Holt and Laury (2002) – the PGp. Andersen et al. (2008a) test consistency of the PGp method within a 17-month time frame. They find some variation in risk attitudes over time, but do not detect a general tendency for risk attitudes to increase or decrease. This result was confirmed in Andersen et al. (2008b). Yet there is a gap in the academic literature on the time stability of different methods and their representation we are eager to fill.

> **Hypothesis H3.1:** On the contrary to predictions by the normative theory, the different risk elicitation methods deliver highly heterogeneous risk attitude parameters.

If this proves to be true, then an objective benchmark – which resembles real-life settings – should serve as a deal-breaker: it shall help to assess the particular methods' predictive power in externally valid settings.

> **Hypothesis H3.2:** The different risk elicitation methods have different predictive power in auction and investment settings.

Interestingly, more work has been done on the field of between-method consistency. Fausti and Gillespie (2000) compare risk preference elicitation methods with hypothetical questions using results from a mail survey. Isaac and James (2000) conclude that the risk attitude and the relative ranking of subjects is different in the Becker-DeGroot-Marschak procedure and in the first-price sealed-bid auction setting. Berg et al. (2005) confirm that the assessment of risk preferences varies generally across institutions in auction settings. In another comparison study, Bruner (2009) shows that changing the probabilities versus varying the payoffs leads to different levels of risk aversion in the PG tasks. Moreover, Dave et al. (2010) conclude that subjects show different degrees of risk aversion in the Holt and Laury (2002) – PGp – and in the Eckel and Grossman (2008) – SGall – task. Their results were confirmed by Reynaud and Couture (2012) who used farmers as the subject pool in a field experiment. Bleichrodt (2002) argues that a potential reason for these differences might be attributed to the

fact that the original method by Eckel and Grossman (2008) does not cover the risk seeking domain, which can be included with the slight modification we made upon the implementation of this method. Dulleck et al. (2015) test the method devised by Andreoni and Harbaugh (2010) using a graphical representation against the PGp and describe both a surprisingly high level of within- and inter-method inconsistency. Drichoutis and Lusk (2012) compare the PGp method to a modified version of it where probabilities are held constant. Their analysis reveals that the elicited risk preferences differ from each other both at the individual and at the aggregate level. Most recently, Crosetto and Filippin (2016) compare four different risk preference elicitation methods using their original representation and framing and confirm the relatively high instability across methods.

In parallel, a debate among survey-based and incentivized preference elicitation methods emerged which were present since the survey on questionnaire-based risk elicitation methods by Farquhar (1984). Eckel and Grossman (2002) conclude that non-incentivized survey-based methods provide misleading conclusions for incentivized real-world settings. In line with this finding, Anderson and Mellor (2009) claim that non-salient survey-based elicitation methods and the PGp method yield different results. On the contrary, Lönnqvist et al. (2015) provide evidence that the survey-based measure, which Dohmen et al. (2011) had implemented, explains decisions in the trust game better than the SGsure task. Charness and Viceisza (2016) provide evidence from developing countries that hypothetical willingness-to-risk questions and the PGp task deliver deviating results.

Hypothesis H3.3: The methods are different in terms of within-method consistency.

3.3.5. Further Considerations

A recent stream of literature broadens the horizon of investigation to further aspects of different elicitation methods: Weber et al. (2002) show that people have different risk attitudes in various fields of life, thus risk preferences seem to be domain-specific. Lönnqvist et al. (2015) document no significant connection between the HLp task and personality traits. Dohmen et al. (2010) record a connection between risk preferences and cognitive ability, which is questioned by Andersson et al. (2016). Hey et al. (2009) investigate noise and bias under four different elicitation procedures and emphasize that elicitation methods should be regarded as strongly context-specific measures. Harrison and Rutström (2008) provide an overview and a broader

summary of elicitation methods under laboratory conditions, whereas Charness et al. (2013) survey several risk preference elicitation methods based on their advantages and disadvantages.

In addition, there is evidence that framing and representation matters. Wilkinson and Wills (2005) advise against using pie charts which show probabilities and payoffs since human beings are not good at estimating angles. Hershey et al. (1982) identify important sources of bias to be taken into account and pitfalls to be avoided when designing elicitation tasks. Most importantly, these include task framing, differences between the gain and loss domains and the variation of outcome and probability levels. Eichberger et al. (2003) hint that risk attitudes depend on whether a willingness-to-pay or a price mechanism is used to elicit them. Von Gaudecker et al. (2008) extend this result and show that the same risk elicitation methods for the same subjects deliver different results when using different frameworks – e.g. multiple price list, trade-off method, ordered lotteries, graphical chart representation, etc. This procedural indifference was confirmed by Attema and Brouwer (2013) as well, which implies that risk preferences on an individual level are susceptible to the representation and framing used.

The previous paragraphs lead us to the conclusion that methods should be compared to each other by using the same representation and format. This justifies our decision to compare them using the standard MPL framework which guarantees that the differences cannot be attributed to the different framing and representation of elicitation tasks. However, this comes at the price that we had to change some of the methods slightly, which implies that they are not exactly the same as their originally published versions. We certainly do not claim that the MPL is the only valid framework, but our choice for it seems justified by its common usage and relative simplicity. We consider a future investigation using a different representation technique as a potentially interesting addition. Also, we emphasize that the differences in our results exist among the MPL representations of the methods and they can only be generalized to the original methods to a limited extent; see Table 3.3 for an overview of the link between the MPL representation and the particular method that was published originally. Consequently, the predictive power and the within-method consistency will allow us to provide suggestions on the quality of methods objectively. Furthermore, by taking the particular characteristics of each method, we can determine whether the presence of a certainty equivalent, the ascending or descending nature of numbers in the table or whether the changing of probabilities or payoffs yield more stable results or an elevated level of risk aversion.

Hypothesis H3.4: A clear pattern will emerge which method characteristics foster risk taking and which deliver less consistent results.

Table 3.3: Link between MPL Representation and the Literature

Method	Corresponding literature
SGp	Bruner (2009)
SGhigh	Bruner (2009)
SGlow	-
SGsure	Cohen et al. (1987), Abdellaoui et al. (2011)
SGall	Binswanger (1980), Eckel and Grossman (2002, 2008)
PGp	Holt and Laury (2002, 2005)
PGhigh	Drichoutis and Lusk (2012)
PGlow	Drichoutis and Lusk (2012)
PGall	Lejuez et al. (2002), Sabater-Grande and Georgantzis (2002), Andreoni and Harbaugh (2010), Crosetto and Filippin (2013)
Questionnaire	Weber et al. (2002), Dohmen et al. (2011)

Notes: On the left, this table lists all MPL and questionnaire methods, on the right the corresponding literature is provided.

3.4. Experimental Design

We provide a laboratory experiment to compare different MPL risk elicitation methods. Subjects answered the risk elicitation questions first. Then, benchmark games were presented to them to gauge predictive power, which was followed by a non-incentivized questionnaire. We provide a detailed description on the exact procedures of each part in the later paragraphs.

We conducted ten sessions at the Vienna Center for Experimental Economics (VCEE) with 96 subjects. One subject has been excluded from our subject pool considering that he/she was unable to answer the control questions correctly upon multiple trials. Our 12 sessions lasted about 2 hours. The range of earnings was between 3€ and 50€, which implies an average payment of 20.78€ with a standard deviation of 10.1€. We calibrated these payments similarly to previous studies – see Bruner (2009) or Abdellaoui et al. (2011), among others. Average earnings were about €9.5 in the risk task and about 8.3€ in the benchmark games plus a €3.00 show-up fee. Harrison et al. (2009) provide evidence that the existence of a show-up fee could lead to an elevated level of risk aversion in the subject pool. In our experiment, this moderate show-up fee was only pointed out to the subjects after they had made their decisions in the risk elicitation methods and the benchmark games. Thus, it could not have distorted their

preferences. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007), and ORSEE (Greiner, 2015) was used for recruiting subjects.

We employed a within-subject design, meaning that each subject took decisions in each and every task as in other comparison studies – Eckel and Grossman (2008) or Crosetto and Filippin (2016). This design property rules out that the methods differ due to heterogeneity between subjects, but it comes with the drawback that methods which were encountered later might deliver noisier or different results due to fatigue or other factors, as the answer to a particular method could also be a function of the previously seen MPLs. Consequently, we included the order in which a method appeared in all regressions as controls wherever possible, compared the variance in earlier and latter methods and tested for order effects; no significant effects were found (Table 8.1.12 in Appendix 8.1.5). To avoid biases, a random number generator determined the order of methods for each subject separately in the beginning of each session. Each subject encountered the methods in a unique random order and each order was used only once during the entire experiment.

After having received instructions on screen and in written form, subjects went through the nine incentivized risk elicitation methods. In order to avoid potential incentive effects mentioned by Holt and Laury (2002), the expected earnings for a risk-neutral individual were equal in every method. Furthermore, to avoid potential biases due to the different reactions to gains and losses (Hershey et al., 1982), each of our lotteries is set in the gains domain. Andersen et al. (2006) confirmed previous evidence (Poulton, 1989) that there is a slight tendency of anchoring and choosing a switching point around the middle for risk elicitation tasks. In order to counteract anchoring and one-directional distortion of preferences as a consequence of this unavoidable pull-to-center effect, each risk elicitation task appeared randomly either top-down or bottom-up. Depending on randomization, out of nine potential switching opportunities, the fourth or the sixth option were the risk-neutral switching points. An example for the difference between the top-down and bottom-up representation can be found in Table 8.1.10 in the Appendix 8.1.4.

Subjects also had the opportunity to look at their given answer and they could have modified it right after each decision if they wished to do so. After having made a decision in each method, we asked subjects the following question: “On a scale from 1 to 10, how difficult was it for you to make a decision in the previous setting?”. With this question we assessed self-perceived complexity of the tasks, since there is evidence in the literature (Mador et al., 2000)

that subjects make noisier decisions if the complexity of a lottery increases, and therefore a less complex method is preferred. Moreover, Dave et al. (2010) outline the trade-offs between noise, accuracy and subjects' mathematics skills. They suggest that it is a good strategy to make MPL tasks simpler for subjects. In this spirit, we asked our subjects to indicate the row in which they switched from the "LEFT" column to the "RIGHT" column, thereby enforcing a single switching point (SSP). Using this framework, subjects were not required to make a decision for each and every row in every method, which would have meant more than 100 monotonous, repetitive binary choices throughout the experiment. Additionally, this approach ensures that the subjects were guaranteed to give answers without preference reversals. We consider this option more viable than accepting multiple switching points – thus allowing inconsistent choices – and using the total number of "safe" choices to determine a subject's risk coefficient interval. The SSP has been used several times, see Gonzalez and Wu (1999) or Jacobson and Petrie (2009) for examples.

By enforcing a SSP, we faced a trade-off between potential boredom and the non-detection of people with inconsistent preferences. Furthermore, some of the reported within-method instability might stem from "fat preferences" or indifference between two or more options. However, the SSP can further be justified because only a small proportion of subjects expressed multiple switching points in earlier studies – e.g. 8.5% in Dave et al. (2010) and 6.6% in Holt and Laury (2002), so this design choice is highly unlikely to drive our results.

In order to test within-method consistency, three of the nine methods were randomly chosen and presented to the subjects again, without telling them that they had already encountered that particular method. The exact number of occasions a particular method was selected a second time is roughly equal across methods: PGhigh(33), PGlow(25), PGp(26), PGall(39), SGhigh(30), SGsure(33), SGlow(27), SGp(40), SGall(35). Repeating all methods was not feasible due to fatigue concerns, as the experiment is already quite long. This approach allows us to test both within-method and inter-method consistency. Our subjects were allowed to change their answers here once as well. The perceived complexity of the risk elicitation procedures was also elicited again. Control questions were used for the preference elicitation methods and for each benchmark game in order to verify that subjects understood the task they were about to perform. Subjects had to answer them correctly in order to participate in the experiment. See the Appendix 8.1.2 for the exact text of these questions.

We incorporated the random lottery incentive system by Cubitt et al. (1998). Thus, the computer chose one of the twelve risk preference methods and one of the eight rows within that particular method on a random basis to be payoff-relevant. Additionally, one of the three benchmark games was chosen to be payoff-relevant as well. The random lottery incentive system helped us to keep the costs at a reasonable level while we could have had similarly sized (Bruner, 2009) or even larger stakes for the elicitation tasks compared to previous studies (Holt and Laury, 2002). In addition, it also mitigated potential income effects. Nevertheless, we note that the random lottery incentive system might be a potential caveat for Chapter 3, since Cox et al. (2015) document somewhat different behavior under various payment mechanisms. However, it is more likely that the differences they document are attributed to the general noisiness of the risk elicitation methods.

As far as hedging behavior is concerned, Blanco et al. (2010) provide evidence that hedging and the corresponding biased beliefs and actions can only be problematic if the hedging opportunities are highly transparent. Taking this consideration into account, we provided feedback on the outcome of the risk elicitation tasks only at the end of the experiment. Thus, it was not possible for subjects to create a portfolio and use hedging behavior over different parts of the experiment.

On top of the risk elicitation tasks, we used three benchmark games that resemble real-life settings as well as situations relevant to economists. As behavior in these settings should only depend on risk attitudes, they will serve as benchmarks to contribute to the debate which risk elicitation methods are appropriate to predict behavior in these games. The benchmark games appeared in a randomized order. First, we used the same investment task as Charness and Gneezy (2010). Here, subjects could decide how much they wanted to invest in stocks and bonds out of an endowment of 10€. Subjects knew that any investment in bonds is a safe investment, and therefore they received the same amount they had invested in bonds as income. Additionally, the amount they invested in stocks was to be multiplied by 2.5 or lost completely with equal chance. Under EUT, this setting implies that both risk neutral and risk seeking decision makers should invest the entire amount. Thus, in order to be able to differentiate between them, we introduced another investment setting where the potential payment for stocks was 1.5 times the invested amount.

The third benchmark game was a first-price sealed-bid auction against a computerized opponent in line with Walker et al. (1987). Subjects could bid between 0.00€ and 20.00€ of

their endowment, and they knew that the computer bid any amount between 0.00€ and 20.00€ with equal chance. The potential earnings (E_1 for subject 1) according to the bids (x_1;x_2) are:

$$\begin{cases} 20 - x_1 & if\ x_1 > x_2 \\ 0 & if\ x_1 < x_2 \\ 20 - x_i\ or\ 0\ (with\ 50\%\ chance) & if\ x_1 = x_2 \end{cases} \tag{3.7}$$

Our benchmark games are deliberately chosen in such a way that risk is clearly relevant in the games, while being one step away from the artificial risk elicitation mechanisms. Therefore, all benchmark games are framed heavily, while we ensure still that risk attitudes should be the only factor driving the subjects' decisions. The investment settings are very similar to the risk elicitation mechanisms described above in the sense that they resemble an SG method (with the difference that you choose your sure payoff and your lottery at the same time). The auction is more complex, as the optimal risk-neutral solution is harder to compute, but here you basically choose your own lottery, too. We therefore expect stronger correlation with the MPL methods for the investment games.

The experiment concluded with an extensive questionnaire. In order to incorporate survey-based measures, we asked subjects to provide an answer on a ten-point Likert-scale to the following two questions in line with Dohmen et al. (2011): "In general, are you a person who is fully prepared to take risk or do you try to avoid taking risks?" and "In financial situations, are you a person who is fully prepared to take risks or do you try to avoid taking risks?". The perceived complexity of these questions was elicited as well. In the questionnaire, we elicited the following socioeconomic factors: age, gender, field of study, years of university education, nationality, high school grades in mathematics, monthly income and monthly expenditure. Furthermore, we elicited cognitive ability by conducting a cognitive reflection test (Frederick, 2005). Lastly, we assessed subjects' personalities in line with Rammstedt and John (2007) who provide a short measure of personality traits according to the BIG5 methodology introduced by Costa and McCrae (1992). In the BIG5, personality is measured along five dimensions: Agreeableness, Conscientiousness, Extraversion, Neuroticism and Openness.

3.5. Results

We will establish in sections 3.5.1 and 3.5.2 that the elicited risk parameter is highly dependent on the particular variant of the MPL because the overall distributions of switching points are very diverse and the rank correlations between the different methods are low in most cases.

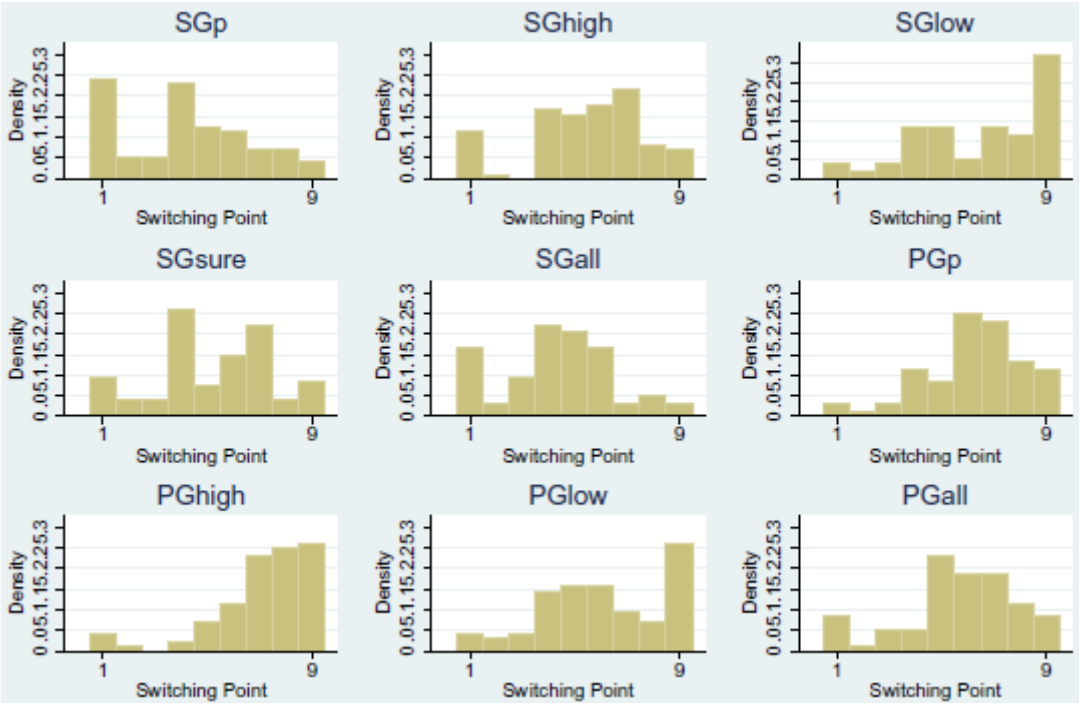
Section 3.5.3 analyzes the common features across methods. In section 3.5.4 we apply multiple measures to determine method quality. To this end, we first use benchmark games to let the data speak which risk elicitation methods predict behavior in these games best. In Section 3.5.4, we will show which method produces the most stable results overall. Section 3.5.5 concludes with the result that the PGhigh method is the most stable method and that it has the highest predictive power.

3.5.1. The Overall Distributions are Different

According to EUT, a subject's behavior does not depend on which parameters are changed from row to row, as his underlying risk parameter value is constant. As the different versions of the MPL are calculated in such a way that the same switching point implies the same risk parameter interval, a consistent individual should have the same switching point in all versions of the MPL. This implies that the distributions of switching points should be the same across methods, barring some noise.

First, see Figure 3.1 for a graphical representation of the distributions. It is clearly visible at first glance that the distributions are not the same across all methods, which confirms H3.1.

Figure 3.1: Distributions of Risk Preferences



Notes: A low value indicates risk loving and a high value indicates risk averse behavior; x-axis: Switching points (e.g. risk preferences) of subjects, where 1 means that a subject switches from left to right in the first row and 9 means that a subject never switches; y-axis: The frequency of switching points.

For example, in the SGp method, most subjects would be classified as highly risk loving, whereas in the PGhigh method the majority of subjects would be described as risk averse. To verify whether distributions across methods are the same, we conduct two tests: A Friedman test, which shows that the means are not the same across methods ($p<0.0001$), and a Kruskal-Wallis test, which shows that the distribution of answers is not the same across methods ($p<0.0001$). We conclude that the switching points are, contrary to standard theory but in accordance with the literature, dependent upon the version of the particular MPL variation used.

To see which specific versions are significantly different from each other, we conduct a series of Mann-Whitney-Wilcoxon tests (MWW tests), the natural pairwise analogue to the Kruskal-Wallis test. We use the MWW test to give a comparison of the distributions, as a difference in distributions is a more meaningful statistic here than a comparison of means. The p-values of the pairwise tests can be found in Table 3.4.

Table 3.4: Pairwise Wilcoxon Tests for Equality of Distributions

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall	GQ
SGhigh	0.00***									
SGlow	0.00***	0.00***								
SGsure	0.00***	0.37	0.00***							
SGall	0.79	0.00***	0.00***	0.01**						
PGp	0.00***	0.01**	0.28	0.00***	0.00***					
PGhigh	0.00***	0.00***	0.02*	0.00***	0.00***	0.00***				
PGlow	0.00***	0.02*	0.23	0.01**	0.00***	0.68	0.00***			
PGall	0.00***	0.31	0.02*	0.04*	0.00***	0.08	0.00***	0.39		
GQ	0.02*	0.03*	0.00***	0.29	0.04*	0.00***	0.00***	0.00***	0.01**	
FQ	0.00***	0.64	0.01**	0.29	0.00***	0.02*	0.00***	0.04*	0.36	0.00***

Notes: p-values of pairwise Wilcoxon tests are displayed; GQ: General question; FQ: Financial question. Stars are given as follows: *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$.

Out of 55 pairwise comparisons, 28 comparisons indicate that methods are different at $p<0.001$. 34 (43) instances suggest that methods are different at $p<0.01$ (0.05) significance levels. Note that one should be careful while reading this table and the ones following because of the presence of the multiple testing problem; therefore, we introduce a new notation in the tables: p-values lower than 0.001 are denoted by three stars, p-values lower than 0.01 are denoted by two stars and p-values lower than 0.05 are denoted by one star. A value of $p<0.001$ can be interpreted as significant, even when using the conservative Bonferroni correction – see Abdi (2007). To make sure that the differing results are not a product of fatigue or order effects, we also test whether the CRRA-coefficients of methods that were encountered later in the

experiment exhibit biases or more noise; the resulting tests shows no significant order effects overall and across methods (Table 8.1.12 in Appendix 8.1.5).

We conclude that the methods deliver significantly different results, and that the different versions of the MPL cannot be used interchangeably, as the estimated risk preference parameter depends heavily on the version used. Subjects can easily be classified as risk loving in one version and as risk averse in another. Of course we do not know a subject's true risk preference, and therefore any of the methods might be able to classify a subject correctly. To provide an answer to this puzzle, see Section 3.5.4, where we conduct a quality assessment of the different methods.

3.5.2. The Rank Correlations are Low

In this section we look at the rank correlation coefficients between the different methods and the questionnaire answers. If there are high rank correlations between the risk elicitation methods, one might argue that it is irrelevant which one is used if one intends to control for risk attitudes under any given circumstance. Rank correlations between the MPL methods and the questionnaire measures can be found in Table 3.5.

Table 3.5: Spearman Rank Correlation Coefficients

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall	GQ
SGhigh	0.46									
SGlow	0.33***	0.44***								
SGsure	0.05	0.22*	0.26*							
SGall	0.03	0.18	-0.03	0.19						
PGp	0.17	0.15	0.17	0.21*	-0.04					
PGhigh	0.20	0.39***	0.21*	0.03	0.21*	0.25*				
PGlow	0.31**	0.28**	0.25*	0.19	-0.02	0.13	0.21*			
PGall	0.24	0.21*	-0.01	0.08	0.19	0.04	-0.01	0.08		
GQ	0.15	0.13*	0.06	-0.12	0.11	0.02	0.14	0.04	0.06	
FQ	0.26*	0.23*	0.29*	0.18	0.10	-0.04	0.04	0.24*	0.13	0.46***

Notes: This table includes the nine different methods and the questionnaires (GQ: General questionnaire, FQ: Financial questionnaire). Stars are given as follows: *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$.

We see that some of the correlations are significant, but only 11% of all pairwise comparisons in total if we test conservatively at $p<0.001$ because of the multiple testing problem. Pay special attention to the fact that PGp, the most widely used method today, has no significant rank correlations with any of the other methods. Also, the financial questionnaire (FQ) results have

much higher correlations with the other methods than the general questionnaire (GQ) results, strengthening the argument that risk attitudes are domain specific. See also Table 8.1.11 in the Appendix for standard correlations, which basically gives the same results as Table 3.5. These findings provide further evidence that the elicitation procedure should be chosen with care as the elicited risk aversion coefficient and also the relative ranking of subjects varies within broad boundaries.

3.5.3. Method Similarities

We have established in sections 3.5.1 and 3.5.2 that there are significant differences in the distributions of the risk elicitation methods. There are, however, some similarities that can be observed across methods: In Table 3.6, we classify MPLs according to whether the high payoff, the low payoff, the probabilities or the certainty equivalents change in the MPL table, whether the method has a certainty equivalent and whether the table was presented in a top-down or bottom-up format. Furthermore, we control for age, gender, cognitive reflection scores and the order in which the tables were presented.

Table 3.6: Similarities across All Methods

	Not repeated	Repeated
High payoff changes	-0.11**	0.05
Low payoff changes	0.21***	0.21***
Probability changes	-0.31***	-0.34***
Certainty equivalent changes	0.09*	-0.04
Has a certainty equivalent	-0.50***	-0.50***
Top-down representation	0.09**	0.04
Constant	0.84	0.21
Adjusted R^2	0.12	0.19
Number of observations	864	288

Notes: This table shows the OLS regressions clustered by individual subjects with one observation being the outcome from one answer by one subject in one method. The dependent variable is the resulting CRRA-coefficient, with low scores indicating risk-loving behavior; independent variables on the left are dummies. Non-significant controls for age, gender, order, BIG5 scores, income and CRT scores are included in the regressions but omitted in the table. The first column gives results for the first time subjects encountered one of the nine methods, the second column for the repeated measurements. Stars are given as follows (differently than in the other tables, due to the absence of a multiple testing problem): *: $p<0.10$; **: $p<0.05$; ***: $p<0.01$.

Column 1 shows the results for the first time a method was encountered, and column 2 for the repeated measurements. To provide an answer to H3.4, we see that generally, methods that change the probabilities or methods that have a certainty equivalent classify subjects as more risk loving, while methods that change the low payoffs classify subjects as more risk-

averse. In two MPL methods, PGall and SGall, multiple characteristics of the MPL table were changed at the same time. Consequently, in these methods the effects add up. When a method is presented to subjects for the first time, changing the high payoff also classifies them as more risk-loving, while presenting the table with ascending numbers seems to classify subjects as more risk-averse, although these two effects seem to vanish when the same tables are presented to the individuals again. Note that we neither observe order effects nor the significance of control variables.

3.5.4. Method Quality Indicators

We use two avenues to measure a method's quality: Its predictive power and its stability. In order to see which method predicts behavior best in our benchmark games, we look at three statistics: the predictive power by simple OLS regression, the predictive power by Spearman rank correlation, and the absolute average deviation from the prediction.

In Table 3.7, we see the outcome of OLS regressions in the upper part, while controlling for personality measures and socioeconomic variables. In the lower part of Table 3.7 you see Spearman rank correlation coefficients. We include these because besides the absolute size of the elicited coefficients, the correct rank ordering of subjects is essential, since these methods are often used to control for the role of risk attitude in various settings. The OLS regression can be understood as follows: The dependent variable is the outcome of a particular benchmark game, and the independent variables include the outcome in terms of the elicited risk aversion parameter ρ corresponding to one of the risk elicitation methods plus all controls mentioned above. The results do not change qualitatively if we use the switching points as independent variables instead of ρ. We also note that it is not possible to add controls in the Spearman rank correlation. The resulting coefficients in the investment games are negative because a higher ρ implies risk-averse behavior, and therefore lower investments and bids in the benchmark games; the reverse is true for the auction. The corresponding adjusted R^2 values can be found in parentheses below the coefficients. The OLS regression equation is then given by

$$BG_{i,j} = \beta_0 + \beta_1 MPL_j + \sum_{k=2}^{6} \beta_k BIG5_k + \sum_{l=7}^{11} \beta_l SE_l + \beta_{12}\, CRT + \varepsilon_i \qquad (3.8)$$

where *i* denotes the index of benchmark games (*BG*), *j* denotes the index of risk measures, *MPL* denotes the outcome of a risk elicitation method, *BIG5* denotes personality measures according

to the BIG5 test, *SE* denotes the socioeconomic variables and *CRT* denotes the number of correct answers in the cognitive reflection test.

Table 3.7: Explanatory Power

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall	GQ	FQ
OLS coefficients											
Auction	0.68	0.52	0.57	-0.03	-0.30	0.03	0.00	0.58	-0.52	0.13	-0.08
	(0.05)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
Inv. Low	-0.09	-0.94*	-0.48	0.01	-0.67	-0.39	-1.48***	-0.23	-0.44	0.30*	0.09
	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.08)	(0.00)	(0.00)	(0.02)	(0.00)
Inv. High	-0.90**	-0.68	-0.28	0.22	0.58	-0.46	-0.66	-0.65	0.20	0.28**	0.25**
	(0.16)	(0.12)	(0.09)	(0.09)	(0.11)	(0.09)	(0.11)	(0.12)	(0.08)	(0.13)	(0.13)
Spearman rank correlation coefficients											
Auction	0.23*	0.09	0.14	0.17	0.07	0.11	0.06	0.16	-0.13	0.10	0.11
Inv. Low	0.02	0.19	0.17	0.06	0.06	0.11	0.36***	0.12	0.04	0.19	0.11
Inv High	0.28**	0.28**	0.05	0.00	0.03	0.13	0.26**	0.23*	0.09	0.31**	0.28**

Notes: In the OLS regression, the dependent variable is the outcome in one of the four benchmark games, the independent variables are the outcome in terms of CRRA-coefficients from one method plus controls (age, gender, BIG5, CRT test, income, years of university education). The adjusted R^2 value for the regression can be found below a coefficient. Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

Additionally, we can calculate a point prediction in each of the benchmark games for each risk elicitation method (but not for the questionnaires). In Table 3.8 we report the absolute average deviations from these predictions, averaged over all three benchmark games according to the formula

$$AAD = \frac{\left(\sum_{i=1}^{n}|H_i - H_i^*| + \sum_{i=1}^{n}|L_i - L_i^*| + \sum_{i=1}^{n}\frac{|A_i - A_i^*|}{2}\right)}{3n} \tag{3.9}$$

where H_i denotes high investment game outcomes, H^* is the high investment game predictions (L stands for investment low and A for auction) and n is the number of participants in our experiment. Note that we divide the deviation in the auction game by 2 because the choice range in the auction game is twice as high.

In the auction, none of the methods produce statistically significant results in the OLS regression. This is puzzling, as the auction can by itself be seen as a risk elicitation procedure, albeit with heavy framing. Recent literature, however, provides evidence that not only risk attitudes but also other factors like regret aversion (Engelbrecht-Wiggans and Katok, 2009)

could drive behavior in auctions. As far as the Spearman rank correlation is concerned, the SGp method is the only one that is rank correlated ($p<0.05$) with auction behavior.

Table 3.8: Deviations from Predictions

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall
Deviation	1.91	2.41	2.19	2.03	2.11	2.27	1.75	2.17	2.11

Notes: Absolute average deviations from the predictions in the benchmark games.

In the investment games, the methods produce much better results. In the low investment setting, PGhigh has the biggest explanatory power, with SGhigh being a close follower. Note that it is surprising that PGhigh is the best predictor both in the regression and the rank correlation, as the investment games can be interpreted as standard gamble methods themselves, so one would expect one of these methods to perform best. In the high investment setting, many methods (PGhigh, PGlow, SGhigh, SGlow, SGp, and the questionnaires) are able to explain part of the variance, with SGp being the one giving the best results ($p<0.01$). Note that in this setting, survey-based measures perform very well, so questionnaire measures seem to serve as good proxies for subjects' risk preferences in some circumstances. Note that the adjusted R^2 values are relatively low in general, since we added the above mentioned controls to our regressions, which are not able to pick up much of the variation. In all regressions, none of the controls were significant at $p=0.05$. This implies that behavior seems to be driven primarily by risk attitudes. As far as the deviations from the predictions are concerned, PGhigh performs best with an average deviation of 1.75 across all benchmark games with SGp and SGlow also having low deviations. Note that one might be concerned with this analysis for the following reason: If a method generally classifies subjects as risk-averse, it is not surprising that it explains behavior well in the low investment setting, as subjects naturally behave risk-averse in this setting due to the parameters. However, this critique is not valid for any method that provides good predictions across multiple benchmark games (e.g. PGhigh).

In conclusion regarding H3.2, PGhigh and SGp yield the best results in explaining behavior, with PGhigh having the lowest deviation from the predicted behavior in the benchmark games. We conclude that PGhigh has the highest predictive power with SGp being a close runner-up. Additionally, we relax our assumptions on CRRA and perform robustness checks taking CARA, DRRA, DARA, IRRA and IARA into account in Tables 8.1.14 – 8.1.25 in the Appendix 8.1.5. We used subjects' self-reported monthly income and expenditure as a proxy for their wealth. This was necessary, since our subject pool consisted of students who are

not expected to have any wealth, but their monthly income or expenditure can serve as a good proxy for this purpose. The reason for this is that their monthly income is, in general, highly correlated with their wealth (Persson and Tabellini, 1994) and social class. Furthermore, due to ample evidence on the violations of EUT, we provide the same regressions by taking probability weighting, prospect theory and cumulative prospect theory into consideration. For prospect theory and cumulative prospect theory as well as probability weighting functions, we used the functional form and parameters provided in Tversky and Kahneman (1992) and in Quiggin (1982) to create the tables in the Appendix 8.1.6. If we consider other specifications for PT provided in Camerer and Ho (1994) and in Wu and Gonzalez (1996), as well as for PW in Prelec (1998), we do not get different results. The specifications and parameters for DRRA and IRRA can be found in Andersen et al. (2012) and in Saha (1993), respectively, for IARA and DARA in Saha (1993).

The results show that our findings remain quantitatively and qualitatively the same under different specifications. In general, we see similar explanatory power and – in the vast majority of cases – the same significance levels for the PGhigh and SGp methods, which confirm our findings. In some specifications, we even see that the coefficient for PGhigh becomes significant (for example under DRRA for the investment games with low stakes). Nevertheless, a further justification is that some of the other methods (SGhigh, PGlow and questionnaire methods) lose significance under some of the above mentioned alternative theoretical foundations and functional forms (for example under PT, IRRA or IARA).

We now evaluate the stability of the different MPL representations. Remember that after our subjects had gone through all nine MPL methods, three elicitation procedures were randomly selected and presented to them again. A method can be described as stable if the given answers between the first and the second time a method was encountered are very similar. To analyze this similarity, we use three criteria: the equality of overall distribution, the equality of rank ordering and the absolute average deviation between the first and second answers. For reasons of completeness, we also report the perceived complexity of each method. We do not use complexity as a stability measure, as the impact of a higher perceived complexity is not clear. On the one hand, one might argue that a higher measure in this category implies noisier behavior, on the other hand one might argue that a subject takes more time thinking about the problem at hand.

Table 3.9 reports these measures and it is immediately clear that the methods also have a large variance in terms of stability as H3.3 suggested. In the first column we give p-values from the Kolmogorov-Smirnov test that evaluates whether the distributions of the first and the second time a method is encountered are the same. A significant p-value means that the distributions are significantly different from each other, indicating a low stability of overall distribution across a 30-minute time period. The second column gives the rank correlation between the first and the second time a method was encountered. This measure is important because if a method's overall distribution merely shifted up or down without changing the rank ordering of subjects, this method can also be described as stable since the ordering of subjects remains the same. The third column reports the absolute average deviation (AAD) of our subjects' answers when a particular method is presented to them again, compared to the first time. A lower value is therefore better. The last column gives the means of the perceived complexity of each method on a 1 to 10 Likert scale. To visualize these results, we also report the distributions of the differences in switching points between the first and second time a method is encountered in Figure 3.2.

Table 3.9: Stability Measures

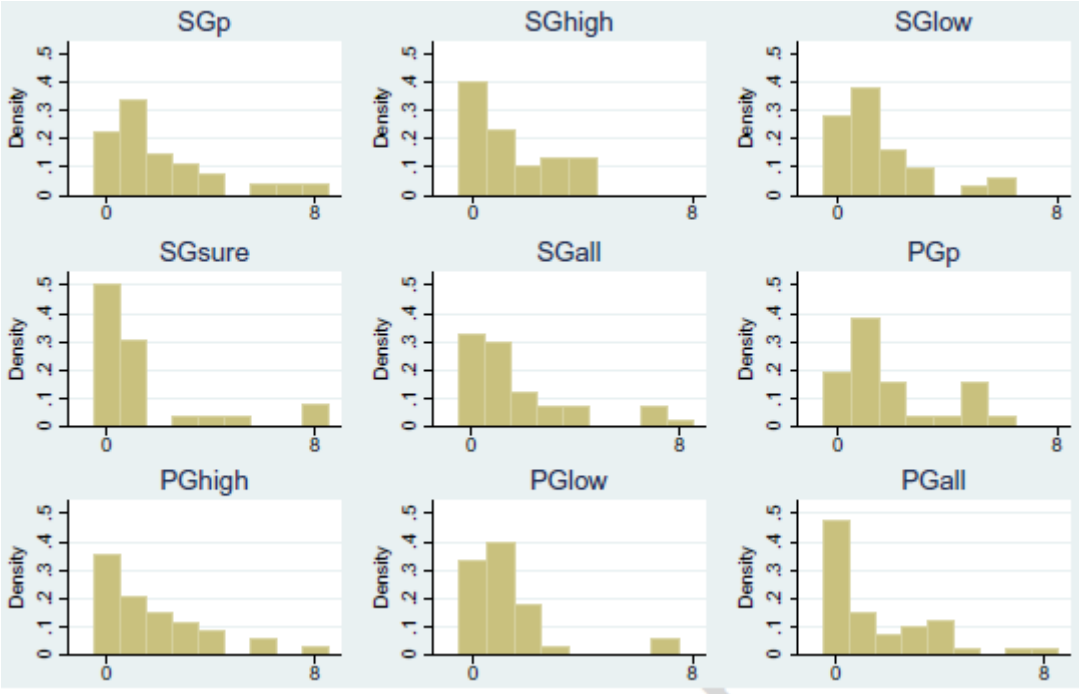
Method	KS-test	Rank correlation	AAD	Complexity
SGp	**0.46**	**0.51*****	1.60	3.42
SGsure	0.01	**0.51*****	**1.37**	3.92
SGhigh	**0.64**	**0.39****	1.48	3.97
SGlow	0.01	0.35	1.96	3.20
SGall	0.01	0.16	1.80	4.81
PGp	**0.24**	0.23	**1.33**	4.21
PGhigh	**0.88**	**0.45*****	**1.24**	3.78
PGlow	0.01	0.25	2.04	4.29
PGall	0.01	0.19	1.85	5.75

Notes: First column: p-values for the Kolmogorov-Smirnov test of equality of distributions. Second column: rank correlation between the distributions of first and second answers (stars indicate significant rank correlation). Third column: absolute average deviation (AAD) between the first and the second decision for the same method. Fourth column: this indicates a subject's perceived complexity of a method; Stars are given as follows: *: $p<0.1$; **: $p<0.05$; ***: $p<0.01$.

Any method that does not yield stable results over a 30-minute time period cannot be described as stable, and stability is a highly preferable characteristic in a risk measure. For the KS-test (column 1 in Table 3.9), the stability relative to the other methods is indicated by a non-significant result: For any methods with a significant result, the overall distributions of answers are different between the first and the second time a method was encountered. Four of the methods have a non-significant p-value: SGp, SGhigh, PGp, PGhigh. A significant rank

correlation (column 2 in Table 3.9) also indicates a stable risk measure, as there is no change in the rank ordering. We see that three of those four methods have significant rank correlations: SGp, SGsure and PGhigh. A low absolute average deviation in answers is also an indicator of a stable risk measure, and the method with the lowest deviation is PGhigh, followed by PGp and SGsure. Concerning the complexity, we see that a method that is perceived as less complex does not necessarily imply more stability in answers, as SGlow has the lowest complexity rating, yet it is classified as unstable in all three categories.

Figure 3.2: Differences in Switching Points between the First and the Second Time



Notes: A low value indicates low differences. X-axis: absolute difference between switching points (e.g. risk preferences) of subjects between the first and the second time the method was encountered. Y-axis: the frequency of switching points.

However, a general tendency can be observed that low complexity indicates more stability. As far as the potential relationship between stability and the control variables (*CRT* and *BIG5* scores, age, gender, income, years of university education) is concerned, no significant effects have been found, so the results will be omitted here. Finally, we mention that the slight differences in the number of observations between the repetitions of particular methods – caused by the pseudo-random number generator – do not drive within-method consistency. We conclude that PGhigh is the most stable method, as it is the only method that performs well in all three categories, with the overall distributions of switching points not being significantly different, having high rank correlations and low average deviation. SGp, SGsure and PGp perform well in two of the three categories.

3.5.5. Results Conclusion

In the benchmark games, as far as predictive power is concerned, we conclude that PGhigh has the highest predictive power with SGp being a close second, irrespective of the assumed functional form or theoretical framework. For non-incentivized surveys, our data shows that eliciting preferences with general and financial questions is a relatively good predictor compared to several incentivized elicitation methods. We conclude that only the PGhigh (Drichoutis and Lusk, 2012), PGp (Holt and Laury, 2002), SGsure (Abdellaoui et al., 2011) and SGp (Bruner, 2009) methods lead to consistent results within a 30-minute time frame, and the PGhigh method is by far the most consistent. The PGhigh method's performance is superior to the other methods in terms of deviations from normative predictions, overall and relative stability across time, etc. So it emerges as an objective winner for the best candidate. Our findings are further supported by the fact that we controlled for personality traits, various socioeconomic factors and cognitive reflection in our analyses. Finally, Table 8.1.13 in Appendix 8.1.5 compares the results from our MPL methods to the results in previously published studies. We found in several cases significant differences between our results and the previously published ones, which makes the need for a stable method even more pronounced. Therefore, we conclude that while SGp also has high predictive power and good stability in answers, the most stable MPL method with the highest predictive power is PGhigh, which corresponds to the method derived by Drichoutis and Lusk (2012) in our alternative interpretation.

3.6. Conclusion

We conducted a holistic assessment and analysis of MPL risk elicitation methods that are present in the economics literature with a sophisticated experimental design using a unified framework and representation method. Previous findings in the literature (Crosetto and Filippin, 2016) indicate that between-method consistency of particular methods is low. We confirm this finding by extending our analysis to all popular methods using the same representation. Furthermore, we show that distributional differences among methods are far from negligible. In addition, we investigate the time consistency of all these methods and document substantial differences in a 30-minute time period for most of the methods. All this implies that an arbitrary selection of a particular risk assessment method can lead to questionable results and misleading

revealed preferences. Thus, it matters which elicitation method is used by researchers in order to control for risk and other preferences.

Our main takeaway is that we provide a suggestion which elicitation method should be used based on objective criteria that assess within-method as well as between-method consistency and validity in real-world settings such as investments and auctions. Our suggestion is to use the PGhigh method by Drichoutis and Lusk (2012). This particular method performs best if we look at the absolute deviations from the normative predictions in benchmark games and also in terms of rank correlations. Furthermore, it yields highly correlated results within a 30-minute time frame in terms of individual deviations and overall distribution. These findings remain robust – in some cases even more pronounced – if we relax our assumptions on CRRA to alternative functions such as CARA, DRRA, DARA, IRRA and IARA. Moreover, our conclusions remain the same if we allow subjective probability weighting or if we estimate risk attitude parameters in line with prospect theory or cumulative prospect theory.

In a broader context, one should take care when selecting a risk elicitation method, especially if one aims to control for risk attitudes in potential real-world contexts such as investment into assets. To be taken into consideration are the nature of the task one intends to control for, trade-off effects between noise, exactness and simplicity. Moreover, we find that changing both the potential rewards and probabilities is perceived as relatively complex by subjects and yields inconsistent results. A further point to consider is that varying the potentially achievable minimum payoff seems to induce more risk-averse behavior while the presence of a certainty equivalent fosters risk taking. Cognitive ability, personality traits and other socioeconomic factors seem to be related neither to risk aversion nor to the extent of the consistency we measured. The debate between changing the probabilities or rewards (Bruner, 2009) seems to be far from settled as one of the methods in each context (PGhigh and SGp) delivers promising results. In addition, our findings might provide guidance in implementing other elicitation methods in the MPL format – e.g. loss aversion (Gächter et al., 2010) willingness to pay (Kahneman et al., 1990), individual discount rate (Harrison et al., 2002) – in terms of whether to vary probabilities, rewards or to use a certainty equivalent. On a final note we suggest that the relatively high variation in risk preferences across and within particular methods might not be a mere artifact – especially in light of further recent evidence on time preferences (Andreoni et al., 2015). We encourage further research to shed light on the consistency of other preference elicitation mechanisms.

Chapter 4

4. Biased Probability Judgment and Other Bias in the Newsvendor Setting

4.1. Introductory Remarks

Matching supply with demand is an extremely challenging task in several sectors including manufacturing, wholesale and retail operations. It is commonly assumed that the primary goal of a supply chain member is to achieve this alignment while maximizing its – or the entire chain's – total profit over a particular time horizon. Keeping this goal in their mind, decision-makers place ordering decisions to satisfy the demand. In the academic literature, the newsvendor setting has served as the workhorse for modelling real-world settings involving a single period and perishable products. The normative solution to the problem has been extended to include risk aversion (Eeckhoudt et al., 1995), overconfidence (Ren and Croson, 2013), etc. Furthermore, several pieces of empirical literature investigated the actual behavior in the newsvendor setting with laboratory experiments and highlighted the role of specific preferences and particular biases, all of which can be measured. Chapter 3 provided insights on methodological issues regarding preference elicitation. This chapter will use our previous conclusions upon the implementation of preference elicitation tasks.

Considering the abundance of co-existing explanations and that the majority of articles focuses on just one or two factors separately, the exact role of the behavioral drivers in the newsvendor – and potentially in other supply chain – settings remains unclear. The literature remains inconclusive about whether the documented results are robust if we consider all these explanations together. Are these factors competing or complementary? For example, the question arises of whether overconfidence remains a significant explanatory factor if we control for all the other documented factors. Furthermore, we argue that the previously documented explanatory factors are not collectively exhaustive. For this specific reason, we consider the previously observed factors jointly while incorporating new ones into our analysis.

For example, a large part of the variation can be explained by personality traits in several economics-related settings, such as criminal activity or job market success (Almlund et al., 2011). Personality characteristics also correlate with behavior in a particular production-

related setting (Strohhecker and Größler, 2013), where the relationship between the actual profits and the corresponding personality traits were examined. The authors claim that openness has a significant and positive impact on performance. Nevertheless, their approach measured this effect in a relatively complex environment with setup costs and lost contribution margins. These additional layers of complexity could in fact be behavioral drivers themselves based on the phenomenon that people tend to attach additional psychological costs (Ho et al., 2010) to several cost factors.

Furthermore, decision-makers made repeated ordering decisions in all of the previously published newsvendor experiments. They were told that the demand comes from a particular – in most cases uniform or normal – distribution in each period and that the demand pattern across these periods is not correlated. All these empirical studies document the presence of demand chasing: Data shows that some decision-makers tend to move their orders closer to the previous period's demand realization – with another proportion moving away from it – across periods. This phenomenon can also be observed in the real world and it can be characterized by the following simple anecdotal evidence: The owner of a hot dog stand might order a larger batch of sausages and bread if (s)he observed a relatively large demand in the previous period(s), since (s)he infers that this demand spike is the result of a particular trend that will prevail in the future. If (s)he speculates that the observed demand spike was an outlier and that it will have a negative effect on the upcoming periods' demand, (s)he will, of course, based on his/her belief, order less than usual. Consequently, demand chasing might have a rationale in the real world where the underlying demand distribution is hard to recognize, not to mention if we consider the different beliefs one might have regarding the demand trend.

There is evidence that demand chasing is not limited to simple, isolated instances such as the hot dog stand example, since it can be observed in the ordering practices of large companies and thus more complex multi-echelon supply chains in the retail and manufacturing industries (Cachon et al., 2007). The doubts about the independence of subsequent periods' events – and about the respective probability that they will occur in the future – have been documented in other settings related to operations management, such as making forecasts based on previous periods' outcomes (Kremer et al., 2011). The basic question is whether we see such behavior under a controlled laboratory setting when the demand distribution is known and where no increasing or decreasing demand trends are present.

In particular, we intend to shed light on the drivers of the aforementioned demand chasing behavior. One might ask why particular decision-makers tend to move their orders

away from the demand that was observed in the previous period, whereas others chase it. Can the direction and extent of these adjustments be explained by a modified framework that is based on an out-of-context measure for biased probability judgment? We find that this measure provides answers to both the direction and extent of demand chasing: Decision-makers who are biased in accordance with the hot hand fallacy will adjust their orders towards the previous demand realization, whereas decision-makers who have a significant gambler's fallacy bias will move away from it. Thus, the out-of-context measure for biased probability judgment predicts the upcoming period's order quantity relatively well.

Ultimately, the underlying reasons behind these adjustments and whether or not newsvendor demand chasing can be reconciled under this particular theoretical framework remains unexplained, since it is hard – if not impossible – to apply control mechanisms for beliefs and for all of the complicating factors in the real world. We therefore conduct a laboratory experiment. Based on the aforementioned hot dog stand example, we keep in mind that the underlying extrapolation of past demands to the future is appealing and that it could pay off in the real world, whereas there is no rationale for it in a controlled laboratory environment with an independent and identically distributed demand sequence.

Finally, the relevance of this bias can be questioned if demand chasing is mitigated by giving perfect information on the optimal behavior or by learning. We observe that demand chasing is still present to the same extent after information on the optimal order quantity has been provided or after several periods. Thus, it cannot be attributed to the subjects' inability to calculate the normatively optimal order quantity.

4.2. Literature Review and Hypotheses

In the newsvendor model, the decision-maker is confronted with ordering a particular amount of perishable items for a single period where the demand (D) is stochastic. The underlying demand distribution and its respective parameters are known. For a risk-neutral decision maker, the order quantity (Q) leading to the highest expected profit P is determined by the critical fractile considering the sales price (p), the procurement price (c) and the salvage value (g):

$$Q^* = F^{-1}\left(\frac{p-c}{p-g}\right) \tag{4.1}$$

where F^{-1} denotes the inverse of the cumulative distribution function of D. We distinguish between high margin settings (HMS) – where $p\text{-}c \geq c\text{-}g$ – and low margin settings (LMS) –

where p-c < c-g. Let us denote $E(D)$ as the expected demand and σ as its standard deviation. Equation (4.1) implies that $E(D) \leq Q^*$ in the HMS, whereas $E(D) > Q^*$ holds in the LMS.

Based on the following formula, the expected profit $P(Q)$ can be calculated:

$$P(Q) = \int_{D=0}^{\infty} [p \cdot MIN(Q, D) + g \cdot MAX(0, Q - D) - c \cdot Q] \cdot f(D) dD \qquad (4.2)$$

Where $f(D)$ is the density function of D. For a detailed discussion, we advise the reader to refer to Choi (2012) who provides a summary of the theoretical work on the newsvendor problem and its extensions.

Schweitzer and Cachon (2000) conducted the first experimental investigation of the newsvendor problem. They found that the average orders of subjects are closer to $E(D)$ than to Q^*, especially during the initial periods of their experiment. Due to imperfect learning, subjects managed to overcome this pull-to-center bias to a limited extent after 20 periods. Schweitzer and Cachon (2000) ruled out risk aversion, loss aversion and waste aversion as potential causes of the observed systematic deviations. They found that only ex post inventory error minimizing behavior and demand chasing are consistent with the average order quantity they observed. Nevertheless, there might be individual differences across subjects, which would imply that these factors also play a role. For example, even if the vast majority orders less (more) in high (low) margin settings than risk/loss-aversion would predict, extremely risk/loss-averse subjects might order tendentiously less than their less biased or neutral peers. This is the reason why we argue that factors need to be examined individually. Consequently, several later articles dealt with the extent of systematic individual differences in newsvendor order quantities and their possible origin: Chen and Davis (2014) provide an extensive review on the behavioral newsvendor. For a selected list of potential explanations, the reader is referred to the subsequent paragraphs.

(i) Mean anchoring and the pull-to-center effect have been verified by Bostian et al. (2008) in average orders. However, Lau et al. (2014) highlighted that the vast majority of individual period-specific orders are outside the pull-to-center interval defined by the mean demand and the optimal order quantity. We differentiate between externally provided anchoring (*AE*) (Jacowitz and Kahneman, 1995) and self-generated anchoring (*AS*) (Epley and Gilovich, 2005), so we elicit both for control purposes. Anchoring to externally provided pieces of information measures how susceptible a particular subject is to moving his/her estimate in the direction of an arbitrary figure provided by someone else, whereas self-generated anchoring measures the extent of a person's bias towards the

first intuitive reference point that comes to his/her mind, which, in our case, is *E(D)*. Notice that there is no externally provided anchor in our setting. Nevertheless, we wanted to include this measure as a control mechanism as in Ren and Croson (2013).

(ii) Ren and Croson (2013) provided empirical evidence that systematic deviations from Q^* were related to the extent of overconfidence (*OC*). Their model specified that the overconfident newsvendor ordered closer to *E(D)*. We elicited overconfidence similarly to their practices based on Russo and Schoemaker (1990).

(iii) Moritz et al. (2013) suggested that cognitive ability can be a good predictor for decision-making in the newsvendor setting because they discovered a correlation to cognitive ability, with cognitively able subjects ordering closer to Q^*. In our analysis, *CRT* denotes the number of correct answers to the three questions in the cognitive reflection test (Frederick, 2005). The random error model by Su (2008) is a complementary explanation relying on cognitive limitations that was tested by Kremer et al. (2010). They held bounded rationality responsible for order quantities that deviate from the normative solution in a way that the period-specific order quantity is modelled as a random variable depending on one's cognitive limitations.

(iv) Eeckhoudt et al. (1995) derived the normative risk-averse order quantities assuming constant absolute risk aversion. De Véricourt et al. (2013) documented that risk-averse subjects order less than risk-loving ones. We denote *RI* as a measure for risk attitude, which was elicited with a multiple price list (Drichoutis and Lusk, 2012) in line with our findings in Chapter 3.

(v) Kremer et al. (2014) verified the intention to minimize ex post inventory errors. Their approach can be regarded as a special case of the reference-dependent newsvendor model (Ho et al., 2010), which implies ordering further away from Q^* and closer to *E(D)*. Our focus is on whether regret aversion (Loomes and Sugden, 1982) can explain this phenomenon in the same way as it did in bidding in auctions (Engelbrecht-Wiggans and Katok, 2008). We elicited individual regret aversion (*RE*) based on Raeva et al. (2010). We apply a slightly modified practice so that we have a scale variable instead of a binary one.

(vi) Schweitzer and Cachon (2000) rejected loss-aversion based on aggregate data. In our investigation, we allow individual differences in loss aversion to have an influence on the actually observed order quantities (Wang and Webster, 2009; Herweg, 2013). The impact of loss aversion was documented in various economics-related settings, for instance in investing behavior (Gneezy and Potters, 1997). We elicited loss aversion (*L*) with a

multiple price list in line with Gächter et al. (2010) based on our findings in Chapter 3. One might argue that this approach measures both risk and loss aversion. This is a potential limitation of loss aversion elicitation methods in general; however, Gächter et al. (2010) provide an extensive list of references, all of which agree that this approach measures loss aversion, rather than risk aversion.

(vii) Personality traits explained several real-world outcomes and results in laboratory experiments, even after controlling for various factors: Borghans et al. (2008) provide an extensive review on the topic. We assessed personality in line with Rammstedt and John (2007): Openness (*OP*), conscientiousness (*CO*), extraversion (*EX*), agreeableness (*AG*) and neuroticism (*NE*).

Hypothesis H4.1: Out-of-context individual loss aversion and particular personality traits are significant explanatory factors in decision-making in the newsvendor setting, even after controlling for all the previously documented findings.

(viii) Several papers document a correlation between performance on a particular task – such as real-effort tasks (Segal, 2012), home ownership, investment decisions (Stephens and Tyran, 2015) and income. A similar relation has been documented between wealth – which is usually highly correlated with income (Persson and Tabellini, 1994) – and the quality and consistency of decision-making (Choi et al., 2014). We elicited income (*I*) as a self-reported measure. Furthermore, we elicited the following factors for control purposes. First, we measured the earnings that the participants wished to achieve during the newsvendor task (*SF*). This allows us to see whether the subjects' order quantities were consistent with achieving a particular earnings target, which would be an indicator of satisficing behavior (Bordley and LiCalzi, 2000). We also elicited self-reported complexity (*CM*) of newsvendor decisions on a ten-point Likert-scale. Other factors we used were age (*A*), gender (*G*) and the years of university education (*Y*) as self-reported socioeconomic characteristics.

Hypothesis H4.2: People with a higher income will order nearer to the optimal order quantity even after controlling for all the previously documented findings.

(ix) The evolution of order quantities across periods was also investigated in the past. Demand chasing was documented by Schweitzer and Cachon (2000). They reported that most

subjects' upcoming order quantities tend to fluctuate across periods depending on the previous period's demand realization. They quantify an adjustment parameter (α_i) across periods with $Q_{i,t} = Q_{i,t-1} + \alpha_i \cdot (D_{i,t-1}-Q_{i,t-1})$. This formula represents the direction and extent of demand chasing and provides a forecast for the next period's order quantity. A positive (negative) value implies that particular adjustments towards (away from) the previous demand realization are made. Similarly, the size of the parameter shows the strength of this bias. It seems plausible that the direction and extent of the individual adjustments in newsvendor orders across periods might be attributed to biased probability judgment. Thus, an out-of-context measure for biased probability judgment (BP_i) correlates with the α_i parameter. In that sense, the predictions by $Q_{i,t} = Q_{i,t-1} + \alpha_i \cdot (D_{i,t-1}-Q_{i,t-1})$ and $Q_{i,t} = Q_{i,t-1} + BP_i \cdot (D_{i,t-1}-Q_{i,t-1})$ should correlate with each other, implying that *BP* can be used as an effective predictor for demand chasing and the upcoming order quantities. In general, this suggests that if a person adjusts his/her forecast and the corresponding behavior away from the previous demand realization, (s)he has a gambler's fallacy bias. The logic behind this is exactly in line with what the hot dog stand owner might believe: The past realization of a probabilistic event makes a similar realization in the next period highly unlikely. This bias has been documented in several contexts, such as betting on lottery numbers (Laplace, 1814), gambling (Croson and Sundali, 2005) and financial literacy (Lusardi and Mitchell, 2008). The opposite behavior, adjusting towards the previous demand realization – assuming that a similar streak of events will occur again – would imply bias in line with the hot hand fallacy. This has also been documented, among other areas, in sports (Gilovich et al., 1985), casino betting (Sundali and Croson, 2006) or for people acquiring an insurance after catastrophic events (Kunreuther, 2006). Rabin and Vayanos (2010) reconciled these two frameworks and claimed that they are ramifications of the same phenomenon: Biased probability judgment. We measure this in line with Dohmen et al. (2009).

Hypothesis H4.3: Out-of-context biased probability judgment can explain the direction and extent of order quantity adjustments in the newsvendor setting.

Furthermore, the question arises whether these biases can be mitigated by specific institutions or de-biasing strategies. Empirical work documents the successful role of communication (Bolton and Katok, 2008), competition (Ovchinnikov et al., 2015), task

decomposition (Lee and Siemsen, 2016) and information (Bolton et al., 2012) in de-biasing newsvendor decision-makers. Furthermore, Gavirneni and Isen (2010) claimed that a large proportion of their participants had identified the underage and overage costs correctly, but then failed to convert this information into the optimal order quantity. Finally, Croson et al. (2014) provided evidence that information can almost completely de-bias decision-makers, even in more complex multi-echelon non-perishable contexts. Our approach complements these papers by revealing whether information and learning have de-biasing effects on demand chasing. Based on the aforementioned evidence, we infer that demand chasing will not prevail after information is provided on Q^*. Notice that our approach does not cover the reduction of bias in out-of-context measurement methodologies such as biased probability judgment, loss aversion, regret aversion, etc. Instead, our focus is on de-biasing particular operations-related phenomena – in our case demand chasing. Thus, we proceed in the same way as the above mentioned studies.

Hypothesis H4.4: Newsvendor demand chasing disappears after perfect information is provided on the optimal order quantity.

4.3. Experimental Design

We conducted a laboratory experiment with 123 subjects, using a between-subject design at the Vienna Center for Experimental Economics. The spread of earnings was between €9.00 and €42.50, which means a payment of €19.36 on average. Sessions lasted 2 hours on average. The experiment was programmed and conducted with z-Tree (Fischbacher, 2007), and the recruitment occurred with ORSEE (Greiner, 2015). We decided to exclude business and economics students, as introductory courses on operations management usually contain lectures on the nature of the newsvendor setting. One subject had to be sent because he/she was unable to answer our control questions correctly. We implemented the random lottery incentive system (Cubitt et al., 1998), which suggests incentivizing preference elicitation procedures, as well as one or two random periods of repeated games in order to avoid transparent hedging effects and to keep subjects focused throughout the entire experiment. Subjects received their instructions, including the conversion rate of experimental currency (EC) to euro, at the beginning of the experiment. The earnings were listed in EC during the experiment, and converted to euro in the end. The experiment consisted of three different parts: Part One, Part Two and Part Three, and it ended with a questionnaire. Each part had control questions to

ensure that the subjects understood the task. Appendices 8.2.1-8.2.8 contain the text of the instructions and the detailed procedures.

4.3.1. Part One

In Part One, six tasks were presented to the subjects. They were told that one of the tasks will be randomly chosen as payoff-relevant. Each of the six tasks elicited individual preferences in the domains of risk attitude, loss aversion, biased probability judgment, overconfidence, self-generated and externally provided anchoring based on standard methodology. These tasks had roughly the same stakes as the newsvendor task. We incorporate all the aforementioned factors into our analysis later in equation (4.3) – see below. A description of our out-of-context individual-level variables with the corresponding meaning, assessment, calculation methodology and descriptive statistics is summarized in Table 4.1. We determined each factor independently from the newsvendor setting. For now, we note that we imitated commonly used elicitation procedures for our individual factors, keeping in mind that, to avoid fatigue, the experimental sessions should not take longer than 2 hours. However, we agree that this is a limitation to the results in Chapter 5, since, in most cases, these procedures rely on a single item.

4.3.2. Part Two

In Part Two, the subjects received instructions for the newsvendor task. They were told that they were about to participate in a repeated setting for 35 periods where they had to take the role of a store owner who ordered and sold 'WIDGETS'. The newsvendor setting's design parameters followed those of Benzion et al. (2008). Our instructions made it clear that the demand is independent and identically uniformly distributed between 0 and 300. In practice, different random demand sequences were generated in advance for each subject. We verified that each of the sequences came from a uniform distribution between 0 and 300 (Kolmogorov-Smirnov test, $p>0.24$). We decided to use multiple demand sequences so that we can claim that our results are not driven by (a) particular demand sequence(s). In order to avoid extreme perceptions that might result from some demand sequences, we ensured that the demand patterns were not negatively or positively correlated and that the mean demand was between 145 and 155 for each of them.

Table 4.1: Descriptive Statistics of Elicited Preferences

Factor	Methodology	Calculation method	Desc.	Rel. st. dev.	Min.	Max.
Ext. p. anch.	Jacowitz and Kahneman (1995)	(True value - answer) divided by (true value - anchor)	0.55 (1.17)	2.13	-2.97	3.77
Self-gen. a.	Epley and Gilovich (2005)	(True value - answer) divided by (true value - anchor)	0.57 (1.59)	2.79	-2.85	9.03
Overconf.	Russo and Schoumaker (1990)	Estimated correct answers - actual correct answers	3.29 (2.24)	0.68	-2.00	9.00
Cognitive ab.	Frederick (2005)	4-point scale based on the number of correct answers	1.50 (1.21)	0.81	0.00	3.00
Risk attitude	Drichoutis and Lusk (2012)	Constant absolute risk aversion	0.01 (0.02)	2.00	-0.07	0.07
Loss av.	Gächter et al. (2010)	Linear functional form under prospect theory	1.76 (1.27)	1.34	1.00	6.75
Regret av.	Raeva et al. (2010)	10-point Likert-scale question	3.15 (2.29)	0.73	1.00	10.00
Openness	Rammstedt and John (2007)	Average of two 5-point Likert scale questions	3.07 (0.82)	0.27	1.00	4.50
Conscient.	Rammstedt and John (2007)	Average of two 5-point Likert scale questions	2.88 (0.78)	0.27	1.50	4.50
Extravers.	Rammstedt and John (2007)	Average of two 5-point Likert scale questions	2.89 (0.84)	0.29	0.50	4.50
Agreeabl.	Rammstedt and John (2007)	Average of two 5-point Likert scale questions	2.09 (0.84)	0.40	0.50	4.50
Neuroticism	Rammstedt and John (2007)	Average of two 5-point Likert scale questions	2.28 (0.99)	0.43	0.50	4.50
Income	Self-reported with a question	-	716 (607)	0.85	0.00	6200
Age	Self-reported with a question	-	24.89 (3.34)	0.13	18.00	37.00
Gender	Self-reported with a question	-	0.49 (0.50)	1.02	0.00	1.00
Y. of u. ed.	Self-reported with a question	-	5.48 (3.25)	0.59	1.00	14.00
Satisficing	Bordley and LiCalzi (2000)	Self-reported target income	2.43 (2.73)	1.12	0.00	13.33
Complexity	Self-reported with a question	10-point Likert-scale question	4.00 (2.24)	0.56	1.00	9.00
Biased prob. judg.	Dohmen et al. (2009)	Belief difference between the first and second cointoss	-0.02 (0.16)	6.56	-0.50	0.50

Notes: This table summarizes our elicited factors, the corresponding literature and descriptive statistics. The column 'Desc.' contains descriptive statistics: Means and standard deviations in parentheses. 'Rel. st. dev.' is the relative standard deviation. 'Min.' ('Max.') abbreviates the smallest (largest) value in our sample.

The subjects received information about the sales price (p=12 EC), the salvage value (g=0 EC) and the procurement price (LMS: c=9 EC; HMS: c=3 EC). 62 subjects participated in the HMS setting and 60 took part in the LMS setting. The normatively optimal order quantities based on equation (4.1) were 225 in the HMS and 75 in the LMS, respectively. The subjects received feedback on the actual demand, the realized sales and their profit/loss after each period. Notice that, in order to avoid experimenter demand effects on demand chasing, the newsvendor decision screens did not contain information about the previous period's demand.

4.3.3. Part Three

In Part Three, we introduced three different treatment variations based on the additional amount of available information. The same instructions were provided as in Part Two, with the following exceptions: In the control treatment (CT), the newsvendor game was played with exactly the same specifications as in Part Two. In Treatment 1 (T1), subjects received the following information on Q^*: 'The game was simulated one million times with random order quantities and random demand. The order quantity providing the highest average payoff is given below: The profit-maximizing order quantity is: Q^*'. Then, the newsvendor game was played for additional 25 periods. In Treatment 2 (T2), five different order quantities were provided to our subjects. These corresponded to the previously revealed preferences in the domains of regret aversion, risk attitude, overconfidence and attitude towards losses. In addition, information on Q^* was provided to subjects in T2. Afterwards, the newsvendor game was played again for 25 periods in each treatment. We were motivated to choose this particular design, since we wanted to have sufficient observations in both margin settings. In addition, we also wanted to check whether providing information on the optimal behavior – and the behavior corresponding to the elicited preferences – can mitigate the role of particular biases/preferences, most importantly demand chasing. We had 62 observations in the HMS (CT: 20, T1: 21, T2: 21) and 60 observations in the LMS (CT: 21, T1: 20, T2: 19). To avoid income effects, the feedback on the earnings was provided only after Part Three was finished.

4.3.4. Questionnaire

The experiment concluded with a questionnaire. The subjects received an additional €10.00 on top of their earnings for filling it in. This helped us to avoid a potentially negative overall payment resulting from the newsvendor game and the loss aversion task. There is existing

literature (Harrison et al., 2009) which indicates that a show-up fee promised in advance might distort subjects' preferences. Therefore, we announced the existence of our show-up fee only after subjects had finished Part Three. Moreover, to minimize the chance that participants anticipate a show-up fee, we decided not to invite subjects who had already taken part in at least two experiments before ours.

4.4. Results

4.4.1. Novel Factors

We built regression models to verify our hypotheses. Equation (4.3) incorporates all the previously mentioned factors. We used subindices *t* and *i* for the current period and the individual subject. Intuitively, βs are the regression coefficients with the corresponding indices; *T* indicates the number of periods that are considered in the regression:

$$\left(Q^* - \frac{\sum_{t=1}^{T} Q_{i,t}}{T}\right) = \beta_0 + \beta_1 AE_i + \beta_2 AS_i + \beta_3 OC_i + \beta_4 (4 - CRT_i) + \beta_5 RI_i + \beta_6 L_i + \beta_7 RE_i + \beta_8 OP_i + \beta_9 CO_i + \beta_{10} EX_i + \beta_{11} AG_i + \beta_{12} NE_i + \beta_{13} I_i + \beta_{14} A_i + \beta_{15} G_i + \beta_{16} Y_i + \beta_{17} SF_i + \beta_{18} CM_i + \beta_{19} \cdot \left[BP_i \cdot \frac{\sum_{t=1}^{T} (Q_{i,t-1} - D_{i,t-1})}{T}\right] + \beta_{20} \cdot \frac{\sum_{t=1}^{T} (Q^* - Q_{i,t-1})}{T} + \varepsilon_i \quad (4.3)$$

We used the deviations from the optimal order quantity as our dependent variable rather than the expected profit. This is the only approach that allows us to introduce and measure the role of order quantity adjustments and thus demand chasing and biased probability judgment. On the one hand, we were interested in our newly introduced behavioral factors such as personality traits, individual loss aversion, etc. On the other hand, we wanted to see demand chasing and the effect of biased probability judgment. This is the reason why we ran our regressions with and without the chasing components. To see if our results are robust, we ran our regressions for the HMS and LMS separately. As equation (4.3) shows, we use average order quantities across the indicated number of periods instead of each individual order quantity $Q_{i,t}$. Note that the same subjects' order quantities in various periods cannot be regarded as independent observations. This is the reason why we rely on the aforementioned approach, which has also been used in other studies (Bolton and Katok, 2008; Bolton et al., 2012; De Véricourt et al., 2013; Moritz et al., 2013). All of the above mentioned choices are standard practice and they have also been applied in several studies (Bolton et al., 2012; Moritz et al., 2013; etc.). We also note that the regressions that include demand chasing contain an autoregressive component. This caveat cannot be avoided if we want to include our observations in conventional regressions that can make our results comparable to the existing

literature. Furthermore, this is the only way we can compare the effect sizes of biased probability judgment and the other factors. To find a workaround to this caveat and to ensure the robustness of the connection between demand chasing and biased probability judgment, we will conduct several sorts of additional analyses later.

Table 4.2 summarizes our results: It includes all factors with a significant explanatory power when they are considered separately. We see that the explanatory power of these factors – with the exception of overconfidence, which is responsible for a large proportion of the variation – is relatively small. Nevertheless, we find that the novel factors – loss aversion, income and extraversion – are significant. We also note that there are quite pronounced differences in terms of the significance of particular factors between the HMS and the LMS, which is not uncommon in the relevant newsvendor literature. A different experimental manipulation would be necessary to shed light on the exact reasons. For now, we note that similar results were published on overconfidence (Ren and Croson, 2013), cognitive ability (Moritz et al., 2013) and gender (De Véricourt et al., 2013). In addition, this thesis is the first that experimentally examines the role of income, extraversion or individual loss aversion for the newsvendor setting. Nonetheless, we have to incorporate multiple factors in our regressions in order to see whether these correlations are robust.

Table 4.2: Regressions Introducing Significant Elicited Factors Separately

Factor	1	2	3	4	5	6	7	8
Constant	-61.7***	28.8***	47.4***	-53.1***	98.4***	71.6***	-58.2***	29.0**
	(3.22)	(10.05)	(5.72)	(5.14)	(16.84)	(6.31)	(7.22)	(14.32)
Overconf. (β_3)	0.77***							
	(0.08)							
Cogn. ab. (β_4)		0.35***						
		(0.12)						
Loss av. (β_6)			0.27**	-0.29**				
			(0.12)	(0.13)				
Extrav. (β_{10})					-0.33***			
					(0.12)			
Income (β_{13})						-0.43***	-0.28***	
						(0.12)	(0.11)	
Gender (β_{15})								0.25**
								(0.13)
Setting	LMS	HMS	HMS	LMS	HMS	HMS	LMS	HMS
Sign.	0.01***	0.01***	0.04**	0.03**	0.01***	0.01***	0.01***	0.05**
Adj. R^2	0.58	0.11	0.06	0.07	0.09	0.17	0.11	0.05
Observ.	60	62	62	60	62	62	60	62

Notes: This table presents the results of regressions showing the averages (standard errors) of standardized regression coefficients (β) corresponding to the independent variables listed in the leftmost column. The dependent variable is the individual average of deviations from the normatively optimal order quantity over all the indicated periods. Stars are given as follows: *: p<0.1; **: p<0.05; ***: p<0.01.

Tables 4.3 and 4.4 give further insight into our results by taking various explanatory factors into consideration. Column 1 in both tables confirms H4.1 and H4.2: Income ($p \leq 0.05$) and extraversion ($p \leq 0.05$) remain significant after considering the explanatory factors jointly.

We see that extroverted people seem to outperform their introverted peers. A potential explanation for this phenomenon is that introverted and shy people tend to behave in a conformist way. They are usually less willing to deviate from behavioral patterns that seem trivial (ordering closer to *E(D)*), whereas the less trivial alternative would be to order more/less depending on the margin. That is the reason why this coefficient has a negative value in both the HMS and the LMS. We also see that the more loss averse a subject is, the less (s)he orders. This is reflected by the different signs for the loss aversion coefficients in the HMS and the LMS. Moreover, we also confirm H4.2: People with a higher income have a better on-task performance, which is also observed in other economics-related settings. This implies that there are non-negligible income – and potentially wealth – effects. This finding is robust if we exclude the top/bottom decile or people claiming an income of zero.

Table 4.3: Regressions Introducing Elicited Factors Jointly – High Margin Setting

Factor	1	2	3	4	5
Constant (β_0)	54.93	6.86	7.39	27.11	-20.22
	(65.98)	(12.38)	(12.72)	(26.54)	(32.32)
Externally provided anchoring (β_1)	0.03	0.02	0.02	-0.03	0.07
	(0.13)	(0.03)	(0.03)	(0.06)	(0.07)
Self-generated anchoring (β_2)	0.07	-0.02	-0.02	0.05	-0.12
	(0.13)	(0.03)	(0.03)	(0.06)	(0.07)
Overconfidence (β_3)	0.07	0.01	0.00	0.00	0.02
	(0.12)	(0.04)	(0.04)	(0.07)	(0.09)
Cognitive ability (β_4)	0.44***	0.04	0.04	0.02	0.07
	(0.16)	(0.03)	(0.03)	(0.07)	(0.08)
Risk attitude (β_5)	0.08	0.08***	0.08**	0.08	0.06
	(0.14)	(0.03)	(0.03)	(0.06)	(0.07)
Loss aversion (β_6)	0.12	0.07**	0.07**	0.03	0.11
	(0.14)	(0.03)	(0.03)	(0.06)	(0.07)
Regret aversion (β_7)	0.19	0.00	0.00	-0.04	0.05
	(0.14)	(0.02)	(0.03)	(0.06)	(0.05)
Openness (β_8)	-0.09	0.02	0.02	0.04	-0.01
	(0.13)	(0.03)	(0.03)	(0.05)	(0.06)
Conscientious. (β_9)	0.22	0.04	0.01	-0.01	-0.03
	(0.15)	(0.03)	(0.03)	(0.06)	(0.07)
Extraversion (β_{10})	-0.36**	-0.11**	-0.04	-0.05	-0.01
	(0.15)	(0.05)	(0.03)	(0.06)	(0.07)
Agreeableness (β_{11})	-0.13	0.00	0.00	0.06	-0.09
	(0.13)	(0.03)	(0.03)	(0.06)	(0.07)
Neuroticism (β_{12})	0.00	-0.07	-0.06	-0.12	0.03
	(0.17)	(0.03)	(0.04)	(0.07)	(0.09)

Income (β_{13})	-0.34***	-0.02	-0.02	-0.03	-0.01
	(0.14)	(0.03)	(0.03)	(0.06)	(0.08)
Age (β_{14})	-0.06	-0.04	-0.03	-0.13	0.10
	(0.23)	(0.04)	(0.04)	(0.09)	(0.10)
Gender (β_{15})	-0.34**	-0.05	-0.02	-0.05	-0.03
	(0.14)	(0.05)	(0.03)	(0.07)	(0.08)
Years of university education (β_{16})	0.11	0.07	0.07	0.15	-0.06
	(0.25)	(0.04)	(0.05)	(0.09)	(0.11)
Satisficing (β_{17})	0.14	0.04	0.04	-0.04	0.14**
	(0.14)	(0.03)	(0.03)	(0.06)	(0.07)
Complexity (β_{18})	0.05	0.01	0.01	0.06	-0.05
	(0.14)	(0.03)	(0.03)	(0.06)	(0.07)
Deviation in t-1 without biased probability judgment (β_{19})		0.03			
		(0.03)			
Deviation in t-1 with biased probability judgment (β_{19})			0.21***	0.20***	0.18***
			(0.09)	(0.08)	(0.07)
Deviation in the previous period (β_{20})		0.92***	0.80***	0.77***	0.70***
		(0.08)	(0.07)	(0.07)	(0.06)
AICC	198.44	117.78	114.20	116.35	132.62
BIC	214.35	135.87	132.30	134.44	150.71
Adjusted R^2	0.33	0.82	0.88	0.86	0.78
Number of observations	62	62	62	62	62
Periods	All	All	All	1-35	36-60

Notes: This table presents the averages (standard errors) of standardized regression coefficients (β). The dependent variable is the individual average of deviations from the optimal order quantity over all the indicated periods. Stars are given as follows: *: p<0.1; **: p<0.05; ***: p<0.01.

Table 4.4: Regressions Introducing Elicited Factors Jointly – Low Margin Setting

Factor	1	2	3	4	5
Constant (β_0)	-28.99	-17.90	4.95	13.55	-18.97***
	(32.38)	(29.92)	(17.32)	(20.91)	(29.27)
Externally provided anchoring (β_1)	0.11	0.05	0.04	0.04	0.04
	(0.10)	(0.07)	(0.05)	(0.06)	(0.07)
Self-generated anchoring (β_2)	-0.04	0.01	0.01	-0.01	0.01
	(0.10)	(0.07)	(0.05)	(0.06)	(0.07)
Overconfidence (β_3)	0.83***	0.11	0.09	0.07	0.09
	(0.10)	(0.11)	(0.07)	(0.09)	(0.11)
Cognitive ability (β_4)	0.02	0.00	0.00	0.00	0.00
	(0.10)	(0.07)	(0.04)	(0.05)	(0.06)
Risk attitude (β_5)	-0.19**	0.07	0.05	0.05	0.06
	(0.09)	(0.06)	(0.05)	(0.05)	(0.06)
Loss aversion (β_6)	-0.06	-0.04	-0.05	-0.05	-0.05
	(0.11)	(0.07)	(0.05)	(0.07)	(0.07)
Regret aversion (β_7)	-0.23**	-0.11	-0.11**	-0.11*	-0.13*
	(0.11)	(0.08)	(0.06)	(0.06)	(0.08)
Openness (β_8)	-0.06	-0.01	0.00	0.02	-0.02
	(0.10)	(0.07)	(0.04)	(0.06)	(0.07)
Conscientious. (β_9)	-0.08	-0.03	0.02	0.03	-0.03

	(0.10)	(0.07)	(0.05)	(0.06)	(0.07)
Extraversion (β_{10})	0.04	-0.01	-0.05	-0.10	-0.01
	(0.10)	(0.07)	(0.06)	(0.06)	(0.07)
Agreeableness (β_{11})	-0.06	-0.04	0.00	0.02	-0.03
	(0.09)	(0.07)	(0.04)	(0.06)	(0.08)
Neuroticism (β_{12})	0.07	0.03	-0.01	-0.06	0.07
	(0.10)	(0.05)	(0.05)	(0.07)	(0.08)
Income (β_{13})	-0.24**	0.06	-0.05	-0.04	-0.05
	(0.10)	(0.07)	(0.04)	(0.06)	(0.07)
Age (β_{14})	0.00	-0.02	0.01	0.02	-0.03
	(0.14)	(0.10)	(0.06)	(0.09)	(0.10)
Gender (β_{15})	0.06	0.06	-0.03	0.02	-0.09
	(0.11)	(0.09)	(0.05)	(0.06)	(0.07)
Years of university education (β_{16})	0.00	0.03	-0.01	-0.00	-0.01
	(0.14)	(0.07)	(0.05)	(0.08)	(0.07)
Satisficing (β_{17})	-0.13	0.05	0.00	-0.02	0.02
	(0.10)	(0.07)	(0.04)	(0.06)	(0.07)
Complexity (β_{18})	0.13	0.04	0.02	0.04	0.05
	(0.10)	(0.07)	(0.05)	(0.06)	(0.07)
Deviation in t-1 without biased probability judgment (β_{19})		-0.05 (0.08)			
Deviation in t-1 with biased probability judgment (β_{19})			0.16** (0.09)	0.16** (0.10)	0.14* (0.09)
Deviation in the previous period (β_{20})		0.94*** (0.12)	0.89*** (0.09)	0.88*** (0.10)	0.92*** (0.10)
AICC	199.83	188.85	178.95	177.90	185.05
BIC	412.52	381.38	358.60	356.16	372.63
Adjusted R^2	0.60	0.76	0.83	0.85	0.80
Number of observations	60	60	60	60	60
Periods	All	All	All	1-35	36-60

Notes: This table presents the averages (standard errors) of standardized regression coefficients (β). The dependent variable is the individual average of deviations from the optimal order quantity over all the indicated periods. Stars are given as follows: *: p<0.1; **: p<0.05; ***: p<0.01.

4.4.2. Biased Probability Judgment and Demand Chasing

Next, we analyzed our data, adding the conventional demand chasing component to our regression equations. In the special case of t=1, we cannot measure the effect of demand chasing, so we excluded the first period. The results in column 2 in Tables 4.3 and 4.4 indicate that just the deviation to Q^* in the previous period by itself cannot explain the extent and direction of the deviation in the current period, since it is not significant. However, the average of the previous period's deviation plays an extremely strong role (p=0.01) and the significance of previous explanatory factors often vanishes as a result. This is intuitive, since a large proportion of subjects did not modify their orders frequently, instead they ordered relatively

constant amounts during the experiment. As a consequence, the deviations between Q^* and the average order quantity were very similar in *t-1* and *t*.

The approach presented in column 2 in Tables 4.3 and 4.4 assumes that the adjustment factor α_i is the same for all subjects. Consequently, only the variation in the mismatch between the order quantity and the demand drives the chasing phenomenon here. We see that this factor is insignificant for both margins. The explanation for this is that some subjects adjust their orders towards the previous demand realization, whereas other subjects move away from that quantity, not to mention that there are people who do not make any adjustments at all. Eventually, all these factors neutralize or cancel out in the end. We change this assumption and allow heterogeneous adjustment parameters that can have different sizes and directions. Thus, we use BP_i as a proxy for α_i as displayed in equation (4.3). Subjects with $BP_i>0$ are expected to move their orders towards the previously realized demand realization in every period, whereas subjects with $BP_i<0$ are supposed to choose an order quantity away from the previous period's demand. This model assumes that unbiased subjects ($BP_i=0$) do not vary their order quantities. We see that the chasing factor becomes highly significant ($p\leq0.01$) and picks up the largest part of the variation – except for the previous period's average deviation – in the regression in column 3 in Tables 4.3 and 4.4. The significance of the aforementioned adjusted chasing factor does not change if we use the goodness of fit (adjusted R^2), the information criterion (AICC) or the model overfit prevention approaches for excluding variables from our regression.

Due to the aforementioned autoregressive nature of the regressions, we are interested in whether β_{19} remains significant if we investigate its effect separately for each period with all the previous control factors that we introduced in equation (4.3) still active. The difference between the previous period's order quantity and the demand by not taking biased probability judgment – thus the model excluding BP_i – into account is significant only in 1 (p=0.01), 7 (p=0.05) and 11 (p=0.1) case(s) out of 118 periods in HMS and LMS taken together, which is equal to the amount of expected false positives.

This changes dramatically if we take biased probability judgment into account, since the coefficient is significant in 78 (p=0.01), 97 (p=0.05) and 107 (p=0.1) cases. From these findings, we conclude that the difference between the previous order quantity and the demand realization by itself cannot explain the large within-subject variation of order quantities among periods. However, the direction and extent of biased probability judgment are crucial factors as H4.3 stated. Furthermore, note that the significance level of this factor does not vanish after

perfect information was provided on Q^*, which is contrary to H4.4. Columns 4 and 5 in Tables 4.3 and 4.4 confirm that this parameter is significant after information has been provided to our subjects, which means that the maximum effect information could have on demand chasing is negligible – if there is any at all. Consequently, providing information on the optimal order quantity does not have the expected de-biasing effect, which implies a high relevance for demand chasing.

Nevertheless, it can be argued that the degree of demand chasing changes across periods: Our subjects might use this strategy less after a couple of periods or after they received information on Q^*. The paired sample t-test ($p=0.35$; $N=122$) indicates that the extent of demand chasing is identical in Part Two and Part Three if we compare the average values of demand chasing in both parts. One might still argue that some individuals might learn, whereas others do not. Thus, we compared demand chasing at the individual level for each subject in Part Two ($N=34$) and Part Three ($N=24$). The Wilcoxon signed rank tests indicate that there were significant ($p=0.05$) differences between Part Two and Part Three only in 16 out of 122 cases, which is only slightly higher than the expected false positive rate of 5%.

Figures 4.1a and 4.1b provide further graphical evidence for the connection between biased probability judgment and demand chasing. The figures show the values of the elicited probability judgment parameters and the demand chasing coefficients we estimated across 59 periods for each subject. Figure 4.1a shows that the direction and extent of biased probability judgment resonate well with demand chasing. Moreover, comparing Figure 4.1a to Figure 4.1b suggests that subjects who have an unbiased perception of probabilities exhibit demand chasing to a smaller extent than their biased peers.

Figure 4.1a: Demand Chasing of Subjects Having Biased Probability Judgments

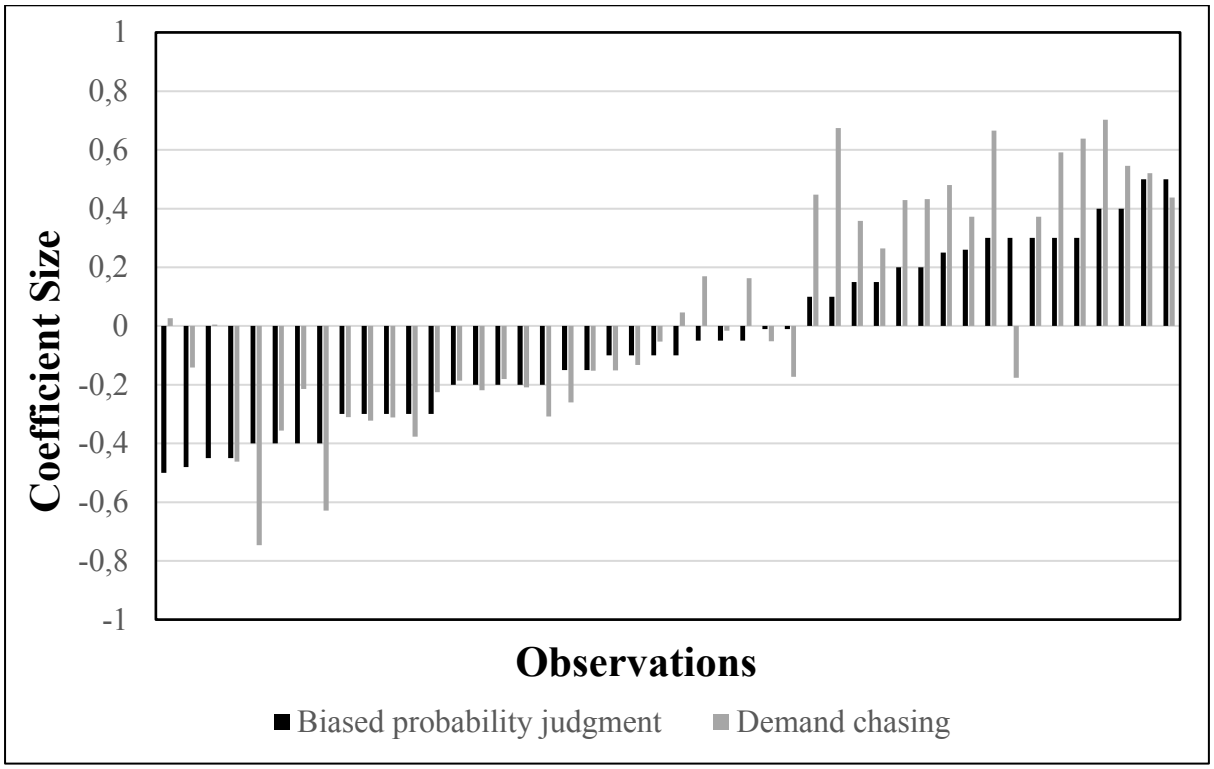
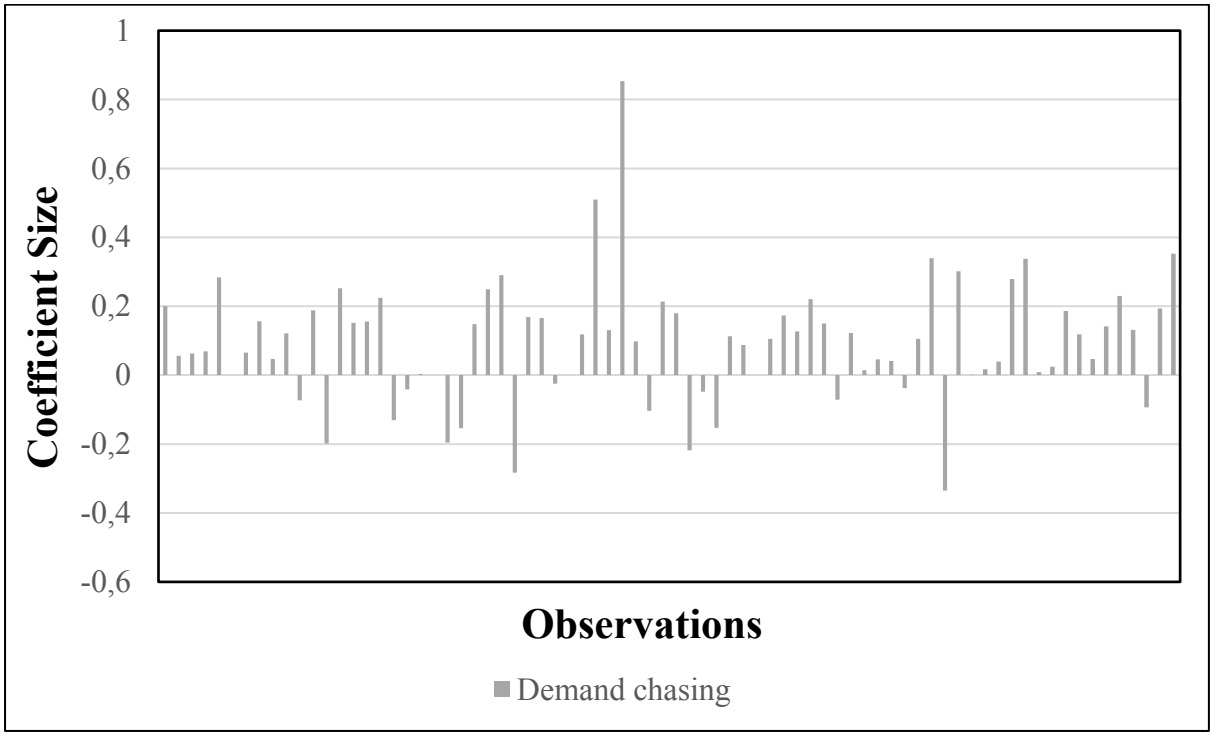


Figure 4.1b: Demand Chasing of Subjects Having Unbiased Probability Judgment



Notes: These figures provide evidence for the connection between the extent of biased probability judgment and demand chasing behavior assessed in line with the suggestions by Lau and Bearden (2013) according to the correlation approach (Bolton and Katok, 2008). Figure 4.1b does not show biased probability judgment as these values are equal to zero for this subsample.

4.4.3. Robustness Checks

Notice that our sample size of 122 subjects is not an outlier, since we reconfirmed previous newsvendor findings on mean anchoring (Schweitzer and Cachon, 2000), risk attitude (De Véricourt et al., 2013), overconfidence (Ren and Croson, 2013), cognitive ability (Moritz et al., 2013), demand chasing (Schweitzer and Cachon, 2000), etc. Similarly, the sample size in the regressions is in line with statistical recommendations (Francis, 2013; Tabachnick and Fidell, 2013) and it is larger than several previous newsvendor studies (Schweitzer and Cachon, 2000; Bolton and Katok, 2008; Bostian et al., 2008; Benzion et al., 2008; Gavirneni and Isen, 2010; Kremer et al., 2010; Bolton et al., 2012; Lee and Siemsen, 2016). Also, we rejected the presence of multi-collinearity in our regressions: The variance inflation factors are below 5 in all cases, whereas there is consensus in the literature that problems arise only if values are above 5. One exception is the years of university education and age, which, thanks to our subject pool, is in general highly correlated. Even in this case, we only observe values slightly above 5 in a limited number of cases. As they are not remotely significant in any of the regressions, this is not a problematic issue.

In addition, we checked for any potential nonlinear (logarithmic, exponential, cubic, quadratic, etc.) relationships between the dependent variable and the independent variables. There is only one instance where we get a robust and significant nonlinear relationship between $D_{i,t-1}$-$Q_{i,t}$ and a particular out-of-context factor: We observed a quadratic relationship in the case of externally provided anchoring ($p=0.03$). However, this factor is not even remotely significant in the multi-factor regression models.

Notice that we used the regression approach in order to compare the effect of biased probability judgment with demand chasing. Nevertheless, Lau and Bearden (2013) criticized this approach, claiming that it might falsely classify subjects as demand chasers who actually follow a decision process in line with the quantal response model. They compared all the four potential demand chasing measurement approaches: Changes towards/against (Benzion et al., 2008; Kremer et al., 2010), adjustment ratio (Schweitzer and Cachon, 2000), regression (Moritz et al., 2013) and correlation (Bolton and Katok, 2008). They found out that the correlation between $Q_{i,t}$ and $D_{i,t-1}$ is the least unbiased estimator for demand chasing, followed closely by the regression approach. Thus, we investigated the connection between demand chasing defined according to the aforementioned correlation approach and biased probability judgment

as a further robustness check: These two measures are highly correlated (ρ=0.60; Part Two separately: ρ=0.53; Part Three separately: ρ=0.47; p=0.01) with each other.

Besides, our results regarding demand chasing are in line with previous findings. The regression approach documented an average chasing coefficient of 0.16, whereas they were 0.15 in Bostian et al. (2008). Bolton and Katok (2008) reported 21% positive and about 16% negative correlations in their corresponding treatments, whereas our sample had 24% and 12% respectively. In addition, our data shows a picture similar to Dohmen et al. (2009) with regard to biased probability judgment: They found that the gambler's (hot-hand) fallacy is exhibited by 21.1% (8.8%) of their sample. These proportions were 23.0% and 12.3% for our respondents.

4.5. Conclusions

Chapter 4 investigates the presence of preferences and biases in the newsvendor context. We find limited evidence that extroverted subjects have systematically fewer deviations from the optimal order quantity than their introverted peers, especially in the high margin setting. This is in line with the intuition that introverted people have a tendency to behave in a conformist way and prefer trivial solutions, whereas extroverted people are more willing to try non-trivial, non-conventional solutions and, as a result, deviate more from the intuitive mean anchor point. This draws the attention to the importance of assessing personality traits when it comes to hiring decisions.

We show the role of differences in individual attitudes towards loss aversion, namely that even though the orders are not consistent with the predictions loss-aversion formulas would indicate, loss-averse subjects tend to order less than their neutral peers.

In addition, we saw that the higher the self-reported income, the less deviations from the normatively optimal order quantities we get. This implies that there are non-negligible income – and potentially wealth – effects: People with a higher income make better decisions in the newsvendor context. However, this finding should be interpreted with care, as the self-reported income of our subjects did not have a clear relation to the nature of the newsvendor task. It is not clear whether people with a higher income order closer to the optimal quantity because they can afford larger fluctuations in their potential earnings. An alternative explanation is that smaller deviations from the optimal order quantity are due to this group's ability, which is reflected by their general superior performance on the task. It is hard to claim

causality as there seems to be a chicken-egg problem between income and high job-specific performance in the real-world, as well.

Foremost and most interestingly, we conclude that biased probability judgment and the underlying gambler's and hot hand fallacies are responsible for the within-subject order quantity adjustments across periods. This means that subjects have a biased belief on the next period's demand and, and as a result, move their order quantities away from/towards the previously realized order quantity. All this occurs because they assume that the upcoming demand realization is not independent from the current one. The direction and extent of out-of-context biased probability judgment is a good indicator for whether these adjustments will occur towards or away from the previous demand realization and how large they will be. This observation is consistent with the prevalent and widely accepted ex post inventory error minimizing explanation: People who make biased decisions assume that the next period's order quantity is positively or negatively correlated with the last one. Therefore, they have a biased demand forecast for the upcoming period and adjust their order quantity accordingly. However, there is a relevant caveat to this result: The random demand is generated from particular distributions for the overwhelming majority of newsvendor experiments, which means that the bias in the ordering behavior can be verified. In real life, the true demand distribution is usually unknown and the beliefs on it are even more heterogeneous. For example, when the demand is high, it is unclear whether it reflects a trend in the underlying demand pattern, or if it is just a random outlier realization from an unchanged distribution. If there is an ongoing trend in the true demand – which nowadays is often the case in our rapidly changing, turbulent environment – then demand chasing could be a preferable strategy. Nonetheless, due to the fact that we are not aware of the true demand's nature and its parameters, it is hard to verify this phenomenon in reality. Thus, we emphasize that the nature of the exact setting has to be kept in mind before we start to de-bias decision-makers.

Our experimental design also has some other limitations. Even though we followed commonly used elicitation procedures closely, the actual impact of particular individual factors could have been quantified more precisely if different factors had been elicited with multiple measures. Unfortunately, we were faced with a trade-off between quantifying several factors, the degree of conciseness and the duration of the experiment. As in other newsvendor studies, we draw the attention to the fact that a clear direction of causality between some of our elicited factors – e. g. income – and the performance in the newsvendor setting cannot be proven. Therefore, our results should be taken with a grain of salt.

Moreover, the relatively low proportion of subjects ordering the normatively optimal quantity – compared to Bolton et al. (2012) – could very well originate in the payment scheme or the different way we provided information to our subjects. The intuitive answer to the question why our approach is able to grasp the subjects' ordering preferences in a more meaningful way is the following: The random lottery incentive system has no diluting effects in this case. In particular, risk attitudes, regret aversion, loss aversion and other preferences are not reflected in the order quantities under the pay-all-period scheme, since the difference between the expected and the actually realized earnings is strongly reduced due to the fact that all periods are relevant. Thus, we encourage further methodology-oriented research on whether different payment schemes have an effect in operations-related contexts. Also, we encourage further research to see if focused de-biasing efforts and specific training on the interpretation of independent probabilistic events prove to be beneficial when it comes to decreasing demand chasing and biased probability judgment.

Chapter 5

5. Behavior in Inventory Rationing Games

5.1. Introductory Remarks

We discovered the most prevalent driving factors of newsvendor decisions in Chapter 4. Nevertheless, operations-related decisions usually are not exercised in an isolated context in the real world. There are multiple companies in a supply network which interact with each other. This results in a flow of goods from the producer to the customer in the end. Therefore, as an effort to increase external validity, we take a step closer to settings that are observable every day: We examine operations-related decisions in a network context. We do not focus on individual behavioral drivers here, but rather on what the actual behavior is in two-echelon supply chains. Our ultimate focus is on whether the ordering pattern of newsvendors is different in an isolated context from a setting which has a larger degree of resemblance to the real world. In particular, we investigate a market where two newsvendor-type retailers order from the same warehouse that is known to have a stochastic – and in the vast majority of cases – limited supply. The most prevalent examples are before the Christmas rush or after natural disasters due to supply disruptions.

We assume that there is a competition for products on the procurement side for the retailers. On the contrary to this upstream side, there is no downstream competition, implying that the retailers are monopolists on their markets, serving different customers from each other. Well-known examples for such practices are regional sales channels of multinational companies or distribution networks where the retailers have exclusive rights to sell a particular product. In the interest of keeping our investigation simple, we focus on products with a predetermined price. This implies that we do not consider heterogeneous prices or a potential bidding mechanism as a solution to this particular rationing problem for now. An example for our setting could be launching a new product with a homogenous price on the market simultaneously in various regions where excess demand is expected in the first few weeks – such as the new models of iPhones or the new Star Wars movie coming in movie theaters in December 2016.

Furthermore, we also investigate how the warehouses distribute their goods among newsvendor-type retailers in the presence of limited supply. Do they allocate based on the demand distribution or rather based on the actually ordered quantities? What is the rationing

mechanism that they follow? Another key aspect is whether there are any social preferences exhibited by the warehouse when it is aware of the actually realized profit by the retailers in each period. On the one hand, a clear signal for this would be allocating a larger amount of products to the less profitable retailer. On the other hand, the opposite type of behavior might turn up as well: Namely rewarding the more profitable supplier, thus sending a larger batch of products to him/her. We also differentiate two contract types between the warehouses and the retailers: These are the wholesale price and the revenue sharing contracts. We need this distinction to see whether the trade-off between the warehouse's own profit and the retailers' profit influences fairness concerns.

We provide a laboratory experiment to shed light on the aforementioned behavioral aspects. We find that the warehouses tend to ration their products based on the actual orders retailers placed under the wholesale price contract. Thus, surprisingly the demand distribution of retailers is merely taken into consideration. The rationing practice is somewhat different under revenue sharing, since the majority of warehouses makes the split based on the demand distribution of retailers here. This is intuitive, since a self-interested profit-maximizing retailer would behave exactly like this. The behavior is not or only to a very tiny extent driven by fairness or rewarding concerns under either of the contract types – even though such behavior would not be costly at all. This finding is somewhat contradictory to previous evidence (Katok and Pavlov, 2013; Katok et al., 2014) which proposes fairness concerns as a relevant explanatory factor. We argue that the lack of fairness in our setting is driven by the fact that our retailers do not have the opportunity to reject the proposed allocation mechanisms, whereas the retailers in previous studies were able to reject the offered contracts. This logic originates in the observed differences between the ultimatum and dictator games. Another clear distinction to those settings is, however, that there is no allocation mechanism that could be called as "fair" except for the 50-50 split in those games, whereas proportional rationing based on the retailers' demand distributions or their actual order quantities are intuitively emerging anchors here.

Another interesting aspect is whether the size of the retailers matters. Do smaller retailers get the requested amount of products more often than the larger ones? This would imply that the warehouses follow the logic that "at least one of them received what (s)he ordered". We could not document such a tendency or the opposite one at all in our findings, meaning that warehouses make decisions irrespectively of size concerns. In addition, we address the impact of rationing mechanisms and the order quantity on profits. Our findings

document a moderate but significant impact that the rationing mechanism has on the actually observed profit. The gap between the optimal profit and the actually observed profit is about 3-4%. The reason for this is that, due to limited supplies, the retailers manage to sell the products they receive in the vast majority of the cases. Consequently, we do not see systematic profit differences between the warehouses or the entire supply chain regardless of whether they ration based on the retailers' demand distribution or their actual orders. This is true both under the wholesale price or the revenue sharing contracts.

A further prediction in the theoretical literature is that, in order to get a sufficient quantity of products, the retailers should overstate their order quantities to the warehouses when the supply is limited. This non-truthful behavioral pattern is called shortage gaming. It is the result of limited supplies as well as the assumption that the warehouses will ration taking the actually ordered quantities at least partially into consideration. The ultimate question is whether the retailers engage in such type of behavior and how this can be mitigated in order to increase supply chain profit and efficiency. A key contribution this chapter makes is that ours is the first rationing experiment that uses human subjects on both the warehouse and the retailer side. This allows us to draw an unexpected conclusion: Shortage gaming is not present in our setting at all. This is the exact opposite what Cui and Zhang (2016) or Chen et al. (2012) documented with computerized warehouses. As our setting was not designed to disentangle the lack of shortage gaming, we can only pinpoint potential causes which are the unwillingness to overstate orders to the warehouse when (s)he knows the demand distribution, ex post inventory error minimization resulting from the wide-ranged stochastic demand, the uncertainty on the allocation mechanism, the human warehouses or the partner matching procedure.

5.2. Literature Review

The inventory rationing literature dates back to the seminal paper by Eppen and Schrage (1981) who established the proportional allocation mechanism where retailers receive a proportional amount of products based on their ordered quantity. An alternative to this – the lexicographic allocation rule – was proposed by Lagodimos (1992). In these settings, a particular retailer is given priority over the other one(s) based on an arbitrarily chosen criterion. Cachon and Lariviere (1999b) introduced the uniform and the linear allocation mechanisms. The former one allocates the available amount of products equally among the retailers. For the latter one, the half of the difference between the retailers' total orders and the total products available is

deducted from each retailer, supposing that the deduction is less than the smallest order quantity. Cachon and Fisher (2000) concluded that information sharing improves supply chain performance in such single-warehouse multiple-retailer settings. The allocation types were further distinguished and classified by Lagodimos and Koukoumialos (2008). All of the aforementioned allocation mechanisms have later been extended from monopolistic retailers to Cournot markets (Chen et al., 2013; Cho and Tang, 2014) as well.

The natural question arises of which exact rationing procedure will be followed by our human decision-makers. A further interesting aspect is whether any fairness concerns can be discovered in line with what several studies documented in various economics-related contexts (Fehr et al., 2007; Houser et al., 2012; Brock et al., 2013). Fairness had been modelled in line with cornerstone studies in the economics literature (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002), and it was formalized by Cui et al. (2007) for supply chains afterwards. Further theoretical analysis was provided by Wu and Niederhoff (2014) who quantified the effect of fairness on supply chain performance under various settings such as heterogeneous demand distributions.

The first experimental work in this context (Loch and Wu, 2008) showed that relationship and status-seeking aspects matter in a two-stage game when both parties decide on their margins sequentially. This study provided the insight that this double marginalization procedure leads to a considerable efficiency drop in the supply chain's performance. Along these lines, Katok and Pavlov (2013) investigated a two-party two-echelon supply chain where the warehouse decided on his/her sales price and the retailer had the opportunity to accept or reject the warehouse's offer. The latter option resulted in no earnings for both parties. The authors conclude that inequality aversion is mainly responsible for the observed behavior, but incomplete information and bounded rationality also play a minor role. This setting has further been investigated by Katok et al. (2014) who concluded that even mild fairness concerns can coordinate such channels relatively well. In addition, Ho et al. (2014) extended their model with an additional retailer, and the retailers made sequential decisions whether they accept or reject the wholesale price offered by the warehouse. The authors documented that the subsequent offers by the warehouse as well as the decisions by the second retailer are strongly driven by the behavior of the first retailer: This led to the development of a concept they call peer-induced fairness based on the article by Ho and Su (2009). The recent contribution by Brock et al. (2013) is especially interesting for us: These authors documented that both ex ante and ex post

considerations played relevant roles in decision-making that involve stochastic payoffs. In our setting, the ex ante type of logic would mean decision-making based on the retailers' demand distributions, whereas the ex post type of decisions would imply rationing based on the retailers' exact order quantities. Finally, an alternative framework to fairness was provided by Cachon and Lariviere (1999a) where allocations occurred on the basis of past sales. This is a well-known example in the automotive industry where the better performing retailer is given a larger share of products in the future. Lu and Lariviere (2012) showed that such strategies can persist in the long run.

Besides the warehouses' rationing practices, this thesis provides experimental evidence on the ordering behavior of retailers. We have already derived the optimal newsvendor order quantities for the retailers in Chapter 4. This changes in the presence of competition for products: Previous literature (Cachon and Lariviere, 1999b) proved that the retailers should overstate their orders and thus engage in shortage gaming in the presence of limited supplies when the available products are rationed proportionally. Lagodimos and Koukoumialos (2008) extended this proposition to all allocation mechanisms which are at least partially based on the retailers' order quantities. Our setting has human decision-makers which means that there is a particular degree of uncertainty regarding the allocation mechanism. This highlights that the order quantity that reflects the Bayesian Nash equilibrium is the potential maximum order quantity, as long as it is assumed that the retailers' order quantities have at least a tiny influence on the allocation pattern set by the warehouse. Thus, none of the allocation rules which rely at least to some degree on the actual order quantities should induce truthful behavior from the retailers. The existing experimental literature discusses experiments with automated warehouses and predetermined allocation mechanisms. Thus, there is no uncertainty component there. Chen et al. (2012) documented that shortage gaming was commonly exhibited by the retailers in their experiment. They offered bounded rationality as an explanation for it. In a follow-up, Chen and Zhao (2015) discussed that this phenomenon is present when the retailer's demand is stochastic, but the potential mismatch between the demand and the supply is small. Cui and Zhang (2016) proposed cognitive hierarchy as an explanation for the increasing degree of shortage gaming in their data. Most recently, Park et al. (2015) documented that communicating the actual rationing rule has a major impact on the extent of shortage gaming, while Zhao et al. (2016) concluded that the strategy and the direct elicitation method yield to negligible differences in the behavior of retailers based on neuroimaging.

5.3. The Model and the Hypotheses

We consider a two-echelon supply chain with one warehouse and two retailers and limited supplies. The retailers order from the warehouse who rations between them. Further details on the knowledge level and the parameter values is provided in section 5.4 which discusses the exact experimental design. For now, we distinguish two contract types among the parties. An extensive discussion on the wholesale price contract is provided in Lariviere and Porteus (2001). For the detailed properties of the revenue sharing contract, refer to Cachon and Lariviere (2005). Let us denote each particular retailer with *Ri*. The letter *p* stands for the sales price and *c* is the cost the retailer has to pay for each unit. Consequently, *c* is also equivalent to the profit per unit for the warehouse. Similarly to Chapter 4, we assume that the product has no salvage value. O_i is the ordered quantity by a particular retailer and G_i is the amount the warehouse sends to them. We denote the demand by *D*.

In both settings, the expected profit – *E(P)* – of the warehouse (*W*) and the retailer (*R*) can be assessed with the following formulas for each period considering that *s* is the warehouse's revenue share:

$$P_W = c \cdot \sum_{i=1}^{n} G_i + \sum_{i=1}^{n} \left[\int_0^{G_i} s \cdot p \cdot D \cdot f(D) dD + \int_{G_i}^{\infty} s \cdot p \cdot G_i \cdot f(D) dD \right] \quad (5.1)$$

$$P_{Ri} = -c \cdot G_i + \int_0^{G_i} (1-s) \cdot p \cdot D \cdot f(D) dD + \int_{G_i}^{\infty} (1-s) \cdot p \cdot G_i \cdot f(D) dD \quad (5.2)$$

The conditions $0 < s < 1$ satisfy the classical revenue sharing contract (RS). Notice that $s=0$ for the wholesale price contract (WP). Every allocation yields to equal profits for the warehouse under this specification, since the warehouse sells the products at a fixed price to the retailers. On the contrary, if $s \neq 0$, the warehouse's profit maximum is achieved if the non-stock-out probability and thus the service level is equal for both retailers. This allocation pattern can easily be calculated for the uniform distributions $[0, U_{MAX}]$ if *L* is the available number of products at the warehouse:

$$G_i = L \cdot \frac{U_{MAXi}}{\sum_{i=1}^{n} U_{MAXi}} \quad (5.3)$$

Every rationing procedure which is different from equation (5.3) decreases the expected profit of the warehouse under revenue sharing. The larger the deviation, the larger the difference between the non-stock-out probability of the retailers and thus the larger the decrease in the

expected profit of the warehouse altogether. For each single retailer, the larger the amount of products allocated – up to the optimal order quantity (Q^*) – the higher their expected profit will be. Note that the expected profit decreases if the allocated amount of goods exceeds Q^*, but this was hardly possible in our setting since L was calibrated in this spirit. Recall that Q^* can be calculated with the critical ratio for the revenue sharing contract as well similarly to equation (4.1) in Chapter 4:

$$Q^* = F^{-1}\frac{[(1-s)\cdot p - c]}{[(1-s)\cdot p]} \tag{5.4}$$

Next, we introduce the proportional, the uniform and the lexicographic rationing mechanisms in line with Cachon and Lariviere (1999b) and Lagodimos (1992). These are the most popular and the simplest ones, which we expect to be followed. For a more comprehensive derivation and classification of rules, refer to Lagodimos and Koukoumialos (2008). We denote X as the proxy for the basis of rationing – which can be the O_i, U_{MAXi} and $E(P_i)$ among others. The proportional rule implies that:

$$G_i = MIN\left(O_i, \frac{X_i}{\sum_{i=1}^{n} X_i}\right) \tag{5.5}$$

This mechanism yields that individuals who order more will get a larger batch of products if O_i is used as the proxy, meaning that this is not a truth-inducing mechanism. Similarly, we introduce the uniform rule as:

$$G_i = \begin{cases} O_i & if\ \sum_{i=1}^{n} O_i \le L \\ MIN(O_i, \frac{L}{n}) & if\ \sum_{i=1}^{n} O_i > L \end{cases} \tag{5.6}$$

As everyone receives an equal amount of products, there is no incentive to lie under this allocation rule. Let us now denote the subset (j,k … v) of retailers which are ahead of retailer i in the priority list from the total number of retailers (i … n). The priority rationing rule can be characterized as follows:

$$G_i = \begin{cases} G_i & if\ L - \sum_{j=1}^{v} G_j \ge G_i \\ L - \sum_{j=1}^{v} G_j & if\ 0 < L - \sum_{j=1}^{v} G_j < G_i \\ 0 & if\ L - \sum_{j=1}^{v} G_j \le 0 \end{cases} \tag{5.7}$$

It is also evident that this rule does not facilitate truth-telling. We assume that this mechanism will often be used for cases where one of the retailers is smaller than the other one so that the warehouse can at least satisfy one of them fully and thus it can avoid some of the resulting psychological costs.

> **Hypothesis H5.1:** The priority rule will often be used – especially for cases when one of the retailer('s ordered quantity) is considerably smaller.

A further question we are interested in is the exact rationing rule the warehouses will apply. Consequently, if it is the proportional rule, which parameter will be used as the proxy for X? The obvious candidates are the retailers' demand distribution (ex ante logic), the actually ordered quantities (ex post logic) or the profit realized in the previous period (fairness or rewarding). Notice that rationing based on the demand distribution leads to the largest expected profit under all the mechanisms for both the warehouse and jointly for the two retailers. Any deviation from this rule will result in a decrease – except in the wholesale price case – for the warehouse. In this spirit, we consider that the demand distribution will be the basis of rationing practices under revenue sharing, but also that concerns for favoring the retailer which is lagging behind will be visible under the wholesale price contract. It is also noteworthy that the uniform allocation rule rations an equal amount of products to the retailers by definition. This is eventually equivalent to the proportional rationing rule following the ex ante logic if both retailers have an identical demand distribution.

> **Hypothesis H5.2:** Under revenue sharing, the rationing practices will mostly follow the proportional rule based on the retailers' demand distributions.

> **Hypothesis H5.3:** Under wholesale pricing, the rationing policies will mostly reflect fairness concerns by favoring the less well-performing retailer.

Intuitively, based on these two hypotheses, we can formulate an additional one on the efficiency and the resulting profit levels in the supply chain. As the rational, profit-maximizing prediction is expected to be followed more closely under revenue sharing, we think that the mismatch between the demand and the supply will be larger under wholesale pricing and the resulting profit will be less. Similarly, we expect smaller supply chain profits under wholesale pricing for the retailers if the warehouse follows the ex post logic or if the the warehouse has strong fairness concerns.

Hypothesis H5.4: The mismatch is smaller and the supply chain profit level is higher under revenue sharing and demand distribution-based rationing.

The main prediction based on the rationing rules is the following: As long as the retailers believe that their orders have at least a slight impact on the amount of goods they will get, they should overstate their orders. This is the reason why they will engage in shortage gaming, which we expect based on the large product margin and the previous evidence in the literature.

Hypothesis H5.5: The retailers will engage in shortage gaming in each treatment.

5.4. Experimental Design

We conducted an experiment at the Vienna Center for Experimental Economics with 192 subjects. The subjects got randomly assigned to the role of warehouses and retailers upon arrival. Their role was kept the same throughout the experiment. As we needed groups of three, the excess subjects who could not form a group of three were determined randomly and sent home with a €10 show-up fee. After having familiarized themselves with the text of instructions and having seen a few concise examples, subjects answered control questions. They participated in the inventory rationing task for 20 periods afterwards. In each period, retailers posted their orders to the warehouse first. Then the warehouses determined how they rationed between them. The demand was realized for the retailers and feedback was provided after every period. Each period was payoff-relevant, the subjects earned their cumulative profits in the end. This design choice was made deliberately so that each and every rationing decision by the warehouse had implications for the retailers. This way the warehouses could not have speculated that fairness concerns are non-relevant because only one period was payoff-relevant. Finally, everyone filled in a standard questionnaire on socioeconomic characteristics. The experimental sessions lasted 75 minutes on average. Subjects' earnings were €17.19 with a standard deviation of €5.17. The graphical interfaces were programmed in z-Tree (Fischbacher, 2007), and the ORSEE system (Greiner, 2015) was used for recruiting our subjects.

In general, the retailers were told that they compete with another retailer for a limited amount of supply in each period and that both of the retailers would be served by the same warehouse who would allocate the available units between them. The retailers knew their own and the other retailer's demand distribution. They were also aware that they were serving different markets from the other retailer. The amount available at the warehouse level was not

known to the retailers in advance, however, they were given the information that there might not be sufficient products on stock at the warehouse. The retailers could order a maximum amount of 10000 in each period as in the other rationing studies. After ordering occurred, everyone received feedback in each period on what he/she had ordered, what he/she had received and how much his/her profit had been. They did not know how many units the warehouse had had at its disposal or how many units the other retailer had received. The warehouse was told that he/she would receive orders from the retailers and he/she would have to allocate his/her units between them. The warehouse was neither allowed to allocate more to the retailers than what they had ordered nor to withhold any products at his/her stock. Remember that this would not have made sense since we consider perishable items. The warehouse was also aware of the retailers' demand distributions, the amount they ordered and their earnings during the previous period. An on-screen calculator was also provided for assistance. After the warehouse made its rationing decision, feedback was provided to the warehouse on the allocation as well as on his/her own profit and that of the retailers. These pieces of information and the retailers' demand distribution were also listed on the decision screens in the upcoming period. The retailers received information on their own profit and the number of products the warehouse allocated to them. We used a partner matching: Subjects did not know which people they were paired with, but the groups were kept constant throughout the experiment.

In the end, the amount of products was sufficient to satisfy the retailers in 20-25% of the periods depending on the treatment. Note that we had human subjects in the role of retailers, which did not make it possible to control for what they would order in advance. Nevertheless the ratio of the available products and the total demand (r) had a value of 0.55 in our case, which meant that the level of shortage ($1-r$) in Chapter 5 was comparable – r=0.6 in Cui and Zhang (2016) – or larger than – r=0.90 in Chen et al. (2012) and Chen and Zhao (2015) – in previous studies.

We had three different treatments and each subject participated in just one of them. We used Experimental Currency (EC) for all of them, where 800 EC was equivalent to a payment of €1.00. The conversion occurred to the exact fraction. There was a wholesale price contract between the retailers and the warehouse in ID-WP and DIFF-WP. The warehouses sold their products for 4 Experimental Currencies (EC) per unit to retailers who were able to resell some or all of them – depending on the actually realized demand – for 15 EC per unit. This means

that the critical ratio is 0.7333 and Q^*=220 in ID-WP, Q^*=110 for the small retailer and Q^*=330 for the large retailer in DIFF-WP based on equation (5.4). The difference between these two treatments is that ID-WP had two retailers where the demand sequence followed the uniform distribution between 0 and 300 (U~0;300), whereas DIFF-WP – had two differently sized retailers with U~[0;150] and U~[0;450]. The inequality between the retailers was also represented by the different size of the respective boxes on the respective z-Tree screens. The same non-stockout probability is achieved by allocating 50%-50% under ID-WP and 25%-75% under DIFF-WP. A deviation from this rule does not affect the warehouse's profit under wholesale pricing, but affects the expected profit of the retailers: The more a particular retailer gets, the larger its expected profit is considering that the allocation does not exceed Q^*. The retailers' demand sequences were randomly determined and were not correlated. To ensure comparability, the retailers' combined demand was equivalent in each treatment. In addition, we ran the ID-RS treatment with similarly sized retailers and a revenue sharing contract among the parties. The retailer's sales price remained 15 EC here, but they bought the products for only 1 EC from the warehouse. In addition, the revenue stemming from the actual sales were shared by the retailer and the warehouse by 75% to 25% (s=0.25). The critical ratio is 0.9024 and Q^*=271 for ID-RS. The same non-stockout probability is given by rationing 50% of the available products to both retailers. A key difference to ID-WP is that every deviation from this particular rule decreases the warehouse's expected profit in ID-RS. Note that the amount of available products at the warehouse was determined in such a way that it was hardly ever possible to allocate more than Q^* to the retailers. This is a crucial prerequisite for the fairness interpretation we introduced above: Namely that each and every additional (less) unit allocated to retailer R_i is regarded as benevolent (detrimental) in terms of expected profits.

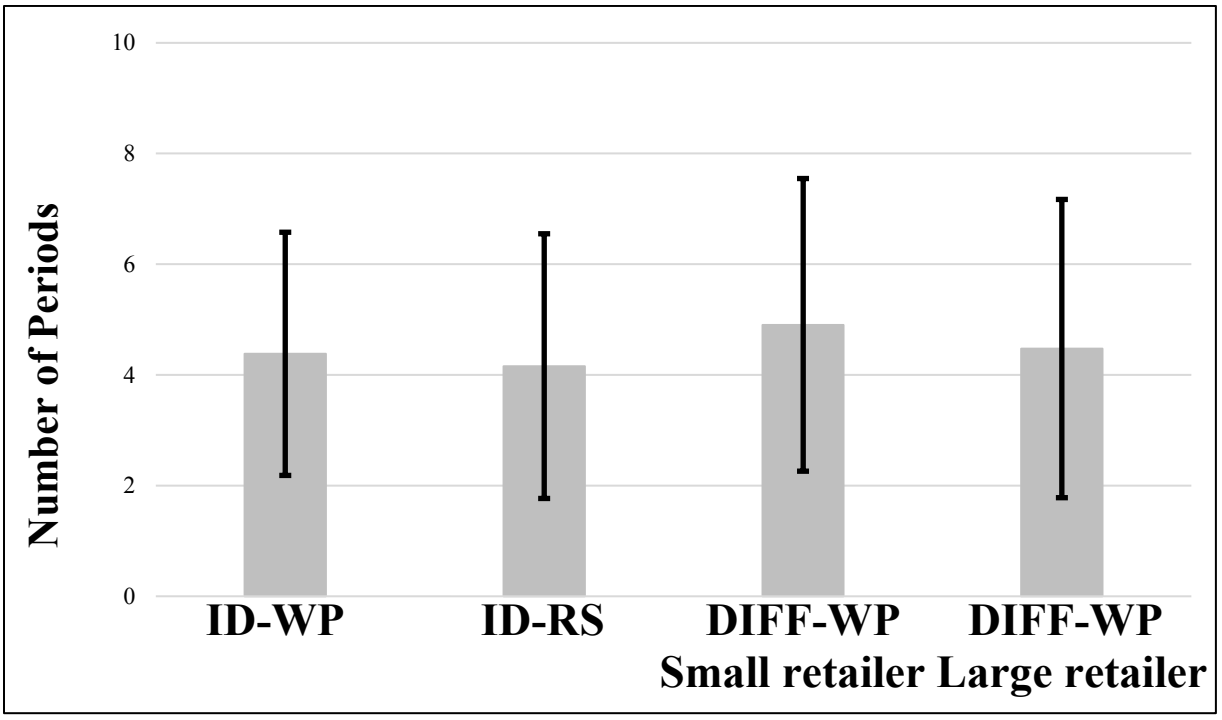
The margin and the revenue share parameters were chosen following real-world industry practices – e. g. the publishing or the flower industry. Furthermore, our parameters also ensured that the average payoffs were roughly equal for the retailers and the warehouse across the treatments ID-W and ID-RS. This could not be guaranteed for DIFF-WP since this treatment had differently-sized retailers. Altogether 21 warehouses and 42 retailers participated in treatments ID-WP and DIFF-WP and 22 warehouses and 44 retailers participated in ID-RS.

5.5. Results

5.5.1. Size Effects

First, Figure 5.1 shows the number of periods when only one of the retailers got fully satisfied, implying that H5.1 is not supported. The respective averages (and standard deviations) are: ID-WP: 4.38 (2.20), ID-RS: 4.16 (2.39), the large retailer in DIFF-WP: 4.48 (2.69), the small retailer in DIFF-WP: 4.90 (2.64). The analysis of variance (ANOVA) confirms that the number of periods is equivalent (p=0.81) across treatments. Furthermore, neither of the pairwise statistical comparisons is significant – the lowest p-value is 0.33. It is obvious that the number of periods is identical across treatments.

Figure 5.1: The Number of Periods When Only One Supplier Got Fully Satisfied



Notes: This figure shows the average number of periods when only one of the retailers got satisfied per treatment. ID-WP abbreviates the treatment with identical retailers under the wholesale price contract. ID-RS stands for the treatment with identical retailers and the revenue sharing contract. The DIFF-WP treatment had a small and a large retailer. Standard deviations were calculated across the subjects in corresponding treatment.

One might argue that there might have been more opportunity to serve at least one of the retailers in some treatments depending on what the retailers' actual orders were. Our results do not change if we control for this and use the corresponding adjusted ratio measure that takes only these periods into consideration. Thus, if the retailer R_i orders the amount of q_i from the warehouse, there is no additional surge in the utility that would stem from shipping q_i to retailer R_i compared to shipping q_i-1. This implies that individuals attach no psychological costs if a

retailer receives less than what (s)he ordered. Note that satisfying at least one of the retailers could have been achieved easily in particular in DIFF-WP, since the large retailer's demand was 3 times as large as the small one's. Next, we explore which rationing rule is used by the warehouses.

5.5.2. The Warehouses' Rationing Mechanism

As we did not observe any tendency to use the lexicographic allocation rule, the warehouse can choose from the three remaining intuitive proportional allocation mechanisms. Recall that the uniform allocation rule is equivalent to the proportional allocation rule based on the retailers' demand distribution where the retailers were identical. The question is whether the warehouse will choose the demand distribution, (U_{MAXi}) the actually ordered quantities (O_{it}) or the realized profit in the previous period (P_{it-1}) as a substitute for X_{it} in equation (5.5). We evaluate this on a period-specific basis with the help of the model confidence set approach (Hansen et al., 2011). This method classifies subjects based on the mismatch compared to all the three models. For each subject and model, we calculate the value of

$$M_i = \sum_{t=1}^{20} |X_{it} - G_{it}| \tag{5.8}$$

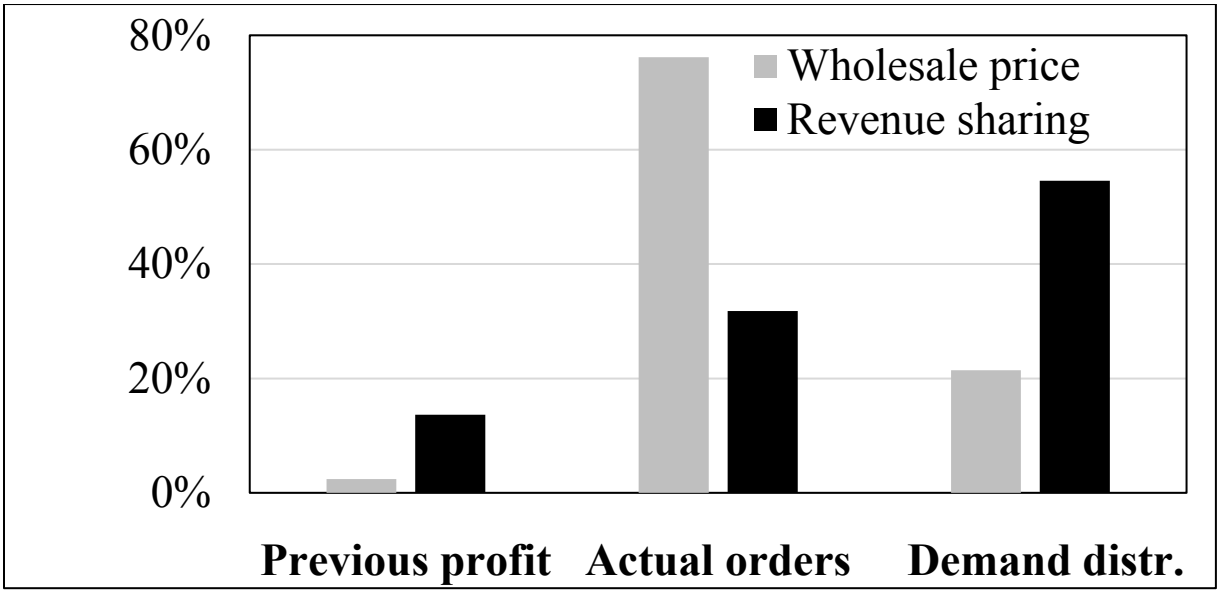
by substituting X_{it} with U_{MAXi}, O_{it} and P_{it-1}. Then we take the smallest M_i value for each subject, which yields the smallest mismatch compared to the predictions. This classifies the subjects according to the three rationing practices.

Figure 5.2 shows that the warehouses use very different rationing mechanisms from each other under wholesale pricing and revenue sharing. The table aggregates the treatments ID-WP and DIFF-WP, since there are no systematic differences between them. The Kolmogorov-Smirnov test shows that the distributions are different (p=0.01). Under the wholesale price contract, 76.19% allocates based on the actual orders posted by the retailers and 21.43% rations taking their demand distributions into account. This changes dramatically under revenue sharing, since 31.82% uses the order quantities and 54.55% rations based on the demand distribution. Interestingly, our individuals' rationing decisions are decisively driven by the retailers' order quantities under wholesale pricing. This is surprising, since every decision screen reminded the warehouses of the retailers' demand distribution, but they decided not to guarantee equal expected profit to their retailers. The main takeaway is that the revenue-sharing contract seems to enhance proportional rationing based on the demand distributions for the

majority of subjects. This coincides with the self-interested behavior, but a considerable proportion still makes his/her decision based on the order quantity here.

We see a negligible proportion (WP: 2.38%; RS: 13.64%) whose rationing decisions reflect the profit in the previous period in both treatments, implying that, in general, such considerations are not behind the observed allocation mechanisms. This is puzzling at first sight, since rewarding the worse or the better performing retailer would not have cost anything to the warehouse. The observed behavior is also quite different from the earlier studies (Loch and Wu, 2008; Ho and Su, 2009; Ho et al., 2014; Katok et al., 2014) which documented fairness concerns. A key difference our setting has compared to the previous ones is that we do not ask the warehouses to set a wholesale price to the retailers, but we request to specify the quantity that ought to be sent to them. A further distinction is that our retailers cannot reject a particular offer, whereas this was clearly the case in the previous papers, which meant zero profit for both parties. This line of thinking is identical to the ultimatum (Güth et al., 1982) and the dictator games (Kahneman et al., 1986): In the former case it is possible to reject an offer, whereas it is not in the latter one. One might argue that we see fair behavior even in the case of dictator games, however, recent evidence (Cappelen et al., 2013) suggests that the observed allocation depends highly on the reference points, namely, whether it is framed as a give or a take. Translating this to our setting means that the distribution- and the order-based allocation patterns emerge as salient reference points in each period. Therefore, subjects seem to choose from these intuitive order- or distribution-based rules.

Figure 5.2: The Basis of Warehouses' Rationing Mechanism Across Treatments



Notes: This table shows which factor serves as the basis of the rationing mechanism under different contract types.

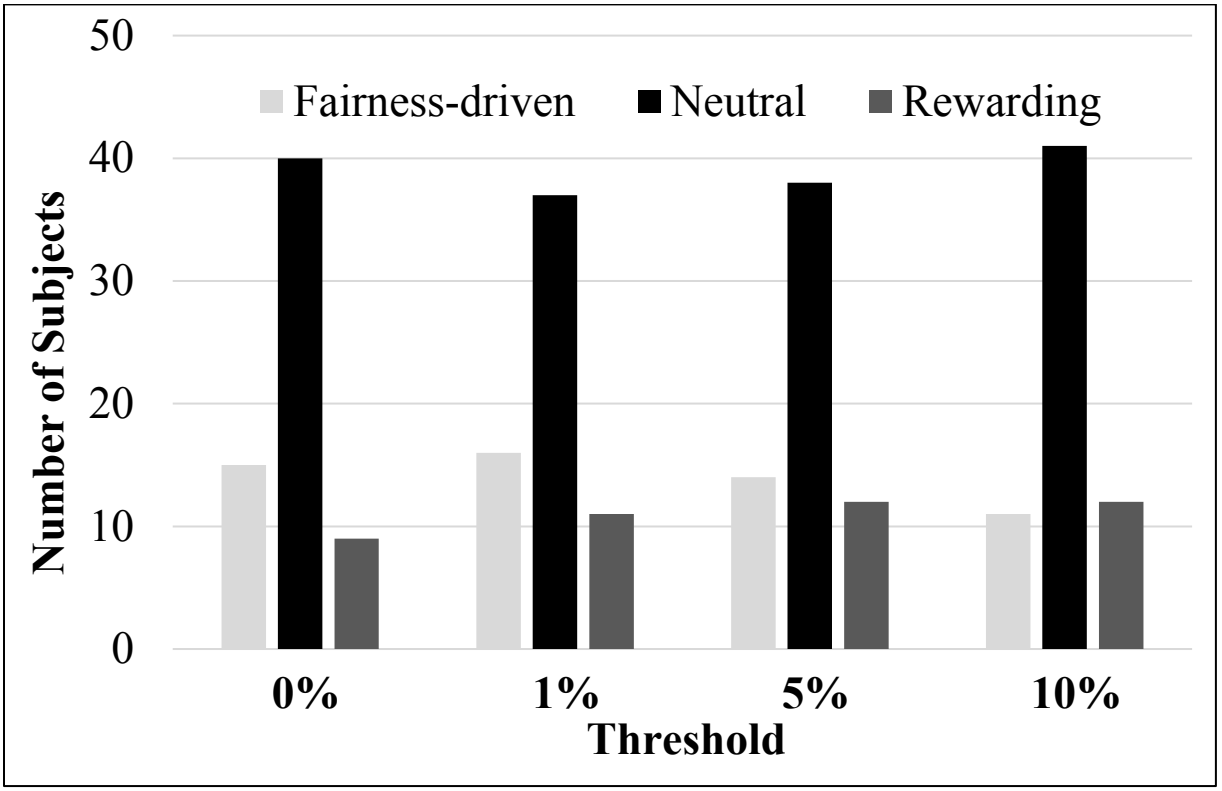
We conducted several robustness checks for our findings. The above mentioned takeaways on the allocation patterns remain the same if we define a quadratic function instead of the linear one in equation (5.8). We get qualitatively the same results if we take the expected profit instead of the actually realized profit into account for the profit-based rule. Notice that if there were inequity aversion concerns by the warehouses, we would not have observed that the allocation was proportional based on the order quantities in the ID-WP treatment. A further evidence that questions the presence of fairness concerns is the DIFF-WP treatment, in particular where the warehouse allowed the larger retailer to make much larger profits than the other retailer or the warehouse itself.

A potential insight might be that there could be slight or moderate fairness or rewarding concerns in our data set, which would imply very small deviations from the distribution- or the order-based rules. Consequently, the model by Hansen et al. (2011) would quantify these deviations in a way that the allocation still occurred based on the orders or the distributions, however, the retailer which is lagging behind would still receive more/less products than the actual order- or demand distribution-based rationing rule would predict period by period. We take this into consideration in the following analysis.

As our subjects might have used rounding upon rationing or might have made calculation errors, we introduced deviation thresholds (e_{it}) of 0%, 1%, 5% and 10% with respect to the proportional rule for each and every subject and period. For example, e_{it}=0% implies that each minor deviation with a minimum of one unit from the exhibited order- or demand distribution-based rationing rule is quantified as showing rewarding or fairness-driven behavior. If e_{it}=10% then those particular periods are not taken into consideration that are within 10% range, implying that periods where $X_{it} - e_{it} \leq G_{it} \leq X_{it} + e_{it}$ are considered as following the actual order- or the demand distribution-based rule. The first indication for the lack of such fairness-driven or rewarding behavior is that, on average out of 20 periods, only 13.5, 11.8, 8.5 and 6.0 periods show deviations in one of the directions under the respective 0%, 1%, 5% and 10% thresholds. Next, we evaluated the periods where a deviation from the proportional or the uniform rule is observable separately by assessing our results for each subject: We calculated the following proportion: We took the number of periods when the retailer that had the larger profit in the previous period was rewarded and divided it by the sum of periods when it was either rewarded or punished. We conjecture that a consistent rewarding behavior would require that a subject should deviate towards the retailer that was better off more than twice as often as

towards the one that lagged behind. Similarly, a fairness-driven warehouse should allocate more units to the retailer that underperformed at least twice as frequently as to the one which was ahead in the previous period. If an individual does not meet this criterion, (s)he is classified as neutral. We show the number of subjects for each threshold level in Figure 5.3. We conclude that these deviations do not show a conscious pattern in any direction and that the vast majority of the subjects fall in the neutral category. These results remain robust if we take the absolute deviations instead of the number of periods or if we take the cumulative profit instead of the previous period's profit into consideration. Moreover, we canot document any systematic differences among our treatments.

Figure 5.3: Archetype of Warehouses Based on the Exhibited Rationing Mechanism



Notes: This table shows the archetype of subjects based on objective criteria which were described in detail in section 5.5.2 of the thesis.

Finally, fairness can be interpreted in a cumulative way, namely that the warehouses might reward one of the retailers in period t and the other one in period t+1. This would imply that there should be a correlation in the direction and extent of deviations with respect to the actual order- or the demand distribution-based rule. We report that there is none, since ρ=0.01.

5.5.3. Profit Implications

We observed that the warehouses follow heterogeneous rationing mechanisms under various contract types. We investigate the profit impact of these choices. Table 5.1 summarizes these results with the averages and standard deviations per subject. First, we compare the warehouses that followed the actual order-based and the demand distribution-based allocation mechanism with each other. The Mann-Whitney-Wilcoxon test (MWW-test) shows that there are significant differences neither for the warehouses (p=0.43) nor for the retailers (p=0.62) between the actually realized profit. A similar picture emerges if we compare the actually realized profit per contract type, in particular for the warehouses (p=0.99) and the retailers (p=0.62) of the ID-RS and ID-WP treatments. Note that we did not include DIFF-WP here, since the retailers had different demand distributions and of course very heterogeneous profit levels in that treatment as displayed in Table 5.1.

Table 5.1: Profit Levels and Statistical Significance

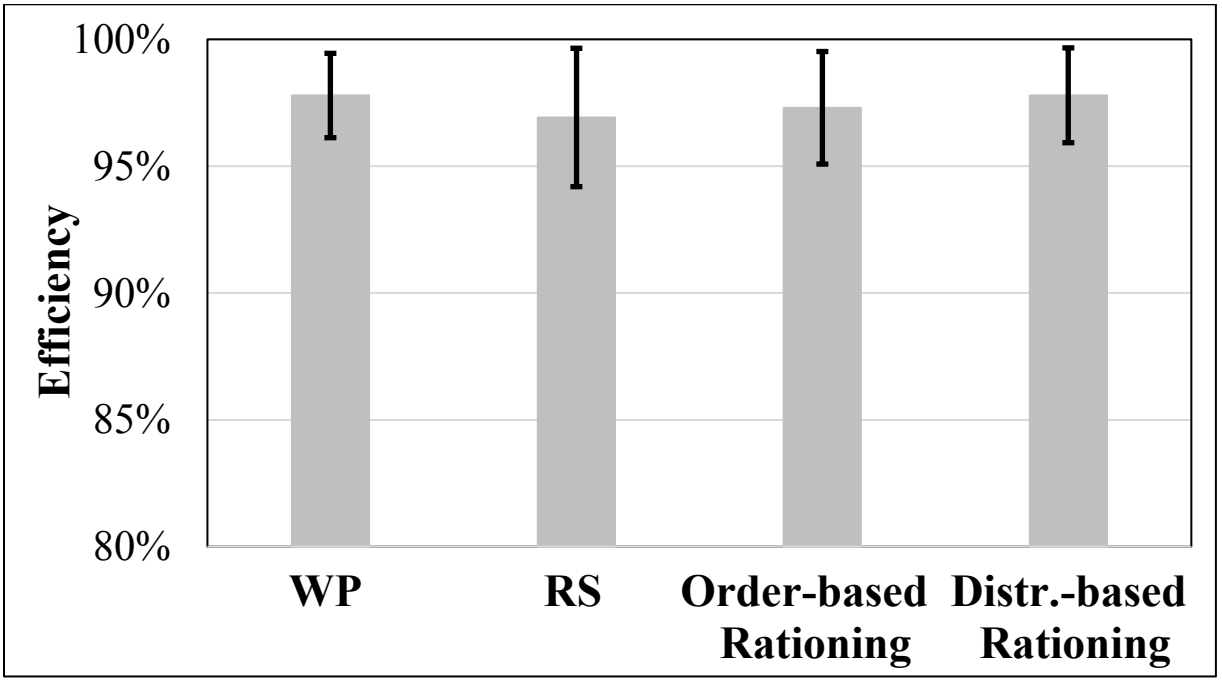
	Actual observations			Optimum
	Warehouses	Retailers	Total supply chain	
Wholesale pricing	14003.43	13477.57	40609.69	41516.27
	(2091.78)	(2655.21)	(4854.27)	(4781.47)
Number of obs.	42	42	42	42
Revenue sharing	14010.01	13810.56	40936.99	42259.73
	(2540.37)	(3547.77)	(6262.60)	(6507.89)
Number of obs.	22	44	22	22
Stat. sign. (p)	0.99	0.62	0.46	-
Order-based rationers	13828.40	13819.97	40277.94	41410.15
	(2046.60)	(3039.66)	(4593.73)	(4766.71)
Number of obs.	39	48	39	39
Distr.-based rationers	14512.04	13569.26	42054.57	42989.52
	(2674.33)	(3135.46)	(6105.85)	(6101.67)
Number of obs.	21	32	21	21
Stat. sign. (p)	0.43	0.72	0.67	-

Notes: This table shows the averages (and the standard deviations) of profits in Experimental Currency. We excluded the profit-based allocators due to their very small number. For the retailers under wholesale pricing, we excluded the DIFF-WP treatment, since the retailers that had rationed based on the demand distributions had ether very large or very small profits due to the 3:1 size difference.

Apparently, it seems that there are only negligible – not even remotely significant – differences among the achieved profits in the supply chains regardless whether revenue sharing

or wholesale pricing is used or the nature of the implemented rationing mechanism. The explanation for this is that the available quantity at the warehouse was relatively low compared to the retailers' average demands. That is why the retailers were able to sell all of their products in most cases. This implies that there are only moderate differences in profits even if the observed allocation was asymmetric between the two retailers. Table 5.1 also reflects these figures per subgroup. Similarly, the figures corresponding to the 'Optimum' were calculated by having taken the available products in each period into consideration. The practice we used here was in line with the profit-maximizing allocation rule. Thus, we followed the proportional allocation based on the demand distributions strictly. We note that neither RS versus ID ($p=0.67$) nor the demand distribution-based versus the actual order-based allocation ($p=0.46$) delivers significant differences between the actual supply chain profits according to the MWW-tests. Thus, we reject H5.4, the impact on profit is minimal. Furthermore, Figure 5.4 shows that there are not large differences with respect to the optimum. It is obvious that the ratio of the actually observed and the theoretically optimal expected profit is statistically identical across treatments, however, the gap to the optimum is highly significant ($p=0.01$) according to the Wilcoxon-tests. Pay special attention to the fact that the differences are only small (about 3%) in each treatment. Again, the explanation is that the retailers were sold out in the overwhelming majority of periods.

Figure 5.4: Actual Versus Optimal Profit Levels

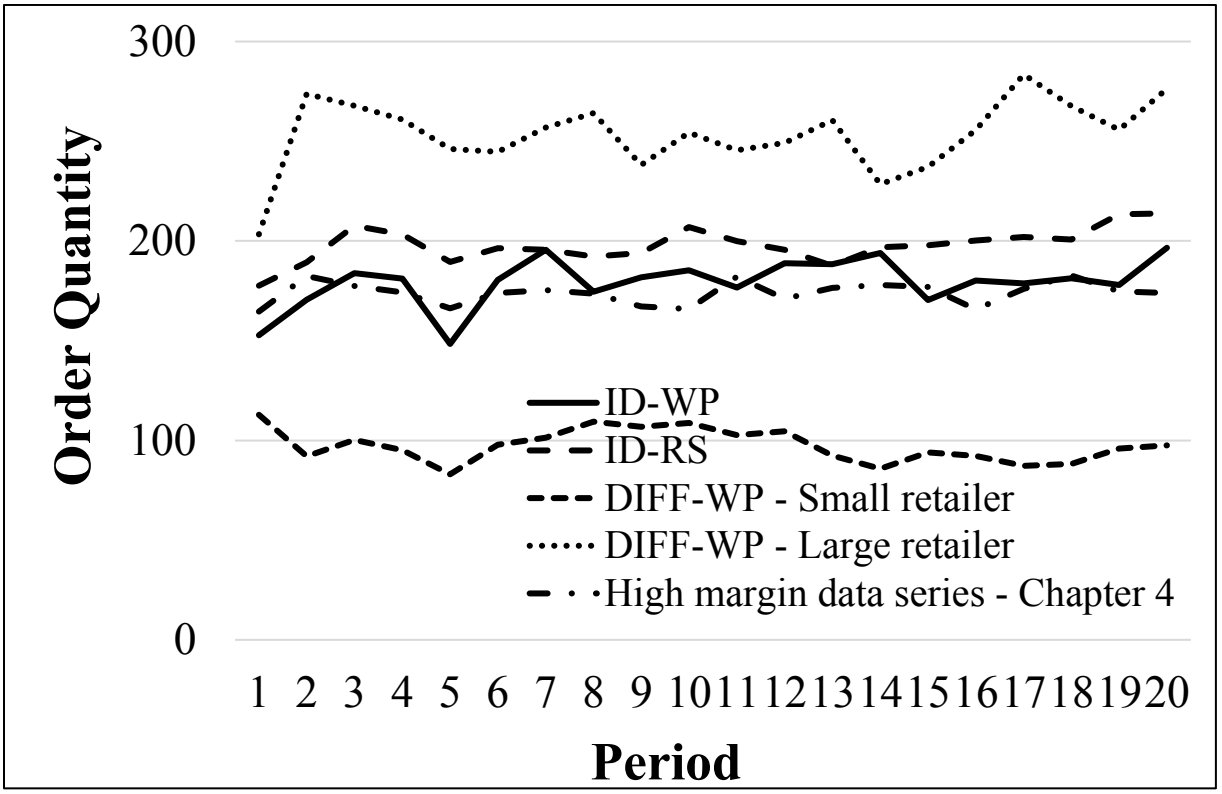


Notes: This figure shows the average percentage of the actually observed profit compared to the optimal profit. The standard deviations are displayed on the subject level.

5.5.4. Shortage Gaming

We now draw our attention to H5.5 and thus to the ordering behavior of the retailers. Figure 5.5 documents that the average order quantities do not show shortage gaming in any of the treatments. Although we see a slightly increasing/decreasing tendency in the order patterns, this is driven by the adjustment towards the optimal order quantity explained in Chapter 4 (Schweitzer and Cachon, 2000) due to learning over time (Bolton and Katok, 2008). The graphs draw a similar picture to previous newsvendor experiments without rationing. The only difference is that large retailers tend to order more and that orders are larger in the revenue sharing treatment. These findings are intuitive, since the critical ratio would predict higher order quantities in these cases.

Figure 5.5: Average Order Quantities Per Treatment

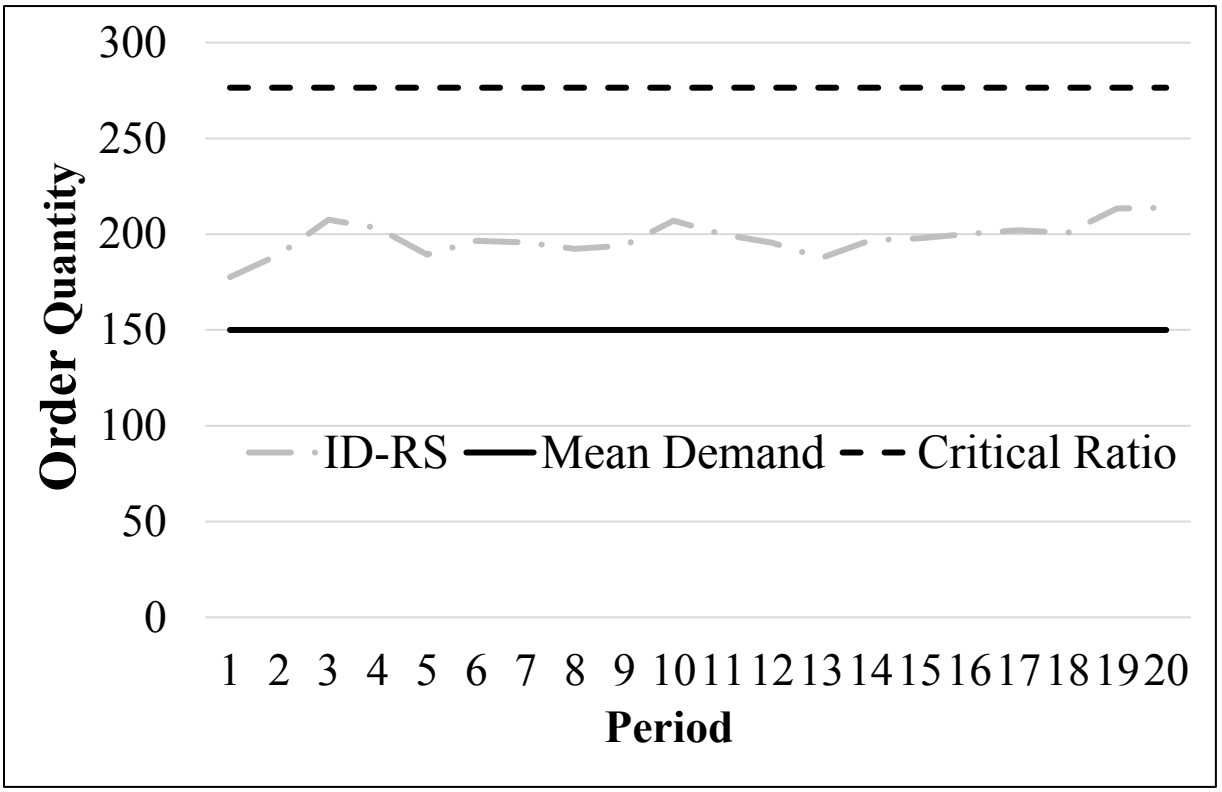


Notes: This table shows the average order quantities over time in different treatments. ID-WP: Identical retailers with the wholesale price contract. ID-RS: Identical retailers with the revenue sharing contract. DIFF-WP: Different retailers where one of the retailers is 3 times larger than the other one under the wholesale price contract. High margin data series – Chapter 4 shows the order quantities of the first 20 periods without limited supply or rationing from Chapter 4.

The lack of shortage gaming is further supported by the fact that the slope of the curves is statistically indistinguishable from the first 20 periods of our high margin data set in Chapter 4. They are 0.92 (3.72) in ID-WP, 0.82 (3.66) in ID-RS, -0.49 (1.99) for the small retailers in

DIFF-WP, 0.92 (4.65) for the large retailers in DIFF-WP and 0.17 (3.18) in the first 20 periods in the high margin treatment in Chapter 4. Pairwise comparisons yield insignificant results according to the MWW-tests ($p>0.37$) and all the slopes of the curves are identical according to the ANOVA ($p>0.53$). We draw the attention to the 'High margin data series – Chapter 4' and 'ID-WP', which are very similar: The critical ratio would indicate an optimal order quantity of 225 in the former one and 220 in the latter one. The lack of shortage gaming is also confirmed in Figure 5.6 where we can see that the average order quantity fluctuates between the mean demand and the optimal order quantity in ID-RS. This is identical to all the previous newsvendor experiments without any rationing. Similarly, we also checked that the correlation between the ordered quantity in the current period is not driven by the level of shortage in the previous period as the correlation between these factors is insignificant and negligible (ρ=-0.08 for ID-RS and ρ=0.04 for ID-WP and DIFF-WP).

Figure 5.6: The Average Order Quantity Over Time in Treatment ID-RS



Notes: This table shows that the average order quantities show, due to the imperfect learning procedure, a slightly increasing tendency in the revenue sharing treatment with identical retailers (ID-RS).

The analogous figures for the other treatments can be found in Appendix 8.3.3. They allow us to draw the same conclusion: Shortage gaming is not present. This sounds surprisingly, since subjects ordered more than 9000 units in previous studies even though they were aware

that their demand was deterministically 50. The initial focus of our experiment was on the warehouse's behavior, therefore our design is not able to pinpoint the exact underlying reasons for this phenomenon. Nevertheless, we can offer several potential explanations for this phenomenon.

The previous rationing experiments had computerized warehouses whereas we had human decision-makers. Thus, an imaginable cause for the lack of shortage gaming can be that retailers exhibit lying aversion towards humans as in other studies reviewed in Gneezy et al. (2013). As a consequence, they are unwilling to report untruthful orders to the human warehouses. Another explanation might be that the warehouse is aware of the retailers' demand distribution and the retailers do not want to look stupid in the warehouse's eyes and they want to maintain a self-image (Ploner and Regner, 2013). It seems plausible that the retailers would have exhibited shortage gaming if they had received information on how many units the warehouse had had at its disposal and how many units the other retailer had ordered. A further explanation might be that the stochastic nature of the retailers' demand might induce some fear to get too many products especially since they can end up with a small demand realization in each period. Consequently, if the retailers follow an ex post inventory error minimizing logic, they will not post large and untruthful orders. Chen and Zhao (2015) tested this and provided a similar experiment where the demand had two stochastic outcomes. Their investigation also showed the presence of shortage gaming in the high margin settings. Furthermore, shortage gaming might as well originate in the random re-matching procedure that was used in previous studies. We used a repeated interaction among parties, which implies that our matching procedure and the resulting reciprocity concerns might as well explain the relatively constant ordering pattern as in Mengel and Peeters (2011). Finally, the last possible explanation could be that the uncertainty regarding the rationing mechanism can partially mitigate shortage gaming as suggested by Park et al. (2015). Nevertheless, this cannot be the comprehensive explanation based on their data and considering the fact that the warehouses made their rationing decisions to a large extent based on the retailers' order quantities. Thus, reversely, retailers should have anticipated the implied odds that this rationing practice will be in fact exhibited and they should have engaged in shortage gaming.

5.6. Conclusion

Our experiment takes a step towards external validity compared to Chapter 4 and assesses behavior in two-echelon supply chains with limited supplies. We provide several empirical insights. First of all, we document that there is no connection between whether a particular retailer gets fully serviced and how large the retailer is. People have neither the preference to satisfy at least one of the retailers fully nor to favor retailers based on their size. So, there seems to be no psychological cost for the warehouse in general if there is a mismatch between the products ordered and shipped. The prevalent rationing practice is the proportional actual order-based rationing rule with some minor unsystematic deviations and errors. Our finding is that individuals follow proportional rationing based on the retailers' actually placed orders under wholesale price contracts. Interestingly, the warehouses know that the demand distribution of the retailers is identical, but they still make their allocation decisions based on the actually observed orders. This implies that the ratio of order quantities is used as a strong anchor even when perfect information on the potential market outcomes is provided. The main takeaway for the retailers is that it would be a successful practice to overstate the number of products they need when the supplies are constrained because it will result in receiving a larger amount of supplies. Nevertheless, this might change if the warehouses suspect untruthful order patterns. The actual order-based rationing remains present to a limited extent even under revenue sharing. Even though this does not coincide with the self-interested behavioral pattern, approx. 30% still follows this particular allocation pattern. The vast majority follows the proportional allocation rule based on the demand distributions under revenue sharing, which coincides with the self-interested prediction. Nevertheless, a key managerial insight is that this leads to only slightly – statistically indistinguishably – larger supply chain profits. This implies that the revenue-sharing contract is a relatively effective coordination mechanism for distribution networks with limited supplies. It induces the optimal allocation mechanism, however, the profit impacts are minimal due to the fact that the limited supplies are expected to get sold even if they are allocated unevenly among retailers. Furthermore, due to the fact that the warehouse's revenue share was only 25%, the impact on the warehouse's profit is small if one of the retailer has some leftovers and the other one eventually runs out of stock. The positive and negative effects of uneven allocation are stronger for the retailers, but, due to the above mentioned reason, the gains on one side are almost equivalent to the losses for the other retailer. All in all, this translates to the fact that the retailers' profit is largely driven by the allocation mechanism, but there is only minimal impact on the entire supply chain.

Interestingly, we could not document the presence of fairness concerns. The warehouses' allocation patterns do not reflect such intentions at all. This draws the attention to the fact that fairness has a negligible role in inventory rationing. Taking one further step along this line, we ask why other researchers managed to document inequity aversion in other operations-related settings such as setting prices in serial supply chains. As our setting is different from theirs in many dimensions, we can only suggest that the key driving factor could be that our setting did not allow the retailers to reject the warehouse's offers. Moreover, there were intuitive anchors present in our setting which were rationing based on the actually ordered quantities or the demand distributions, whereas the other settings did not have any. In a nutshell, fairness concerns do not have a universal role and their presence is dependent on the nature of the setting.

Even though the warehouse relied largely on the retailers' actual orders, we did not see a general tendency from the retailers to behave untruthfully and they did not overstate their orders. As shortage gaming is a widely observed phenomenon in the community that is detrimental to supply chain coordination and efficiency, this finding is very promising. It suggests that particular supply chain design characteristics might be able to mitigate this behavioral regularity. Nevertheless, based on this experiment, the potential list of explanations is still manifold: human decision-makers, lying aversion, the knowledge on the other retailer's orders and the products it received, the stochastic nature of the retailers' demand, the uncertainty regarding the exhibited rationing mechanism or the reciprocity stemming from the repeated interactions between parties.

It is obvious from the current experimental design that we expected neither the lack of shortage gaming nor the absence of conscious fairness concerns. Therefore, our current experimental setup cannot answer some of the above mentioned crucial whys. For the moment we see multiple potential candidates that might explain the documented phenomena. In this spirit, we encourage further research which can define the exact circumstances that induce fairness concerns in operations-related contexts. Similarly, we are eager to find out which specific design choices can lead to the absence of shortage gaming. All of these aspects have further substantial implications for supply chain design.

Chapter 6

6. Behavioral Decisions in Inventory Systems with Dual Sourcing Option

6.1. Introductory Remarks

In this chapter, we investigate ordering behavior when multiple supply modes exist. Our main motivation is that several multibillion companies – such as Chiquita (Tomlin, 2006) and even the US government (Lyon, 2006) – use dual sourcing because of its cost saving potential (Klosterhalfen et al., 2014). A supply base usually consists of potential alternatives with different costs, lead times, etc. Purchasing professionals allocate orders between the available supply modes based on these characteristics. Individual preferences and potential behavioral explanations have received relatively little attention in such dual sourcing contexts, but they have recently managed to explain the observed behavior in other operations-related contexts (De Véricourt et al., 2013; Moritz et al., 2013).

For multi-period inventory control, whether there is only a single or there are multiple sourcing modes, inventory theory contains several analyses on optimal policy structures and optimal parameters for predetermined structures. Despite the fact that dual sourcing has become a common practice, there is little knowledge on behavioral policies in such settings and about the extent to which they influence supply chain costs. There is a consensus in theoretical research that dual sourcing can lead to considerable cost reductions compared to single-sourcing. This means that the normatively optimal cost in any dual sourcing setting is smaller or in the worst case at most equal to that of the single-sourcing alternative. However, behavioral evidence (Bansal and Moritz, 2015; Lee and Siemsen, 2016) suggests that subjects deviate more from the optimal solution if they encounter cognitively more complex settings. Consequently, the gap between the normatively optimal and the actually realized costs could be larger in the dual sourcing setting due to its more complex nature. Additional evidence suggests that the dual sourcing setting can be regarded as cognitively more complex than the single-sourcing one because of the number of decisions: Wood (1986) defines that tasks which have fewer different categories of information and which require less processing capacities are simpler.

Allon and Van Mieghem (2010b) define dual sourcing as ‘a surprisingly complex management problem with potentially high rewards’, which is caused by the tedious process of

computing its effects on the inventory. All this means that it is not clear ex ante whether the dual sourcing setting delivers the projected costs savings in practice because of its additional layer of complexity. Counterintuitively, we observe further cost savings in our experiment in addition to the ones stemming from the normative predictions.

Furthermore, allocating orders between two particular supply modes with different lead times can be seen as a real option in a broader context. Real options are opportunities for executives to change particular characteristics of a decision in response to the changing environment. In the operations context, these are the pipeline and the on-hand inventory, the differences in the lead times of supply modes or the varying demand. There is further evidence from the management literature that executives systematically overvalue real options due to particular concerns such as competition (Lankton and Luft, 2008). Bansal and Moritz (2015) showed that subjects overestimated the monetary value of the flexible manufacturing option and thus overused it – even without the presence of such a concern. It was possible to substitute a particular type of product to the other one in their production-related setting, but not vice versa. They claim that this phenomenon is robust and that it is caused by fundamental and systematic behavioral biases, but the main drivers still remain to be explored. In a broader context, our investigation provides a hint why real options are overused in supply chain (Kremer and Van Wassenhove, 2014) or other business contexts. We find that reference-dependent preferences (Kahneman and Tversky, 1979; Ho et al., 2010) – and the resulting psychological costs assigned to leftovers and shortages – are consistent with the actually exhibited overusage in our sample.

Finally, there are further crucial questions to be answered: Which of the policy candidates derived from the theoretical academic literature is followed under dual sourcing? Do people combine the slow and the fast supply mode or do they tend to have a preferred supply mode? How much randomness can be observed when one follows a particular policy? Are there systematic differences between groups of subjects that follow different policies? To what extent do the corresponding biases drive costs? We observe that subjects tend to choose a preferred supply mode which can differ across periods for the same subject. We show that subjects use order smoothing policies. Their orders are based on the incoming units in the current period and the demand in the previous one. We document that subjects do not take all of the products that are currently in transit into consideration and thus underweight the supply line. We find that their differences in cognitive ability explain the extent to which supply line underweighting,

order smoothing, overusage of the fast supply mode and random errors are present. Moreover, we provide evidence that there are cost differences between subjects with high and low cognitive ability, primarily due to the difference in biases. All this highlights the importance of having multiple supply modes and hiring procurement decision-makers with a high cognitive reflection.

6.2. Literature Review

The most relevant finding on dual sourcing is that it leads to cost reductions in various different normative settings. For a review on multiple sourcing, we advise the reader to refer to Minner (2003). Several competing dual sourcing policies have been derived. Most recently, this stream of literature has focused on comparing the potential performance of different policies under multiple sourcing. Veeraraghavan and Scheller-Wolf (2008) tested the dual-index sourcing policy and claimed that it costs only 1-2% more than the optimal policy obtained by stochastic dynamic programming (SDP), however, it leads to cost reductions of up to 50% compared to single-sourcing. Their results were validated in various environments including non-stationary demand, supply disruptions and demand spikes. Sheopuri et al. (2010) classified various dual-index policies based on whether they have an order up to structure to the regular or to the emergency supply mode. They derived policies that need similar computational requirements as the dual-index policies but performed slightly better. Allon and Van Mieghem (2010a) investigated the cognitively easy and intuitive tailored surge-base policy (Rosenshine and Obee, 1976) where a constant quantity is ordered from the slow and a variable quantity is ordered from the fast supply mode in each period. Their numerical study proved that this policy is near to optimal in terms of total cost. In addition, they derived the normatively optimal allocation to the slow and the fast supply modes. Janakiraman et al. (2015) showed that the cost gap between such surge-base policies and the optimal policy decreases dramatically when the lead time of the regular supplier is relatively large compared to the fast vendor's lead time. Boute and Van Mieghem (2015) included capacity costs in their model and showed that the tailored surge-base policy is a subtype of order smoothing policies under the assumption that the slow supply mode cannot change its production level. They defined the orders in the current period as the weighted average of incoming products in the current period and the demand in the previous one. Boute and Van Mieghem (2015) claimed that these smoothing policies perform similarly to tailored surge-base, single-index and dual-index dual sourcing policies under various demand sequences and even outperform them when capacity costs and lead times are high. They provide a short

review on the smoothing literature and point out that such policies can be observed in the real-world in the retail – and in some cases even in the manufacturing – industry (Cachon et al., 2007). Klosterhalfen et al. (2011) tested various dual sourcing policies in a single-product periodic review system with backorders. They confirmed that following a constant-order policy means a good trade-off in terms of simplicity and efficiency. Jain et al. (2014) used empirical data to investigate the impact of dual sourcing on inventory investments. They find that using dual sourcing practices reduces inventory investments by 11%.

Chapter 6 is also related to the literature on behavioral factors. Among others, it has been shown that ex post inventory error minimization (Ho et al., 2010; Kremer et al., 2014) and cognitive limitations can drive behavior in the newsvendor (Moritz et al., 2013) and more complicated interactive multi-period settings such as the beer distribution game (Narayanan and Moritz, 2015), forecasting tasks (Moritz et al., 2014) or inventory rationing (Cui and Zhang, 2016). The theory on the role of cognitive ability is summarized by Toplak et al. (2011).

Similarly, the random error model (Su, 2008) uses cognitive limitations as its cornerstone hypothesis. It has been tested by Kremer et al. (2010) in the newsvendor and also in a particular dual sourcing setting: Gurnani et al. (2014) investigated perishable items under supply risk and disruptions. They found that the overwhelming majority of their subjects used dual sourcing in cases when single sourcing was theoretically optimal. However, the general aversion to corner solutions and the presence of highly probable supply disruptions (Goldschmidt et al., 2014) could be the main driver of their results. Our repeated setting is different from theirs in the sense that we focus on non-perishable products where it is optimal to combine both supply modes in the majority of the cases. In addition, we take a more holistic approach: We analyze the behavioral choice between different ordering policies, the motives for the exact allocation between the slow and the fast supply modes and the difference between the actually realized and the normatively optimal costs of single- and dual sourcing treatments.

6.3. The Dual Sourcing Setting and Research Hypotheses

We focus on a single-echelon single decision-maker who has a fast, expensive and a slow, relatively inexpensive supply mode with different deterministic cost parameters and lead times in a repeated multi-period, non-perishable context. Leftovers (shortages) are carried over to the next period and on-hand inventory/backorders is/are penalized. Neither supply mode has capacity constraints or disruption risk.

We expect that the gap between the normatively achievable and the actually realized cost increases due to the more complex nature of the dual sourcing setting. In addition, we expect more supply line underweighting, random errors and order smoothing in the complex dual sourcing setting because of this additional layer of complexity.

Hypothesis H6.1: Dual sourcing leads to a larger difference between the normative and the actual costs due to more bias.

Additionally, we expect that individuals prefer the more flexible supply mode and consequently overuse it. We assume that this is consistent with ex post inventory error minimization. Furthermore, it can be explained by preferences with respect to the reference point of not having any shortages or leftovers which have additional psychological costs as in Ho et al. (2010). We investigate whether the ordering policies are consistent with particular preferences that managed to explain behavior in other inventory-related contexts – such as risk attitude (De Véricourt et al., 2013), loss aversion (Schweitzer and Cachon, 2000) or regret aversion. Furthermore, we assume that subjects are generally averse to extreme solutions such as ordering from only one of the suppliers and tend to combine their orders from the two supply modes as in Gurnani et al. (2014).

Hypothesis H6.2: The fast sourcing option is overused and subjects tend to combine the slow and fast supply modes.

We assume that cognitive ability explains the difference in the exhibited ordering pattern since it has been a useful predictor in the newsvendor setting (Moritz et al., 2013). We expect that subjects with high cognitive reflection are less biased. We assume that subjects with low cognitive reflection exhibit more variation in their ordering policy, which is in line with the random error model in Gurnani et al. (2014). Moreover, we expect that this group of subjects underweights the supply line to a larger extent than the people with higher cognitive reflection scores as in Narayanan and Moritz (2015). We assume that subjects with a lower cognitive score overuse the fast supply mode to a greater extent and have higher costs. Due to the relatively complex computational requirements of the dual sourcing policies, we expect that ordering policies are not followed according to their exact definition, but an order smoothing strategy is used as a heuristic instead, which is suggested by Boute and Van Mieghem (2015). We predict that individuals with high cognitive reflection use ordering policies that mimic the optimal single-index policy closer: They rely less on the incoming products in the current period as an

anchor and take the demand in the previous period with a larger share into consideration, whereas their peers with low cognitive reflection exhibit smoothing to a larger extent.

Hypothesis H6.3: Higher cognitive ability yields lower costs, less overusage, less order smoothing, less supply line underweighting and less order quantity randomness.

6.4. The Theoretical Models

6.4.1. The Normative Benchmark

Normative theory covers several competing inventory policies for dual sourcing settings. Scheller-Wolf et al. (2006) derived the single-index policy and the optimal order quantities for the two supply modes. This is optimal for lead times of 0 and 1 and we consider exactly this setting. The single-index policy tracks only the inventory position (*IP*) in every period. The parameters (z_S, z_F) define the regular and emergency order-up-to levels, which can be found with an iterative search procedure for subject *i* in each period *t*. We denote the demand by *D*. S_{it} and F_{it} are the respective quantities ordered from the slow and the fast option with their corresponding lead times (L_S=1 and L_F=0) and purchase prices (p_S and p_F).

$$F_{it} = MAX\{z_F - IP_{it}, 0\} = MAX\{D_{t-1} - (z_S + z_F), 0\} \tag{6.1}$$

$$S_{it} = MAX\{z_S - (IP_{it} + F_{it}), 0\} = MIN\{z_S - z_F, D_{t-1}\} = D_{t-1} - F_{it} \tag{6.2}$$

We consider 3 different treatments: D7T (*N*=19), D10T (*N*=24) and ST (*N*=22). We denote *IL* as the on-hand inventory level, *h* as the unit holding cost per period and *b* as the unit backorder cost per period. The parameters *h*=2, *b*=8 and p_S=6 were kept constant in all treatments. We determined the parameters of our experiment such that the normatively optimal usage proportion of the slow and fast modes resembled real-world practices and lead to an approx. 75%-25% split between the slow and the fast supplier in D7T as in Allon and Van Mieghem (2010a). In treatment D10T, we used p_F=10, which is the lowest integer value when it was never optimal to use the fast mode according to normative theory, whereas it proved to be optimal to use the fast mode under particular circumstances in D7T (p_F=7). Our setting will have a finite horizon of 48 periods, thus we report the optimal parameters (z_S;z_F) based on SDP in line with Whittemore and Saunders (1977) for each treatment:

D7T: Periods 1-44: (10;6), period 45: (9;6), period 46: (9;7), period 47: (-;7), period 48: (0;-).

ST and D10T: Periods 1-45: (11;-), period 46: (9;-), period 47: (6;-), period 48: (-;-).

The subject-specific total cost (C_i) over the planning horizon T is the sum of the holding (C_H), the backorder (C_B), the slow (C_S) and the fast supply mode's (C_F) costs:

$$C_i = \sum_{t=1}^{T}[h \cdot \text{MAX}\{IL_{it}, 0\} - b \cdot \text{MIN}\{IL_{it}, 0\}] + \sum_{t=1}^{T}[p_S \cdot S_{it} + p_F \cdot F_{it}] \qquad (6.3)$$

We call this approach Model 1. As a further step, we calculated the optimal decisions by taking only the actual inventory positions into account. We refer to this approach as Model 2. If our subjects were more often in situations where their inventory position was larger (smaller), this would mean that they should use the fast supply mode more (less) frequently according to Model 2.

6.4.2. Alternative Behavioral Theories

An alternative model by Ho et al. (2010) indicated that psychological costs assigned to shortages and leftovers are responsible for biases in newsvendor orders. Therefore, we modified our theoretical benchmark to take various levels of potential psychological costs into account in Model 3: We calculated the optimal policy and the usage ratios by adjusting the holding and backorder costs in equation (6.3) to $h'=h+h_{ps}$ and $b'=b+b_{ps}$ where h_{ps} and b_{ps} denote the additional psychological costs assigned to leftovers and shortages respectively in line with the estimations by Ho et al. (2010). Our dual sourcing environment was cognitively different from the newsvendor setting, which might imply that a better alignment could be achieved with different parameters. Based on our parametric analysis shown in Figures 8.4.9 and 8.4.10 in the Appendix 8.4.8, the ratio of the ordered products from the slow option divided by the total order quantity is relatively stable for a wide range of psychological costs and it increases/decreases significantly stepwise at particular thresholds. The same could be claimed about the corresponding order-up-to parameters.

Finally, we calculated the optimal ordering decisions taking only the actual situation and the upcoming period into account for each subject. This was necessary to see whether people could have ignored or misjudged the future which would mean that they have done only a myopic optimization as reported by other studies (Thaler et al., 1997). We denote this approach as Model 4. Note that the myopic approach according to this definition would take only the current period into consideration, which implies that it would never be optimal to order from a

supply mode with a lead time of 1 or larger. Therefore, we extend the assumption on myopic optimization by taking both the current period and the next one into account and call this approach Model 5.

6.5. The Experimental Design

We ran an experiment at the Vienna Center for Experimental Economics with 65 students who participated in 3 different treatments according to our between-subject design. Our students spent 1 hour and 55 minutes in the lab and earned €25.56 on average. Our sample sizes in each treatment are in line with other recent studies in behavioral operations management (Croson et al., 2014; Bansal and Moritz, 2015; Ovchinnikov et al., 2015). Subjects participated in an inventory task, a few preference elicitation tasks and a questionnaire. The experiment was programmed with z-Tree (Fischbacher, 2007). ORSEE (Greiner, 2015) was used for recruitment.

In the inventory task, our subjects were told that they owned a warehouse that supplied spare parts and that they had to make decisions on how many units they ordered in 48 consecutive periods. The subjects could order from two options. The orders placed to the fast option were expensive and they were delivered instantly, whereas the slow option was less expensive, but had a lead time of one period. In every period, the sequence of events was the following:

1. The subjects tracked information on the on-hand inventory/backorders and the outstanding order with the slow option.
2. The subjects placed their orders to both options before the demand was realized.
3. The spare parts that were ordered from the fast supply mode during this period as well as the spare parts that had been ordered from the slow supply mode during the last period arrived. The demanded units were shipped. If the demand was not partially or totally satisfied, excess demand was backlogged. The on-hand inventory and backorders were carried over to the next period.
4. Extensive feedback was provided after each period. The subjects saw the demand, their ordered quantities from both options, their inventory level, the corresponding costs in that specific period and their total cost until that particular period. The quantity that had been ordered from the slow option in the previous period was shown separately in the in transit box from the slow option. Multiple examples and control questions were used

to clarify the sequence of events and to indicate that the units in transit were still outstanding, but were going to arrive after the next ordering decision, yet before the demand realization.

Initially, the subjects had 6 units on stock and 4 items in transit from the slow supplier which corresponds to the optimal status. They were told that *D* was randomly determined between 0 and 8 in each period from a uniform distribution and that there is independence across different periods. We used the same uncorrelated (ρ=-0.01) sequence for all subjects to ensure that the different realizations of random numbers do not distort our results. Note that we consider non-perishable items where the excess inventory and backlogs are carried over across periods. In addition, our subjects ordered different quantities of products from the first period on. These are key differences to Chapters 4 and 5 which imply that our subjects had different inventory in the various stages of the pipeline from the second period on even if they had the same demand sequence.

We used a relative rewarding mechanism based on the total cost our subjects accumulated across the 48 periods in the inventory task in line with Croson et al. (2014). The subjects' payment *P* was determined by the following formula:

$$P_i = €15 \cdot \frac{\mathrm{MAX}_{\{j=1,\dots,n\}}\{C_j\} - C_i}{\mathrm{MAX}_{\{j=1,\dots,n\}}\{C_j\} - \mathrm{MIN}_{\{j=1,\dots,n\}}\{C_j\})} \tag{6.4}$$

We elicited the subjects' cognitive reflection (*CRT*) (Frederick, 2005) after the dual sourcing task. In addition, regret aversion was assessed based on a slightly modified mechanism by Raeva et al. (2010). Attitude towards risk (Drichoutis and Lusk, 2012) and losses (Gächter et al., 2010) were measured with multiple price lists based on the findings in Chapter 3. The preference elicitation procedures took place after the inventory task in order to avoid potential priming and framing effects. We avoided income effects, since we gave feedback on the earnings only after we had elicited the subjects' preferences. Each preference elicitation task was incentivized and had roughly the same stake as the inventory task. In the end, our subjects received an additional €5.00 for filling in a questionnaire which surveyed socioeconomic characteristics (age, gender, income, etc.). This flat show-up fee was announced only after the experiment had been finished. Similarly, there is evidence (Croson et al., 2014) that indirect effects which stem from strategic uncertainty about the other players' actions could play a role in inventory-related contexts. Our design rules out this issue because our tasks involved only

individual decisions. To make sure that subjects understood the nature of the dual sourcing tasks, we included several control questions. All subjects managed to provide correct answers to those. In addition to the inventory task, one of the preference elicitation tasks was chosen to be payoff-relevant as it is recommended by the random lottery incentive mechanism (Cubitt et al., 1998).

In line with best practices in statistics, we report the results of two-sided non-parametric tests when at least one of our samples was smaller than 30 independent observations and the results of two-sided parametric tests otherwise.

6.6. Main Results

6.6.1. Cost Comparisons

We calculated the total cost across 48 periods by equation (6.3) for each subject. Table 6.1 gives an overview on the average (standard deviation) of the realized and normative costs per treatment. We now examine the cost differences among treatments. The Mann-Whitney-Wilcoxon tests (MWW-tests) show that the total cost is significantly higher in ST than in D7T ($p=0.03$) and in D10T ($p<0.01$). Consequently, we see lower costs in the dual sourcing treatments. Moreover, there is no difference between the cost of D7T and D10T, which is an indicator that even a very expensive dual sourcing option can drive down costs effectively.

Table 6.1: Comparison of Normatively Achievable and Actually Realized Costs

	Cost	Total (C)	Holding (C_H)	Backorder (C_B)	Slow (C_S)	Fast (C_F)
ST	Norm.	1647.4	329.3	166.1	1152	-
	Actual	2009.1***	310.6	567.3***	1131.3	-
		(448.3)	(200.4)	(579.1)	(33.9)	-
D7T	Norm.	1598.9	317.8	75.8	831.8	373.4
	Actual	1700.4	333.1	222.7**	639.5***	505.1**
		(222.4)	(128.9)	(296.7)	(227.3)	(254.1)
D10T	Norm.	1647.4	329.3***	166.1	1152.0***	0.00***
	Actual	1751.9	435.0	157.7	975.0	184.2
		(138.4)	(146.8)	(146.5)	(130.0)	(208.7)

Notes: The table contains the averages (and the standard deviations) of the actually observed and the normative costs across 48 periods. Significance levels are given as follows: *: p=0.1, **: p=0.05, ***: p=0.01.

In theory, subjects should have the same cost as in the single-sourcing setting according to the normative predictions. The average total cost is 6.6% (6.4%) larger than the theoretically predicted one in D7T (in D10T), but it is not statistically different from it (Wilcoxon signed rank test (W-test), $p>0.31$) – whereas this difference is significant ($p<0.01$) in ST. This was mainly driven by the (W-test, $p=0.01$) backorder cost differences between the single-sourcing and dual sourcing treatments. Despite the fact that the dual sourcing setting is more complex, it leads to a lower gap to the normative predictions, lower standard deviations and better performance than in the single-sourcing setting.

Somewhat surprisingly, the more complex setting with an additional second sourcing option leads to cost reductions on top of the normatively achievable savings – contrarily to H6.1 – even if the second sourcing mode is relatively expensive. Table 6.2 provides evidence on further cost differences among others. In addition, it gives an overview on the main differences and similarities between single- and dual sourcing. We will discuss the meaning and calculation methodology of the various characteristics in detail later in each of the corresponding subsection. Generally, we see that the total cost and the gap between the normatively optimal solution and the actually observed one are significantly higher in ST than in the dual sourcing treatments (MWW-test, $p<0.01$). Interestingly, we observe large difference among the variances of costs: The Levene's test for the equality of variances (L-test) shows that the variance in ST is larger than in D7T or D10T (L-tests, $p<0.01$), whereas there are just modest differences among the treatments where both supply modes were present (L-test, $p\geq0.09$).

The huge standard deviations in ST indicate that there is a relatively large spread in the physical on-hand inventory levels in ST. The very low standard deviation of costs stemming from the slow mode in ST is intuitive, since this is the only option available in this treatment. Consequently, the total orders should be about the same for each subject. In general, we see a higher holding cost in D10T and higher backorder costs in ST and D7T, which can be explained by the fact that average inventory positions after ordering but before demand realization are somewhat – although not significantly – higher in D10T and D7T [ST:6.96 (3.39), D7T: 6.26 (2.21) and D10T: 7.84 (1.83)].

Table 6.2: Single-sourcing Versus Dual Sourcing

	Single-sourcing	Dual sourcing	Diff. significance
Supply line underweighting (γ)	0.39	0.73	0.01
	(0.41)	(0.40)	
Random error coefficient (β)	92.77	102.89	0.73
	(111.07)	(130.49)	
Gap to normatively optimal cost	361.73	133.38	0.01
	(448.34)	(183.01)	
Total cost (C)	2009.10	1729.14	0.01
	(448.35)	(223.62)	
Smoothing parameter (λ)	0.25	0.19	0.27
	(0.21)	(0.15)	

This table shows the averages (and standard deviations) of behavioral characteristics. The significance levels of statistical tests between single- and dual-sourcing are given as follows: *: p=0.1; **: p=0.05, ***: p=0.01.

In line with Narayanan and Moritz (2015), we classify our subjects in groups with high (*CRT*=2 and *CRT*=3, *N*=32) and low (*CRT*=0 and *CRT*=1, *N*=33) cognitive reflection. Table 6.3 summarizes all of our findings that are related to the documented differences in cognitive ability. In parallel to the differences between single- and dual sourcing, we elaborate on these findings later at the end of each respective subsection. For now, we document that individuals with higher cognitive reflection have lower total costs which confirms H6.3 (*CRT*=0 and *CRT*=1, *C*=1876.8 (370.3); *CRT*=2 and *CRT*=3, *C*=1719.8 (134.6); t-test, p=0.03). The averages (and the standard deviations) of the different cost types are (*CRT*=0 and *CRT*=1, C_H=382.4 (218.1), C_B=347.9 (424.8), C_S=863.2 (240.7), C_F=351.2 (258.5); *CRT*=2 and *CRT*=3, C_H=411.9 (145.2), C_B=209.7 (318.7), C_S=915.71 (213.8), C_F=293.9 (259.4)). Cognitive ability affects the differences in each component slightly, since we record only weakly significantly (t-test, p=0.08) larger backorder costs for the group with lower cognitive reflection, whereas the differences between the other components do not prove to be significant separately (a series of t-tests, p≥0.28). Note that the proportion of high-*CRT* subjects is comparable in each treatment and varies between 45.83% and 52.63%, ruling out that it drives our results.

Table 6.3: The Relationship between Cognitive Ability and the Actual Ordering Policy

	CRT = 0 or 1	*CRT* = 2 or 3	Diff. significance
Supply line underweighting (γ)	0.45 (0.42)	0.78 (0.32)	0.03
Random error coefficient (β)	120.30 (142.90)	59.10 (61.61)	0.04
Fast mode's usage proportion in relation to Model 3 predictions	5.71% (17.56%)	-3.20% (16.03%)	0.07
Total cost (C)	1876.78 (370.28)	1719.75 (134.62)	0.03
Coefficient of variation (CoV_S) Slow supply mode	0.81 (0.21)	0.71 (0.20)	0.02
Coefficient of variation (CoV_F) Fast supply mode	2.00 (1.20)	2.77 (1.81)	0.04
Smoothing parameter (λ)	0.25 (0.19)	0.16 (0.14)	0.05
Correlation between orders to the slow (S_{it}) and fast mode (F_{it})	0.29 (0.22)	0.24 (0.26)	0.45

Notes: This table summarizes the main differences across groups with heterogenous cognitive ability (*CRT*) with the average values and their standard deviations in parentheses. The fast mode's usage proportion is displayed using Model 3 as the reference point since, due to the different unit cost of the fast mode, a different proportion is predicted for treatments D7T and D10T.

6.6.2. Supplier Usage

Table 6.4 contains the averages (standard deviations) of the actually observed order quantities from the slow and the fast supply mode for each period and treatment. In addition, it shows the actually observed usage proportions over 48 periods. The respective theoretical predictions are also displayed, which indicate that Model 3 provides the best fit to our observations. Note that the slow and the fast orders add up to 3.74-3.77 in our treatments, which is due to the initial on-hand (6) and pipeline (4) inventory in our system. If we take these into account as orders (or if we truncate the initial 10 and last 5 periods), the treatment averages would be 3.95-3.98, which is the same as the average demand per period (3.97).

Individuals with high cognitive reflection used the fast supply mode 3.20% less often than the aggregate prediction by Model 3 would suggest, whereas their less able peers ordered 5.71% more frequently from it (t-test, p=0.07). This supports the overusage assumptions of

H6.3. Notice that it would be reasonable to assume that high-*CRT* subjects are less biased in their overusage pattern based on the aforementioned complexity argument. Taking this assumption into consideration, we get conventionally significant results with a one sided t-test (p=0.04).

Table 6.4: Actual and Model-predicted Supplier Usage

	Actual orders			Ordering policy characteristics					
	Slow	Fast		Actual	M1	M2	M3	M4	M5
ST	3.75	-	Slow ratio	100.00 (0.00)	100.00	100.00	100.00	100.00	100.00
	(0.11)	-	Inv. pos.	6.90 (3.70)	11	11	8	Never order	5
D7T	2.21	1.53	Slow ratio	59.37 (20.72)	72.22	55.80	58.33	0.00	0.00
	(0.77)	(0.78)	Inv. pos.	7.02 (2.33)	10	10	6	0	5
D10T	3.37	0.40	Slow ratio	89.82 (11.65)	100.00	92.72	91.67	0.00	51.75
	(0.44)	(0.44)	Inv. pos.	8.04 (1.86)	11	11	8	Never order	5

Notes: This table provides data on the averages (and the standard deviations) of the order quantities, the slow supply mode's average usage ratio and the inventory positions. M1 indicates the normatively optimal values according to the original SDP model. M2 takes the reoptimization approach into account. M3 denotes figures in line with the corresponding additional psychological costs. M4 (M5) is the myopic approach that takes only the current (and the next) period into account. The columns under Actual contain the actually observed values.

A further interesting observation is that the normative prediction signals a positive correlation of 0.54-0.59 between the slow and the fast supply mode, depending on the particular dual sourcing treatment. However, our data shows that the mean correlation for our subjects was 0.27 with a standard deviation of 0.24, which is significantly different (t-test, $p<0.01$) from the predictions. We documented even negative correlations for a number of subjects. The main reason why the normative model predicts high correlations is the following: In case of a large demand, a large fast order and a maximum slow order would raise inventory position to a particular level. Thus, in optimal dual sourcing, subjects should combine both modes. However, the majority uses – depending on the situation – either the slow or the fast mode, with some error that implies a small order with the other option. This behavioral pattern is very strong in general, since both the high- and low-*CRT* individuals choose a preferred supplier in each

period: The correlation between the fast and the slow mode is identical in these two groups (MWW-test, p=0.45). Therefore, the subjects seem to relax complexity by choosing a preferred mode. The preferred supply mode can of course vary across periods.

6.6.3. Robustness Checks

We truncate the first ten and the last five periods for all previous analyses and verify that our findings remain the same. Additionally, our detailed investigation (Appendix 8.4.9) rules out potential learning, endgame and autocorrelation effects.

We checked whether our subjects decided based on the physical inventory level or the inventory position: Appendix 8.4.9 elaborates on the validity of the latter assumption. Furthermore, it is not necessarily sufficient to determine the model fit that is based solely on the aggregate inventory position per treatment because the congruence between the individual period-specific decisions and the averages predicted by the various models might not coincide. The estimated average usage proportion can have a good aggregate fit even if the subjects deviate to a considerable extent in both directions across all the periods, which cancels out in the average measures in the end. Appendix 8.4.9 shows that the period-specific fit (Hansen et al., 2011) of Model 3 is superior to the others based on the objective usage ratio and inventory position figures. Thus, H6.2 is sustained, people seem to attach psychological costs to shortages and leftovers in general. This finding can be reconciled with our observations in Chapter 5, since only the lack of additional psychological costs was documented in that section, which stem from not satisfying a particular client.

We conducted robustness checks that provide further support for Model 3. We calculated the normatively optimal period-specific usage ratio for each subject. In addition, we computed the difference between this ratio and the actually observed one for each individual. We report that there is no significant correlation between these figures and the elicited risk attitude, regret and loss aversion measures.

The coefficients of variation in Table 6.5 also support the validity of Model 3, especially in the dual sourcing treatments. For 3 out of the 5 cases, the *CoV*s indicated that the predictions based on the normatively optimal ordering policy are the most accurate estimates after taking the psychological costs (Model 3) into account. The superiority of Model 3 is further supported by a series of W-tests which show that the differences between Model 3 (Model 1; Model 2;

Model 4; Model 5) and the documented behavior are significant at the 5% level in 1 (2;3;5;5) out of 5 cases.

Table 6.5: Optimal and Actual Coefficients of Variations of Order Quantities

	Slow option						Fast option					
	Act.	M1	M2	M3	M4	M5	Act.	M1	M2	M3	M4	M5
ST	0.67 (0.19)	0.68	0.79	0.84	0.00	1.39	-	-	-	-	-	-
D7T	0.91 (0.44)	0.53	0.48	0.60	0.00	0.00	1.47 (0.63)	1.38	1.16	1.48	0.68	1.14
D10T	0.65 (0.16)	0.68	0.72	0.84	0.00	0.00	3.45 (1.80)	0.00	2.12	3.69	0.68	1.14

Notes: This table contains the averages (and the standard deviations) of coefficients of variation. Values under Act. mean the actually observed values in the respective treatments. M1 indicates the normatively optimal values according to the original SDP model. M2 takes the reoptimization approach into account. M3 denotes the figures in line with the corresponding additional psychological costs. M4 (M5) is the myopic approach taking only the current (and the next) period into account.

We now turn our attention to the ordering policies to find out which biases are responsible for the order quantities we actually observed. Previously documented behavioral biases managed to explain behavior in single-sourcing settings. In order to shed light on the behavior in dual sourcing settings, we were interested in whether these biases are present here as well.

6.7. The Ordering Policy

6.7.1. Supply Line Underweighting

First, notice that, by definition, the physical inventory level-based and the inventory position-based approaches represent extremes because the subjects might take only a proportion of their pipeline inventory into account. Therefore, we seek evidence for supply line underweighting. We maximize the congruence between the actual decisions and the theoretical model that quantifies supply line underweighting (Croson et al., 2014). We use the following formula based on previous studies for each period and each subject:

$$S_{it} + F_{it} = MAX\{0, CO_{it} + \alpha_i(S'_i - IL_{it} - \gamma_i S_{t-1}) + \varepsilon_{it}\} \quad (6.5)$$

where α_i is the proportion of the inventory surplus (or shortage) ordered per subject, γ_i is the fraction of the supply line considered by the subject, S'_i denotes the desired inventory position

per subject and CO_{it} assesses the expected incoming orders per period per subject. This latter one can be quantified with δ and inserted in equation (6.5) afterwards as in previous studies (Croson et al., 2014; Narayanan and Moritz, 2015):

$$CO_{it} = \delta_i(S_{it-1} + F_{it}) + (1 - \delta_i)CO_{it-1} \tag{6.6}$$

We use least squares to derive the estimates of the *α, γ, δ and S'* parameters subject to the constraints $0 \leq \alpha \leq 1$, $0 \leq \gamma \leq 1$, $\delta \geq 0$ and $S' \geq 0$. The observed γ values confirm that supply line underweighting is present in our sample as γ is significantly lower than the standard model's predictions (γ=1), but considerably larger than in Croson et al. (2014) or Narayanan and Moritz (2015). This means that a slightly larger fraction of the supply line is considered by our subjects, which is intuitive since our setting has a lead time of 1 period, whereas the lead time is 4 in the aforementioned settings. We confirm that supply line underweighting increases with lead time, since it (γ=0.73 (0.40)) is smaller (MWW-test, p<0.01) in treatments where the fast sourcing option was available compared to ST (γ=0.39 (0.41)). It seems plausible that the existence of the second supply mode, thus an additional option with lower lead time, reduces the degree of uncertainty and increases the planning accuracy. This explains why we observed large standard deviations in the inventory positions and costs – especially backorder costs – in ST compared to the other treatments: Due to increased supply line underweighting, the lack of the flexible supply mode causes more mismatch between supply and demand – especially for low-*CRT* subjects as γ correlated (ρ=0.37, p=0.01) with *CRT*. The γ values in the two cognitive groups (*CRT*=2 or 3: γ=0.78 (0.32); *CRT*=0 or 1: γ=0.45 (0.42)) differ significantly (t-test, p<0.01). We note that the other parameters we estimate with equation (6.5) (α=0.36 (0.26), δ=0.39 (0.41), MSE=5.01, R^2=0.69), are in line with previous studies (Croson et al., 2014; Narayanan and Moritz, 2015). The parameter which quantifies the desired on-hand and on-order inventory (*S'*=7.87 (4.79)) is not identical, which makes sense considering that our setting was different from their setting.

Interestingly, we also find that subjects take only some of the units in transit into consideration, however, they are aware that those units will arrive in the current period before the demand is realized. One explanation for this is consistent with the psychological cost model: These products are not yet in the possession of the company, even though they can be used to satisfy the current period's demand. Notice that our finding on supply line underweighting does not contradict our claim that subjects based their decision on their inventory position in general, because the γ parameters show that a relatively large fraction of the supply line was still taken

into consideration. This explains why the model that relies on inventory position provides a better fit than the physical inventory based one.

6.7.2. Order Smoothing

After we found evidence for underweighting, we investigated the extent of anchoring based on the previous period's orders. Boute and Van Mieghem (2015) introduced such a smoothing policy to single-sourcing:

$$S_{it} = \lambda_i S_{it-1} + (1 - \lambda_i) D_{t-1} \tag{6.7}$$

where λ is the subject-specific smoothing parameter estimated across all periods. Notice that the optimal parameters for the tailored surge-base policy under dual sourcing (Allon and Van Mieghem, 2010a) – where a constant amount is ordered from the slow and a variable amount or zero units are ordered from the fast supply mode – would be equivalent to setting λ_i=1. This means a constant order quantity from the slow supply mode. Similarly, an index-driven policy would imply λ_i=0. Boute and Van Mieghem (2015) consider both supply modes in the same smoothing formula for dual sourcing smoothing policies:

$$S_{it-1} + F_{it} = \lambda_i (S_{it-2} + F_{it-1}) + (1 - \lambda_i) D_{t-1} \tag{6.8}$$

We estimate order smoothing parameters (λ=0.21 (0.17)) by adding an error term and by minimizing the mean squared errors for each subject across the 48 periods. This model achieves a better fit than the one that considered only inventory position (MSE=3.79, R^2=0.77). Our results are significantly different (t-test, p<0.01) from the normative prediction (λ=0), which shows that our subjects use their previous orders as anchors to some extent, but consider the previous demand with a larger weight. This results in an ordering policy that is close to the optimal single-index policy. The order smoothing strategy performs relatively well in terms of costs, confirming the benevolent findings of the simulation results in Boute and Van Mieghem (2015) in practice. Notably, we document that order smoothing is identical (MWW-test, p=0.27) in our dual sourcing (λ=0.19 (0.15)) and single-sourcing treatments (λ=0.25 (0.21)). Our findings on order smoothing do not contradict our previous claims on the lack of learning and autocorrelation, since the estimated λ parameters are relatively low. Furthermore, they are estimated across 48 periods per subject, implying that there is considerable variation for the same subject across periods. In addition, we observe that individuals with high cognitive reflection put less weight (t-test, p=0.05) on their previous orders (λ=0.16 (0.14)) than their

peers who have a low *CRT* score (λ=0.25 (0.19)). Therefore, the orders of our high-*CRT* subjects followed the demand pattern and the optimal single-index policy more closely.

We confirm the presence of order smoothing since the actually observed inventory positions in the Appendix 8.4.9 are not different from the theoretical predictions, which was 6.11 (2.62) (MWW-tests, p=0.45). Furthermore, we draw the attention to the fact that following a clearly defined surge-base policy would imply that CoV_S=0 and CoV_F=1.38 in D7T and CoV_F=1.16 in D10T. Generally, we observed larger coefficients of variation in Table 6.5, but this discrepancy can be explained. Order smoothing predicts a higher CoV_S value, namely a figure between 0 and the theoretical predictions by Model 3 in each treatment. Similarly, this approach predicts a somewhat higher CoV_F compared to Model 3. If we assume the same magnitude of smoothing across all periods, our smoothing estimates (λ=0.21) will also imply higher theoretical predictions: CoV_S=0.55 and CoV_F=2.22. These values are calculated following the formula provided in Boute and Van Mieghem (2015):

$$CoV_{iF} = \frac{\sqrt{\frac{1-\lambda_i}{1+\lambda_i} \cdot \left[1-{\lambda_i}^{2 \cdot (L_S - L_F)}\right] \cdot Var(D)}}{\left[1-{\lambda_i}^{(L_S - L_F)}\right] \cdot E(D)} \tag{6.9}$$

$$CoV_{iS} = \frac{\sqrt{\frac{1-\lambda_i}{1+\lambda_i} \cdot {\lambda_i}^{2 \cdot (L_S - L_F)} \cdot Var(D)}}{{\lambda_i}^{(L_S - L_F)} \cdot E(D)} \tag{6.10}$$

Generally, the *CoV* estimates based on order smoothing are considerably closer to our actual observations than the predictions by Model 3. Yet the actual CoV_S (CoV_F) values are still slightly lower (higher) than the predictions. We turn our attention to the random error model to provide an explanation for this phenomenon.

6.7.3. The Random Error Model

It seems plausible that subjects with high cognitive reflection have more consistent inventory positions before the demand realization than their peers with low *CRT* scores. This would be in line with previous evidence on the random error model (Su, 2008; Kremer et al., 2010). In a nutshell, this model assumes that the inventory position before demand realization and thus the order quantities are random variables depending on one's cognitive limitations. Gurnani et al. (2014) provide the theoretical derivation of the random error model for their dual sourcing experiment. We apply their approach and adjust it to our setting, since we consider costs instead

of profits for each subject across T=48 periods. We can calculate each subject's individual bounded rationality parameter β_i separately by maximizing the following joint log-likelihood formula:

$$LL_i(\beta_i) = \sum_{t=1}^{T} \frac{C_{it}}{\beta_i} - T \cdot log \prod_{t=1}^{T} \left(\sum_{F_{it}=0}^{F_{imax}} \sum_{S_{it}}^{S_{imax}} e^{C_{it}/\beta_i} \right) \quad (6.11)$$

C_{it} is the cost of a particular combination of S_{it} and F_{it}. F_{imax} and S_{imax} are the largest orders from the slow and the fast supply modes per subject, which are determined based on our observations. As opposed to the setup of Gurnani et al. (2014), who investigate perishable items, our leftovers and shortages are carried over across periods. This means that $\sum_{F_{it}=0}^{F_{imax}} \sum_{S_{it}}^{S_{imax}} e^{C_{it}/\beta_i}$ does not change on a periodical basis in their setting. Instead, we must assess it in each period depending on the exact inventory position individually. By doing so, we take a conservative approach and assume that our subjects take an optimal decision in the upcoming periods after the period that is relevant for our calculations. This implies that we take only the direct cost difference that is applicable to the particular period into consideration and we neglect potential indirect effects. Note that our estimated bounded rationality parameters would increase if we assumed the same degree of suboptimal behavior in the subsequent periods. For the treatment without the fast sourcing mode, we used the following formula instead of equation (6.11):

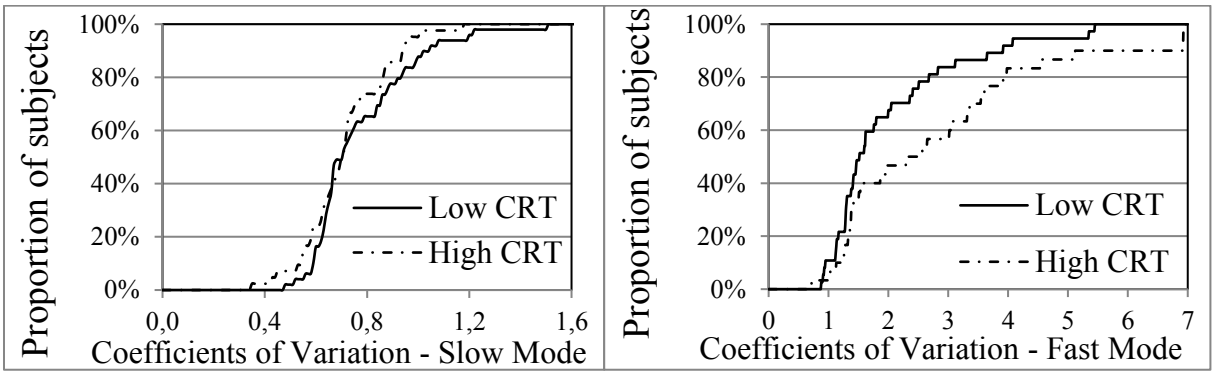
$$LL_i(\beta_i) = \sum_{t=1}^{T} \frac{C_{it}}{\beta_i} - T \cdot log \prod_{t=1}^{T} \left(\sum_{S_{it}}^{S_{imax}} e^{C_{it}/\beta_i} \right) \quad (6.12)$$

The average (and the standard deviation) of our empirical bounded rationality parameters are comparable in the dual sourcing treatments (102.89 (130.49)) to ST (92.77 (111.07)), implying no significant differences (MWW-test, p=0.73) in bounded rationality among our treatments. The values listed above are consistent with the estimates – between 75 and 379 depending on the specific treatment – by Gurnani et al. (2014). The normative decisions without the behavioral model predict that β=0 (or at least reasonably close to it). We note that taking bounded rationality into account gives a better fit, since LL(β_i)=-74.77, whereas LL(β=1)=-1125.59. Moreover, we use the log-likelihood ratio test χ^2= 2·[$LL_i(\beta_i)$ – $LL_i(\beta$=1)] for each subject to test whether the estimated bounded rationality parameters are significant. All values exceed χ^2(p=0.01)=6.635 by far, which means that every subject makes random errors. We discover a significant negative correlation (D7T and D10T: ρ=-0.36, p=0.01; ST: ρ=-0.24, p=0.02) between the empirical β parameters and the *CRT* score. Thus, the extent of the

documented random errors has a connection to cognitive ability. Moreover, the t-test also shows significant differences (p=0.04) between individuals with high (β=59.10 (61.61)) and low (β=120.30 (142.90)) *CRT* scores in accordance with H6.3.

A further piece of evidence for the systematic difference in the policies is that the *CRT* scores correlate with the coefficients of variation of the slow (fast) supply mode negatively (positively) (ρ=-0.24, p=0.01; ρ=0.30, p=0.01). Consequently, we see higher *CoVs* for subjects who had more random errors, which is intuitive considering the larger spread in their orders. Surprisingly, the effect of bounded rationality on the CoV_F is the opposite since this supply mode should be triggered only in emergency cases. As opposed to the slow mode where a positive order should be placed in almost all the periods, the optimal policy and Model 3 predict that 0 units should be ordered from the fast mode in a fairly large proportion of periods. In all these cases, a boundedly rational subject can only make a mistake by ordering 1, 2 etc., which compresses the CoV_F in the end. The same logic applies to supply line underweighting which, due to the more heterogeneous inventory positions and physical inventory level is expected to increase (decrease) the *CoV* for the slow (fast) mode. In line with this prediction, we observe that the *CoV* is significantly higher (lower) for the slow (fast) (t-test, p=0.02 and p=0.04) mode in the case of high-*CRT* subjects: Figures 6.1a and 6.1b depict the differences of the coefficients of variation for the slow and the fast option respectively.

Figure 6.1a and 6.1b: Coefficients of Variation Across Different Groups



Notes: These figures shows the cumulative distributions of the slow and the fast supply mode's coefficients of variation conditional on having a high (*CRT*=2 or *CRT*=3) or a low (*CRT*=0 or *CRT*=1) score on the cognitive reflection test.

Note that it is not possible to define an exact model that quantifies the effect sizes and the relationship between the aforementioned behavioral parameters and the *CoV*s. The main reason is that a specific estimate of β, γ or λ would not necessarily correspond to particular *CoV*s by itself. The reason for this is that the estimation of these parameters is based on multiple periods.

Thus, a large estimate can be a result of small deviations from the normative model in numerous periods, but it can also be driven by large deviations that were mere artifacts in just a few periods.

6.7.4. Joint Effects

One might argue that supply line underweighting, random errors and order smoothing are competing behavioral explanations, or at least that their effects could somewhat overlap or cancel out each other. For example, supply line underweighting could lead to similar fluctuation in the physical inventory levels and the inventory positions as bounded rationality does. Alternatively, smoothing strategies that are based on the previous period's events and supply line underweighting might cause the same observed outcomes. Therefore, we investigate whether the factors are still significant if we consider the aforementioned behavioral explanations simultaneously. We calculate the underweighting and smoothing coefficients jointly by minimizing the mean squared error terms in the following formula for each subject across all periods:

$$\lambda_i(S_{it-1} + F_{it}) + (1 - \lambda_i)D_{t-1} = MAX\{0, CO_{it} + \alpha_i(S'_i - IL_{it} - \gamma_i S_{t-1}) + \varepsilon_{it}\} \quad (6.13)$$

We note that this specification yields that a larger fraction of the supply line is considered in general. However, underweighting is still present (γ=0.67 (0.40)) and γ is significantly different (t-test, p<0.01) from 1. Moreover, the smoothing coefficients (λ=0.49 (0.25)) are greater than our previous estimates, which implies stronger and highly significant (t-test, p<0.01) presence of order smoothing strategies. Taking both factors into consideration improves the fit of our model as we expected (MSE=3.37, R^2=0.80). Next, we predict the optimal order quantities for each subject and for each period. Then we calculate the corresponding expected cost levels in order to obtain the random error coefficients in line with equations (6.11) and (6.12). This estimate tells us whether random errors are still present if we take underweighting and order smoothing into account. The bounded rationality parameters estimated under this specification are somewhat lower than our previous findings (β=69.24 (124.04)), but still relatively large. We estimate that including bounded rationality in the model yields a better fit ($LL(\beta_i)$=-74.85, $LL(\beta=1)$=-1100.59), which means that the presence of bounded rationality is highly significant ($\chi^2=2\cdot[LL_i(\beta_i) - LL_i(\beta=1)]$, p<0.01) for each subject. The previously documented differences between the cognitive groups (t-tests, p<0.01) do not change with this joint estimation.

6.8. Conclusion

Chapter 6 provides new insights into the behavior of single-echelon dual sourcing settings and applies to companies which already use or consider introducing dual sourcing practices. Most importantly, we saw that adding a faster supply mode to the supply network leads to significant cost reductions compared to single-sourcing. This effect stems partly from the expected cost savings that normative theory predicts for the dual sourcing treatments and thus they are driven by the availability of the fast supply mode. More importantly, these savings originate in fewer mismatches between supply and demand, thanks to the additional supply mode with lower lead times and, as a consequence, less supply line underweighting. The behavioral value of this setting is much larger than the higher unit sourcing cost of the fast supply mode. Our results are robust even if the fast supply mode is 66.67% more expensive than the slow one. Intuitively, we see further cost improvements, due to the normative predictions, in treatments where the fast supply mode was less expensive. Nevertheless, there is no further reduction in the gap between the normatively achievable and the actually observed costs. This highlights the importance of having multiple supply sources with different cost and lead time structures in the supply base because, counterintuitively, decision-makers seem to perform better in such complex settings – even if they have a preferred supply mode. Therefore, if the extra cost (e.g. doing research, assessing suppliers, getting quotes, conducting the negotiation procedure, etc.) of enlarging the supply base by adding suppliers with shorter lead times is relatively small, the savings are expected to offset these initial expenditures even in the short run.

We observe that subjects prefer overusing the fast and more expensive supply mode in dual sourcing settings. This overusage is in line with the ex post inventory error minimizing logic, which attaches additional psychological costs to shortages and leftovers. Parametric analyses and robustness checks support the validity of this model: The optimal order quantities and usage ratios remain the same between relatively broad threshold values. The fast supply mode could be considered as a real option that allows the postponement of order quantity decisions by one period, and as a matter of fact it reduces the uncertainty. This leaves more room for achieving a better match. Therefore, Chapter 6 could be regarded as a contribution to the potential reasons why real options are overused.

We see that, in addition to underweighting, the subjects use incoming orders as the anchor for their order placement decisions and they also make random errors. In general, the

extent of the documented biases correlates with the individuals' cognitive reflection, implying that individuals with high cognitive reflection take a larger fraction of units in transit into account. In addition, their inventory replenishment strategy is closer to the optimal single-index policy that has less random errors. In the end, these behavioral factors are the main drivers for the lower total costs. This highlights the importance of having procurement specialists with high cognitive reflection and provides far-reaching implications for hiring decisions similarly to the high margin setting in Chapter 4. In general, our subjects choose a preferred supply mode and they order the majority of the units from the selected supplier irrespective of their cognitive ability. Meanwhile, instead of relying on the optimal combination strategy, they place no (or very low) orders to the other supply mode. This suboptimal strategy still yields considerable improvements, mostly due to less supply line underweighting when compared to single-sourcing.

There are some limitations to our results. The experiment in Chapter 6 was conducted without any training, communication and decision support systems. These institutional frameworks managed to improve performance in other operations-related contexts such as the newsvendor setting (Bolton and Katok, 2008). Thus, it is not clear how large the actual differences are under such circumstances. We encourage further research to explore this avenue.

Chapter 7

7. General Implications

This thesis provides arguments that support the importance of the human factor in operations-related decisions. It documents that human decisions are in most cases far from the normatively optimal outcome. Due to the resulting inefficiency, this implies a considerable profit loss or cost increase in the investigated operations-related contexts. The ultimate focus of this thesis was to shed light on the underlying reasons and preferences in order to provide managerial insights and to identify improvement opportunities.

We conjecture that inefficiency can at least partially be mitigated by effective supply chain and mechanism design. The dual sourcing context provides a key insight here showing that people are not good at backwards induction. Consequently, they cannot calculate the appropriate inventory position and the corresponding order quantities in advance and they do not take products into account which are in transit. The cost of mismatch in the supply line is so high that providing a supply mode with lower lead time results in a large improvement – even if this fast vendor is extremely expensive. The newsvendor experiments deliver yet another key takeaway regarding the ordering mechanism: Most importantly, it is not sufficient to tell people what the optimal strategy is in a stochastic environment – especially if they receive outcome-based feedback on a periodical basis. The main reason for this discovery is that people prefer thinking in patterns and project their false perception because they cannot grasp the nature of independent probabilistic events well. An additional suggestive factor is the outcome-based ex post rationing practice that the warehouses use in their rationing practices in the presence of wholesale price contracts. Nevertheless, this inefficient practice is mitigated to some extent when revenue sharing incentives are provided for the warehouse. Similarly, it seems that the retailers' self-interested shortage gaming practice – which is detrimental to several distribution networks – can also be mitigated, however, it is not yet clear under which exact circumstances.

The other main takeaway of this thesis is regarding the role of preferences and other personal characteristics. We identify several factors as real drivers of the aforementioned ordering decisions. The experiment on the newsvendor setting helps to classify and quantify the strength of these characteristics. It also pinpoints where HR questionnaires, potential de-biasing

efforts and development activities should focus. In addition to identifying biased probability judgment as the root cause of the fluctuation in newsvendor orders across periods, we document that the impact of overconfidence, loss aversion, extraversion and cognitive ability is quite pronounced. Consequently, the thesis shows similar evidence for dual sourcing. People's cognitive ability seems to be a non-negligible factor that determines the extent of biases we observed and it influences the corresponding performance. We have to note that we cannot document any connection to regret or risk aversion in our dual sourcing settings. Further along this line, the thesis provides evidence that the way preferences are measured is not indifferent. The exact (risk and other) preference elicitation procedures might be crucial influential factors for the results that document correlations, especially since the consistency of these measurement methodologies is questionable. We therefore draw the attention to replications of operations-related studies that focus on the presence of particular preferences. Furthermore, the consistency and stability of other preference elicitation mechanisms should also be investigated.

All these aspects mean that we cannot present the philosopher's stone here, thus there is no general recipe for which exact traits or mechanism design elements are relevant and which are not. It depends on the nature of the setting, the cognitive environment, the context of the decision, the rewarding mechanism and even the framing of the task at hand among others. Thus, we are excited to see the forthcoming results in the field of behavioral operations management and hope that our work contributed to the general knowledge about this field.

Chapter 8

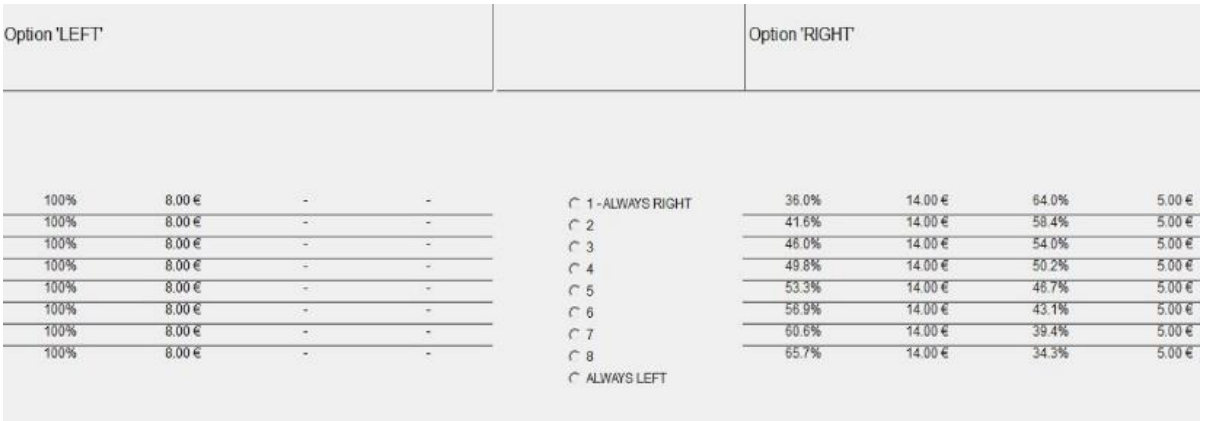
8. Appendix

8.1. How Sensitive Are (Risk) Preferences to Elicitation Mechanisms?

8.1.1. Instructions

Welcome to this experiment in decision making. You will be asked to make a series of choices that will affect your payment at the end of the experiment. Please pay close attention to the instructions, and if you have any questions raise your hand and an employee of the lab will help you with any questions you might have. Also, you will be asked to fill out a short questionnaire, where your answers are not relevant for your payoff. All your decisions and answers during the experiment will stay completely anonymous to everyone. In ’Part 1’, you will make decisions in twelve different situations, where in each situation you will make one choice. A sample decision screen is provided above and on the piece of paper in front of you. Please feel free to take notes on this paper.

Figure 8.1.1: Sample Screen



On each screen you see two columns (’LEFT’ and ’RIGHT’). In each column you have eight rows with different payment possibilities. You have to decide which option you prefer in each row. For example, on the screen provided above, in the first row you will have to choose between:

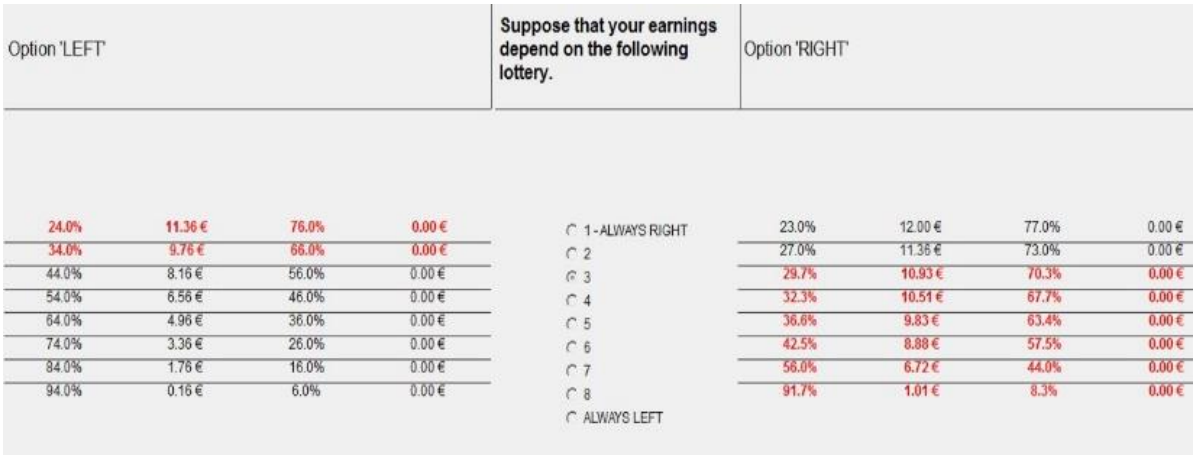
’LEFT’: receive 8€ with 100% probability
’RIGHT’: receive 14€ with 36.0% probability or 5€ with 64.0% probability
In the second row, you would make a choice between:

’LEFT’: receive 8€ with 100% probability
’RIGHT’: receive 14€ with 41.6% probability or 5€ with 58.4% probability

In the center of the screen you will find a number of radio buttons. You can only click one of those buttons on each screen. This button indicates at what row you want to switch from ’LEFT’ to ’RIGHT’.

If you choose row 3, that means you prefer ’LEFT’ for the first 2 rows, but you prefer ’RIGHT’ for rows 3-8.
If you choose row 6, that means you prefer ’LEFT’ for the first 5 rows, but you prefer ’RIGHT’ for rows 6-8.
If you choose row 1 - ALWAYS RIGHT, that means you prefer ’RIGHT’ in every row.
If you choose row ALWAYS LEFT, that means you prefer ’LEFT’ in every row.

Figure 8.1.2 Sample Screen with Color-coding



When you are finished making your decision for a screen, click ’OK’ and you will get to the next screen where you will see your choice again. In case you are not satisfied with your choice, you can change your choice once if you wish to. Your second choice is final and cannot be changed afterwards. After each decision screen, we will ask you how difficult it was for you to make a decision on the previous screen. These questions do not affect your payment. Still, we ask you to answer truthfully.

For your payoff of Part 1, one screen of the twelve is randomly selected by the computer. The computer will also select one of the rows at random. In this row, you have chosen ’LEFT’ or ’RIGHT’. Finally, the computer will randomize between the two possible outcomes based on

the given probabilities. For example, the computer chooses at random the following row to be relevant for your payoff:

'LEFT': 'receive 20€ with 70% probability or 5€ with 30% probability.'
'RIGHT': 'receive 15€ with 60% probability or 10€ with 40% probability.' You have chosen left in that particular row. Therefore, you either get 20€ or 5€, but the probability to get 20€ is higher.

Every screen and every row on each screen has an equal chance to be chosen by the computer to be relevant for your payoff. Considering that every decision you make matters, we advise you to think carefully about each decision you make. After you made these twelve decisions, 'Part 2' will start. Instructions for 'Part 2' will be given on the screens themselves. One situation in 'Part 2' will be randomly selected to affect your payment. 'Part 1' and 'Part 2' are completely independent of each other.

Your final payoff will then be the sum of your payoffs from 'Part 1' and 'Part 2'. If you have no questions, please click 'OK' to answer a couple of control questions. These questions will make sure that you have understood the setup. You cannot commence with the experiment unless you answer the control questions correctly. The experiment will start afterwards! If you have questions, please raise your hand at any time and an experimenter will provide assistance.

Auction Instructions

You will now participate in an auction against a computer opponent over a good that has a value of 20€. You can bid any amount from 0€ to 20€, and you can specify your bid down to the exact Cent. The computer will bid a random number from 0€ to 20€, down to the exact Cent, and each number has an equal probability to be chosen. If your bid is higher than the computer's, you will get 20€ minus your bid as your payoff. If your bid is lower than the computer's, you will get 0€. If your bids are tied, the winner of the auction is selected randomly and you will receive the payoff of 20€ minus your bid with 50% probability.

Example 1: If you bid 12.41€ and the computer bids 16.53€, your payoff is 0€ as the computer's bid is higher than yours.
Example 2: If you bid 18.8€ and the computer bids 0.17€, your payoff is 1.2€ (=20-18.8) from this auction, as your bid is higher than the computer's.

Remember, you can bid any amount from 0€ to 20€. If you win the auction, your payoff is 20€ minus your bid. Now, please type in how much you want to bid for the good.

Investment Game Instructions

You will now have the opportunity to invest an endowment of 10.00€. There are two assets you can invest in: STOCKS and BONDS. The amount you invest in bonds does not give returns. You will get the amount you invested as your payoff for sure. STOCKS can have higher gains than BONDS, but are more risky. The amount you invest in STOCKS has 50% chance to be multiplied by 1.5, and 50% chance to be lost. You can freely allocate your endowment of 10.00€ between the two assets, down to the exact Cent.

Example 1: You invest 10€ in BONDS and 0€ in STOCKS. Your payoff will be 10€.
Example 2: You choose to invest 2.58€ in BONDS and 7.42€ in STOCKS. Your payoff will either be 13.71€ (=7.42*1.5+2.58) with 50% probability, or 2.58€ with 50% probability.
Example 3: You choose to invest all 10€ into STOCKS. Your payoff will either be 15€ (=10*1.5) with 50% probability, or 0€ with 50% probability.

Remember, you will receive the amount you invest in BONDS as your payoff. You will receive the amount you invest in STOCKS 1.5 times with 50% probability, and 0€ with 50% probability. Please choose how much you want to invest in STOCKS (the rest of your endowment will be invested in BONDS):

8.1.2. Control Questions

1.: Suppose that your earnings depend on the following lottery. The computer randomly chooses row ’3’ to be relevant for your payoff. Your choice was to switch from ’LEFT’ to ’RIGHT’ at row ’7’. How high is your payoff in the best case?
2.: Suppose that your earnings depend on the following lottery. What is the probability in % for you getting 4.50€ if the computer randomly chooses row ’8’ to be relevant for your payoff and you clicked radio button ’5’?
3.: Suppose that your earnings depend on the following lottery. You have chosen to switch from ’LEFT to ’RIGHT’ at row ’2’. The computer randomly chose row ’2’ to be relevant for your payoff. What is the probability in % that you will get 8.60€?
4.: If you invest 7€ in STOCKS, what is the probability in % that you will receive 20.5€?

5.: How much do you receive as payoff if you decided to invest 4€ in STOCKS and you prove to be lucky with your investment?

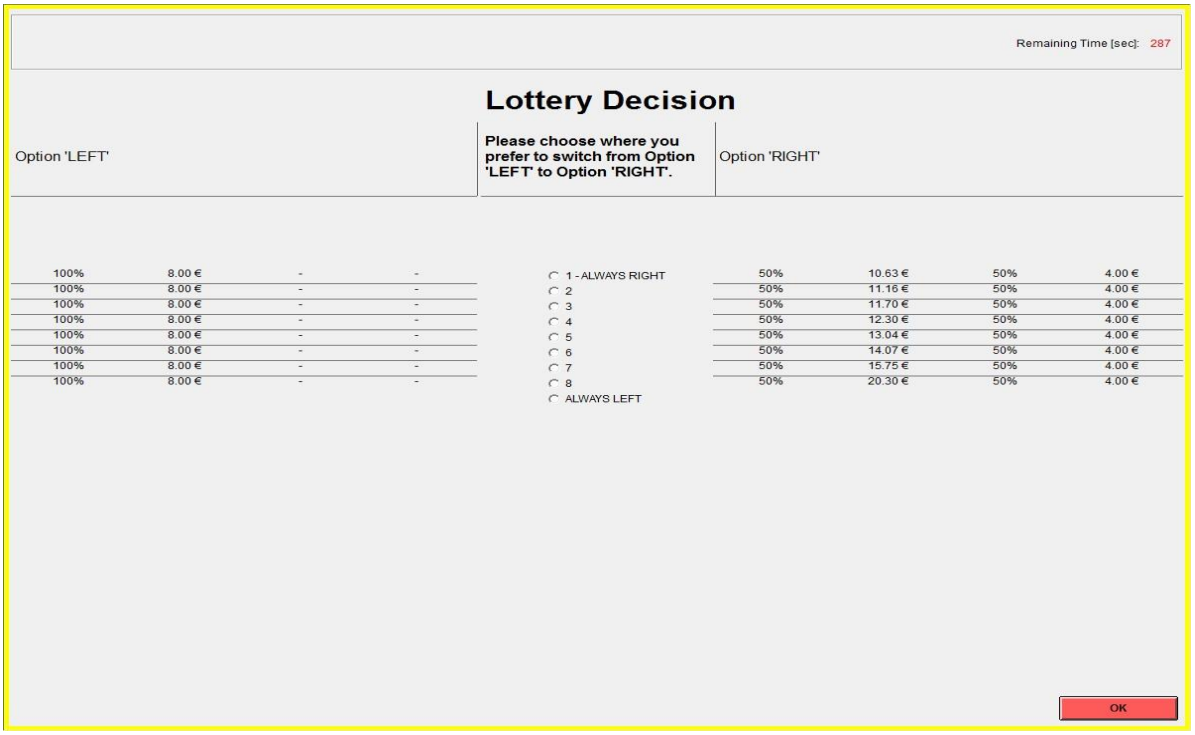
6.: If you invest 9€ in STOCKS, what is the probability in % that you will receive 22.5€?

7.: Your bid is 13.5€ and the computer's bid is 12.0€. How much is your payoff?

8.: Your bid is 8.0€ and the computer's bid is 13.4€. How much is your payoff?

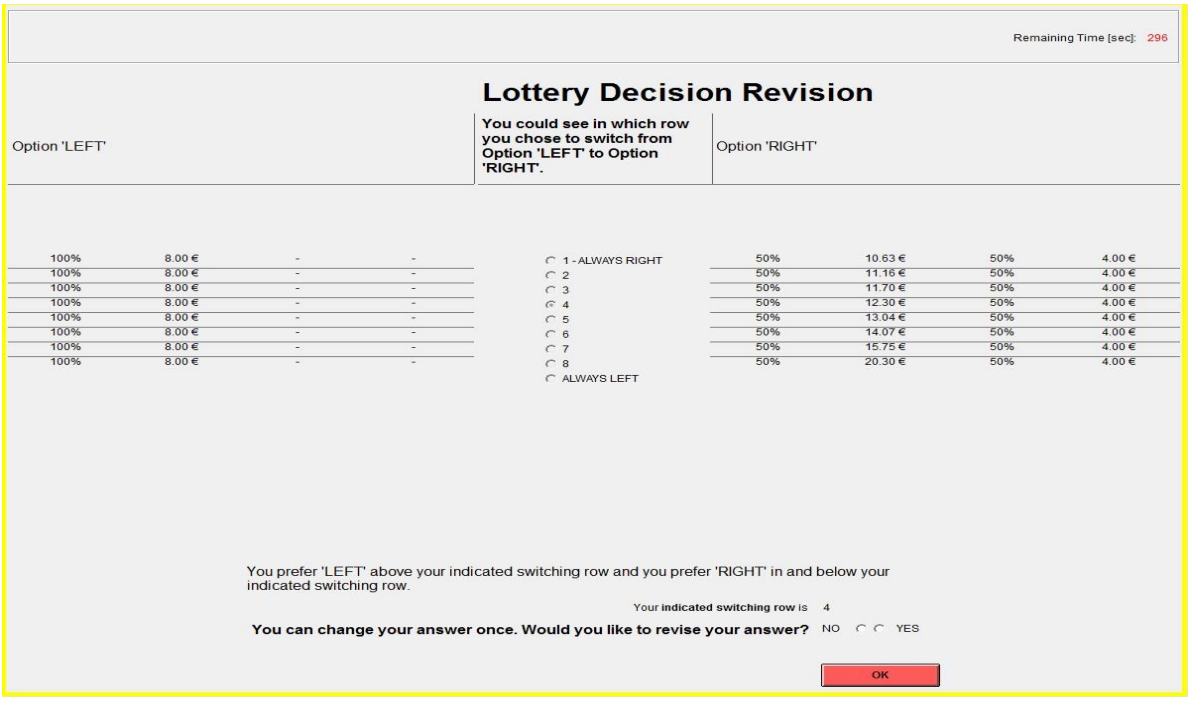
8.1.3. Decision and Feedback Screen

Figure 8.1.3: Decision-making Screen for Lotteries



Notes: Subjects indicated in which row they wanted to switch from the left to the right option by clicking one of the radio buttons in the middle.

Figure 8.1.4: Revision Screen for Lotteries



Notes: Subjects indicated whether they wanted to revise their first decision.

8.1.4. Risk Elicitation Methods

In this section we report the parameters for the gambles as they were presented to subjects. Note that all the preference elicitation methods described here are represented top-down for simplicity reasons. An example of a bottom-up representation of a particular method is provided afterwards.

Table 8.1.1: SGp Method

Left option				Right option			
Probability	Payoff	Probability	Payoff	Probability	Payoff	Probability	Payoff
100%	8.00	-	-	62.0%	4.00	38.0%	12.00
100%	8.00	-	-	56.4%	4.00	43.6%	12.00
100%	8.00	-	-	52.0%	4.00	48.0%	12.00
100%	8.00	-	-	48.1%	4.00	51.9%	12.00
100%	8.00	-	-	44.7%	4.00	55.3%	12.00
100%	8.00	-	-	41.1%	4.00	58.9%	12.00
100%	8.00	-	-	37.4%	4.00	62.6%	12.00
100%	8.00	-	-	32.3%	4.00	67.7%	12.00

Table 8.1.2: SGhigh Method

Left option				Right option			
Probability	Payoff	Probability	Payoff	Probability	Payoff	Probability	Payoff
1	8.00	-	-	50.0%	4.00	50.0%	10.63
1	8.00	-	-	50.0%	4.00	50.0%	11.16
1	8.00	-	-	50.0%	4.00	50.0%	11.70
1	8.00	-	-	50.0%	4.00	50.0%	12.32
1	8.00	-	-	50.0%	4.00	50.0%	13.04
1	8.00	-	-	50.0%	4.00	50.0%	14.07
1	8.00	-	-	50.0%	4.00	50.0%	15.75
1	8.00	-	-	50.0%	4.00	50.0%	20.31

Table 8.1.3: SGlow Method

Left option				Right option			
Probability	Payoff	Probability	Payoff	Probability	Payoff	Probability	Payoff
100%	9.50	-	-	50.0%	6.10	50.0%	12.00
100%	9.50	-	-	50.0%	6.61	50.0%	12.00
100%	9.50	-	-	50.0%	6.89	50.0%	12.00
100%	9.50	-	-	50.0%	7.09	50.0%	12.00
100%	9.50	-	-	50.0%	7.24	50.0%	12.00
100%	9.50	-	-	50.0%	7.38	50.0%	12.00
100%	9.50	-	-	50.0%	7.51	50.0%	12.00
100%	9.50	-	-	50.0%	7.66	50.0%	12.00

Table 8.1.4: SGsure Method

Left option				Right option			
Probability	Payoff	Probability	Payoff	Probability	Payoff	Probability	Payoff
100%	8.91	-	-	50.0%	4.00	50.0%	12.00
100%	8.50	-	-	50.0%	4.00	50.0%	12.00
100%	8.16	-	-	50.0%	4.00	50.0%	12.00
100%	7.85	-	-	50.0%	4.00	50.0%	12.00
100%	7.57	-	-	50.0%	4.00	50.0%	12.00
100%	7.27	-	-	50.0%	4.00	50.0%	12.00
100%	6.96	-	-	50.0%	4.00	50.0%	12.00
100%	6.56	-	-	50.0%	4.00	50.0%	12.00

Table 8.1.5: SGall Method

Left option				Right option			
Probability	Payoff	Probability	Payoff	Probability	Payoff	Probability	Payoff
100%	8.00	-	-	50.0%	10.00	50.0%	5.34
50.0%	10.00	50.0%	5.34	50.0%	11.00	50.0%	3.83
50.0%	11.00	50.0%	3.83	50.0%	12.00	50.0%	2.61
50.0%	12.00	50.0%	2.61	50.0%	13.00	50.0%	1.83
50.0%	13.00	50.0%	1.83	50.0%	14.00	50.0%	1.41
50.0%	14.00	50.0%	1.41	50.0%	15.00	50.0%	1.21
50.0%	15.00	50.0%	1.21	50.0%	16.50	50.0%	1.09
50.0%	16.50	50.0%	1.09	50.0%	20.50	50.0%	1.01

Table 8.1.6: PGp Method

Left option				Right option			
Probability	Payoff	Probability	Payoff	Probability	Payoff	Probability	Payoff
20.0%	9.00	80.0%	7.20	20.0%	17.20	80.0%	0.45
30.0%	9.00	70.0%	7.20	30.0%	17.20	70.0%	0.45
40.0%	9.00	60.0%	7.20	40.0%	17.20	60.0%	0.45
50.0%	9.00	50.0%	7.20	50.0%	17.20	50.0%	0.45
60.0%	9.00	40.0%	7.20	60.0%	17.20	40.0%	0.45
70.0%	9.00	30.0%	7.20	70.0%	17.20	30.0%	0.45
80.0%	9.00	20.0%	7.20	80.0%	17.20	20.0%	0.45
90.0%	9.00	10.0%	7.20	90.0%	17.20	10.0%	0.45

Table 8.1.7: PGhigh Method

Left option				Right option			
Probability	Payoff	Probability	Payoff	Probability	Payoff	Probability	Payoff
50.0%	9.00	50.0%	7.20	50.0%	10.96	50.0%	3.70
50.0%	9.00	50.0%	7.20	50.0%	11.55	50.0%	3.70
50.0%	9.00	50.0%	7.20	50.0%	12.15	50.0%	3.70
50.0%	9.00	50.0%	7.20	50.0%	12.87	50.0%	3.70
50.0%	9.00	50.0%	7.20	50.0%	13.75	50.0%	3.70
50.0%	9.00	50.0%	7.20	50.0%	15.01	50.0%	3.70
50.0%	9.00	50.0%	7.20	50.0%	17.21	50.0%	3.70
50.0%	9.00	50.0%	7.20	50.0%	23.83	50.0%	3.70

Table 8.1.8: PGlow Method

Left option				Right option			
Probability	Payoff	Probability	Payoff	Probability	Payoff	Probability	Payoff
50.0%	16.09	50.0%	7.00	50.0%	17.20	50.0%	3.70
50.0%	15.30	50.0%	7.00	50.0%	17.20	50.0%	3.70
50.0%	14.41	50.0%	7.00	50.0%	17.20	50.0%	3.70
50.0%	13.35	50.0%	7.00	50.0%	17.20	50.0%	3.70
50.0%	12.18	50.0%	7.00	50.0%	17.20	50.0%	3.70
50.0%	10.85	50.0%	7.00	50.0%	17.20	50.0%	3.70
50.0%	9.29	50.0%	7.00	50.0%	17.20	50.0%	3.70
50.0%	7.35	50.0%	7.00	50.0%	17.20	50.0%	3.70

Table 8.1.9: PGall Method

Left option				Right option			
Probability	Payoff	Probability	Payoff	Probability	Payoff	Probability	Payoff
99.0%	7.55	1.0%	0.00	81.0%	8.08	19.0%	0.00
94.0%	7.93	6.0%	0.00	78.0%	8.73	22.0%	0.00
89.0%	8.28	11.0%	0.00	75.0%	9.28	25.0%	0.00
84.0%	8.60	16.0%	0.00	72.0%	9.83	28.0%	0.00
79.0%	8.98	21.0%	0.00	69.0%	10.53	31.0%	0.00
74.0%	9.33	26.0%	0.00	66.0%	11.33	34.0%	0.00
69.0%	9.70	31.0%	0.00	63.0%	12.90	37.0%	0.00
64.0%	10.05	36.0%	0.00	62.0%	28.95	38.0%	0.00

Example for Bottom-Up Representation

Table 8.1.10: PGhigh Method – Bottom-Up Appearance

Left option				Right option			
Probability	Payoff	Probability	Payoff	Probability	Payoff	Probability	Payoff
50.0%	23.83	50.0%	3.70	50.0%	9.00	50.0%	7.20
50.0%	17.21	50.0%	3.70	50.0%	9.00	50.0%	7.20
50.0%	15.01	50.0%	3.70	50.0%	9.00	50.0%	7.20
50.0%	13.75	50.0%	3.70	50.0%	9.00	50.0%	7.20
50.0%	12.87	50.0%	3.70	50.0%	9.00	50.0%	7.20
50.0%	12.15	50.0%	3.70	50.0%	9.00	50.0%	7.20
50.0%	11.55	50.0%	3.70	50.0%	9.00	50.0%	7.20
50.0%	10.96	50.0%	3.70	50.0%	9.00	50.0%	7.20

8.1.5. Robustness Checks

Table 8.1.11 shows standard correlation coefficients between methods; the results are qualitatively the same as in Table 3.5.

Table 8.1.11: Correlation Coefficients between the Methods

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall	GQ
SGhigh	0.46***									
SGlow	0.37***	0.46***								
SGsure	0.01	0.18	0.25*							
SGall	0.02	0.12	0.00	0.16						
PGp	0.13	0.12	0.15	0.23**	-0.08					
PGhigh	0.07	0.27**	0.10	0.26*	0.12	0.26*				
PGlow	0.27**	0.21*	0.20	0.17	-0.04	0.12	0.10			
PGall	0.17	0.16	-0.06	0.04	0.17	-0.01	-0.08	0.03		
GQ	0.12	0.16	0.00	-0.14	0.09	-0.07	0.12	0.02	0.01	
FQ	0.25**	0.17	0.27**	0.19	0.07	-0.05	-0.02	0.20	0.03	0.46**

Notes: This table shows standard correlations as opposed to Spearman rank correlations as in Table 3.5. Our conclusions remain qualitatively the same. Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

Table 8.1.12: Testing for Order Effects

Method	Standard deviation of CRRA coefficient	CRRA coefficient
SGp	0.02	-0.02
SGhigh	-0.01	-0.01
SGlow	-0.02	0.03
SGsure	0.03	-0.04
SGall	-0.01	0.00
PGp	-0.02	0.03
PGhigh	0.00	0.01
PGlow	-0.01	0.02
PGall	-0.01	0.01
Overall	0.00	0.00

Notes: This table reports the coefficients of the OLS regressions. The dependent variable is the standard deviation (column 1) of the CRRA-coefficient or the CRRA score (column 2) across all subjects in one particular order. The independent variable is the order in which a method appeared, i.e. the number of previously encountered MPLs. Stars are given as follows in the absence of the multiple testing problem: *: p<0.1; **: p<0.05; ***: p<0.01.

Table 8.1.12 tests for linear order effects in the data: As the nine methods were presented to subjects in a randomized order, risk aversion or noise might increase or decrease as a function of previously seen MPLs. The coefficients in Table 8.1.11 can be interpreted as a change in the standard deviation (i.e. noise) of CRRA coefficients or changes in the CRRA coefficients

themselves as a function of order, i.e. the number of previously encountered MPLs. No significant effects were found. For robustness checks on the functional forms (PT, CARA, DRRA, IRRA, IARA, DARA and probability weighting), please refer to Tables 8.1.14 – 8.1.25.

Table 8.1.13: Comparison of Results to Previous Studies

Our results			Previous studies				t-test
Method	Mean	St. dev.	Mean	St. dev.	Subjects	Study	p-value
SGp	0.02	0.72	0.45	0.45	157	Bruner (2009)	0.01***
SGhigh	0.40	0.65	0.51	0.59	157	Bruner (2009)	0.17
SGlow	0.69	0.68	-	-	-	-	-
SGsure	0.31	0.66	0.20	0.08	39	Abdellaoui et al. (2011)	0.30
SGall	0.07	0.63	0.60	0.59	256	Eckel and Grossman (2008)	0.01***
			0.73	0.9	30	Reynaud and Couture (2012)	0.01***
			0.69	0.33	444	Crosetto and Filippin (2016)	0.01***
PGp	0.62	0.54	0.32	0.41	175	Holt and Laury (2002)	0.01***
			0.23	0.14	39	Abdellaoui et al. (2011)	0.01***
			0.59	0.07	100	Drichoutis and Lusk (2012)	0.15
			0.39	0.54	78	Dulleck et al. (2015)	0.01***
			0.43	0.60	444	Crosetto and Filippin (2016)	0.01***
			0.62	0.80	268	Andersen et al. (2008a)	1
			0.67	0.57	881	Dave et al. (2010)	0.46
PGhigh	0.87	0.56	0.35	0.18	100	Drichoutis and Lusk (2012)	0.01***
PGlow	0.57	0.67	0.35	0.18	100	Drichoutis and Lusk (2012)	0.01***
PGall	0.47	0.62	0.82	-	86	Lejuez et al. (2002)	-
			1.13	0.64	444	Crosetto and Filippin (2016)	0.01***
			0.70	0.83	444	Crosetto and Filippin (2013)	0.02**

Notes: This table shows the differences between our results and previous studies. We report the means and the standard deviation in terms of the CRRA-coefficients. The number of observations is 96 in our experiment. The PGp method by Crosetto and Filippin (2016) follows the balloon representation by Lejuez et al. (2002). No standard deviation was reported in the original study by Lejuez et al. (2002).

In Table 8.1.13 we see the differences in the mean values of CRRA risk coefficients between previous studies and our results for each method. We find significant differences in many cases. Unfortunately, a comparison of distributions was not possible, as many studies do not report the exact distributions of CRRA coefficients. Note that the studies by Eckel and Grossman (2008), Reynaud and Couture (2012) or Crosetto and Filippin (2016) do not cover the risk loving domain in the SGall task, which primarily drives the differences there: The pull-to-center effect drives the risk estimates to be higher here, which is also suggested by Bleichrodt (2002). The PGall task has a very different framing – bomb and balloon – for each study. Besides these exceptions, the relatively noisy nature of methods is confirmed, which makes the need for a stable method even more pronounced.

8.1.6. Tests Under Different Functional Forms

In this section, we report similar regressions to Table 3.7 with the different assumption that a subject's utility function does not follow CRRA but either PT, CARA, DRRA, IRRA, IARA or DARA. The results remain unchanged or provide an even more favorable picture. PGhigh becomes a significant explanatory factor under DRRA, and several other methods lose their significance under the other specifications. We conclude that our results are robust to the assumption of the underlying utility function. In this section we also report regressions that take probability weighting (PW) into account. Note that, under this specification, the only methods that change compared to the tables in the above section are SGp, PGp and PGall, as these have changing probabilities over different rows. The conclusions we drew for Table 3.7 in the main text remain qualitatively the same if we assume probability weighting.

Table 8.1.14: Explanatory Power – CARA

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall
OLS coefficients									
Auction	5.17	4.76	5.47	0.20	-1.82	0.26	-0.04	5.01	-2.20
	(0.05)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)
Inv. Low	-0.81	-7.65*	-4.57	0.50	-4.51	-2.65	-12.11**	-1,83	-1.72
	(0.00)	(0.03)	(0.00)	(0.00)	(0.01)	(0.00)	(0.07)	(0.00)	(0.00)
Inv. High	-6.87**	-5.48	-2.63	1.82	3.79	-2.93	-5.55	-5.61	0.87
	(0.16)	(0.12)	(0.09)	(0.09)	(0.11)	(0.10)	(0.10)	(0.12)	(0.09)
Spearman rank correlation coefficients									
Auction	0.23*	0.09	0.14	0.17	0.07	0.11	0.06	0.16	-0.13
Inv. Low	0.02	0.19	0.17	0.06	0.06	0.11	0.36***	0.12	0.04
Inv High	0.28**	0.28**	0.05	0.00	0.03	0.13	0.26**	0.23*	0.09

Notes: Compared to Table 3.7, significance remains the same, and R^2 values are lower. Spearman rank correlation coefficients are unchanged due to the nature of the functional form. Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

Table 8.1.15: Explanatory Power – Prospect Theory

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall
OLS coefficients									
Auction	-0.68	-0.53	-0.58	0.02	0.29	-0.03	0.00	-0.59	0.52
	(0.05)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)
Inv. Low	0.09	0.94*	0.48	-0.01	0.66	0.39	1.48**	0.23	0.44
	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.08)	(0.00)	(0.00)
Inv. High	0.91**	0.68	0.28	-0.21	-0.58	0.46	0.66	0.65	-0.20
	(0.16)	(0.12)	(0.09)	(0.09)	(0.11)	(0.09)	(0.11)	(0.12)	(0.08)
Spearman rank correlation coefficients									
Auction	0.23*	0.09	0.14	0.17	0.07	0.11	0.06	0.16	-0.13
Inv. Low	0.02	0.19	0.17	0.06	0.06	0.11	0.36***	0.12	0.04
Inv High	0.28**	0.28**	0.05	0.00	0.03	0.29	0.26**	0.23*	0.09

Notes: Spearman rank correlations and adjusted R^2 values are unchanged compared to Table 3.7 due to the similarity of the functional form to CRRA. The expected signs are the opposite as for the other functional forms. Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

Table 8.1.16: Explanatory Power – DRRA

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall
OLS coefficients									
Auction	0.56	0.45	0.48	-0.07	-0.20	0.01	-0.03	0.42	-0.40
	(0.05)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)
Inv. Low	-0.03	-0.76*	-0.38	-0.02	-0.53	-0.32	-1.26***	-0.16	-0.26
	(0.00)	(0.03)	(0.00)	(0.00)	(0.01)	(0.00)	(0.09)	(0.00)	(0.00)
Inv. High	-0.72**	-0.58	-0.24	0.21	0.46	-0.35	-0.55*	-0.53*	0.13
	(0.17)	(0.12)	(0.09)	(0.09)	(0.11)	(0.09)	(0.11)	(0.12)	(0.08)
Spearman rank correlation coefficients									
Auction	0.24*	0.11	0.17	0.15	0.06	0.12	0.07	0.18	-0.10
Inv. Low	0.02	0.19	0.17	0.06	0.06	0.11	0.37***	0.13	0.02
Inv High	0.28**	0.29*	0.08	0.03	0.03	0.15	0.28**	0.27**	0.10

Notes: Coefficients, p-values and R^2 values are slightly better compared to Table 3.7. Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

Table 8.1.17: Explanatory Power – IRRA

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall
OLS coefficients									
Auction	0.32	0.22	0.24	0.01	-0.11	0.02	0.00	0.25	-0.27
	(0.05)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)
Inv. Low	-0.06	-0.42*	-0.19	-0.02	-0.32	-0.18	-0.67***	-0.09	-0.16
	(0.00)	(0.03)	(0.00)	(0.00)	(0.01)	(0.00)	(0.10)	(0.00)	(0.00)
Inv. High	-0.42**	-0.32	-0.13	-0.12	0.27	-0.21	-0.29	-0.31*	0.06
	(0.17)	(0.12)	(0.09)	(0.09)	(0.11)	(0.10)	(0.11)	(0.13)	(0.08)
Spearman rank correlation coefficients									
Auction	0.23*	0.11	0.16	0.16	0.06	0.12	0.07	0.18	-0.10
Inv. Low	0.02	0.20	0.18	0.06	0.06	0.11	0.37***	0.14	0.03
Inv High	0.30**	0.29**	0.08	0.03	0.03	0.15	0.28**	0.27**	0.10

Notes: Coefficients, p-values and R^2 values are slightly better compared to Table 3.7. Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

Table 8.1.18: Explanatory Power – DARA

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall
OLS coefficients									
Auction	3.51	3.23	3.67	-0.16	-1.16	0.84	-0.14	3.32	-1.93
	(0.05)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)
Inv. Low	-0.43	-5.12**	-2.98	-0.07	-3.07	-1.82	-8.27***	-1.28	-1.73
	(0.00)	(0.04)	(0.00)	(0.00)	(0.01)	(0.00)	(0.08)	(0.00)	(0.00)
Inv. High	-4.59**	-3.68	-1.82	1.12	2.38	-1.76	-3.76	-3.89*	0.27
	(0.17)	(0.12)	(0.09)	(0.09)	(0.10)	(0.09)	(0.11)	(0.13)	(0.08)
Spearman rank correlation coefficients									
Auction	0.25*	0.11	0.18	0.15	0.07	0.12	0.07	0.18	-0.12
Inv. Low	0.00	0.22*	0.18	0.06	0.07	0.11	0.37***	0.12	0.06
Inv High	0.28**	0.31**	0.02	0.09	0.05	0.13	0.29**	0.27**	0.11

Notes: Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

Table 8.1.19: Explanatory Power – IARA

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall
OLS coefficients									
Auction	7.40	6.79	8.00	0.21	-3.01	0.45	0.16	7.37	-2.94
	(0.04)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)
Inv. Low	-1.47	-11.20*	-6.84	0.52	-6.48	-3.74	-17.29**	-2.49	-2.19
	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.06)	(0.00)	(0.00)
Inv. High	-10.09**	-8.00	-3.62	2.72	5.92	-4.29	-8.02	-7.85	1.43
	(0.15)	(0.11)	(0.09)	(0.09)	(0.11)	(0.10)	(0.10)	(0.11)	(0.09)
Spearman rank correlation coefficients									
Auction	0.21*	0.18	0.11	0.17	0.07	0.11	0.03	0.13	-0.13
Inv. Low	0.03	0.16	0.15	0.05	0.05	0.10	0.32***	0.13	0.03
Inv High	0.28**	0.24*	0.01	0.01	0.01	0.12	0.19	0.19	0.07

Notes: In general, we see somewhat lower R^2 values than for DARA. Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

Table 8.1.20: Explanatory Power – CARA plus Probability Weighting

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall
OLS coefficients									
Auction	8.44	4.76	5.47	0.20	-1.82	0.45	-0.04	5.01	-1.83
	(0.05)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)
Inv. Low	-1.03	-7.65*	-4.57	0.50	-4.51	-4.36	-12.11**	-1,83	-1.90
	(0.00)	(0.03)	(0.00)	(0.00)	(0.01)	(0.00)	(0.07)	(0.00)	(0.00)
Inv. High	-11.22**	-5.48	-2.63	1.82	3.79	-4.82	-5.55	-5.61	0.47
	(0.16)	(0.12)	(0.09)	(0.09)	(0.11)	(0.10)	(0.10)	(0.12)	(0.08)
Spearman rank correlation coefficients									
Auction	0.23*	0.09	0.14	0.17	0.07	0.11	0.06	0.16	-0.13
Inv. Low	0.02	0.19	0.17	0.06	0.06	0.11	0.36***	0.12	0.04
Inv High	0.28**	0.28**	0.05	0.00	0.03	0.13	0.26**	0.23*	0.09

Notes: Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

Table 8.1.21: Explanatory Power – Cumulative Prospect Theory

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall
OLS coefficients									
Auction	-1.56	-0.53	-0.58	0.02	0.29	-0.05	0.00	-0.59	0.09
	(0.05)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.02)
Inv. Low	0.19	0.94*	0.48	-0.01	0.66	0.66	1.48**	0.23	0.18
	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.08)	(0.00)	(0.00)
Inv. High	2.06**	0.68	0.28	-0.21	-0.58	0.76	0.66	0.65	-0.11
	(0.16)	(0.12)	(0.09)	(0.09)	(0.11)	(0.09)	(0.11)	(0.12)	(0.08)
Spearman rank correlation coefficients									
Auction	0.23*	0.09	0.14	0.17	0.07	0.11	0.06	0.16	-0.13
Inv. Low	0.02	0.19	0.17	0.06	0.06	0.11	0.36***	0.12	0.04
Inv High	0.28**	0.28**	0.05	0.00	0.03	0.13	0.26**	0.23*	0.09

Notes: The expected signs are the opposite as for the other functional forms. Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

Table 8.1.22: Explanatory Power – DRRA plus Probability Weighting

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall
OLS coefficients									
Auction	1.00	0.45	0.48	-0.07	-0.20	0.02	-0.03	0.42	-0.32
	(0.04)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)
Inv. Low	-0.12	-0.76*	-0.38	-0.02	-0.53	-0.53	-1.26***	-0.16	-0.28
	(0.00)	(0.03)	(0.00)	(0.00)	(0.01)	(0.00)	(0.09)	(0.00)	(0.00)
Inv. High	-1.31	-0.58	-0.24	0.21	0.46	-0.56	-0.55*	-0.53*	0.09
	(0.15)	(0.12)	(0.09)	(0.09)	(0.11)	(0.09)	(0.11)	(0.12)	(0.08)
Spearman rank correlation coefficients									
Auction	0.22*	0.11	0.17	0.15	0.06	0.11	0.07	0.18	-0.10
Inv. Low	0.03	0.19	0.17	0.06	0.06	0.09	0.37***	0.13	0.03
Inv High	0.26*	0.29*	0.08	0.03	0.03	0.12	0.28**	0.27**	0.11

Notes: Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

Table 8.1.23: Explanatory Power – IRRA plus Probability Weighting

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall
OLS coefficients									
Auction	0.76	0.22	0.24	0.01	-0.11	0.01	0.00	0.25	-0.23
	(0.03)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)
Inv. Low	-0.63	-0.42*	-0.19	-0.02	-0.32	-0.27	-0.67***	-0.09	-0.16
	(0.00)	(0.03)	(0.00)	(0.00)	(0.01)	(0.00)	(0.10)	(0.00)	(0.00)
Inv. High	-1.00	-0.32	-0.13	-0.12	0.27	-0.28	-0.29	-0.31*	0.03
	(0.11)	(0.12)	(0.09)	(0.09)	(0.11)	(0.09)	(0.11)	(0.13)	(0.08)
Spearman rank correlation coefficients									
Auction	0.21*	0.11	0.16	0.16	0.06	0.11	0.07	0.18	-0.10
Inv. Low	0.01	0.20	0.18	0.06	0.06	0.09	0.37***	0.14	0.04
Inv High	0.27**	0.29**	0.08	0.03	0.03	0.12	0.28**	0.27**	0.11

Notes: Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

Table 8.1.24: Explanatory Power – DARA plus Probability Weighting

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall
OLS coefficients									
Auction	5.83	3.23	3.67	-0.16	-1.16	0.27	-0.14	3.32	-1.53
	(0.05)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)
Inv. Low	-1.1	-5.12**	-2.98	-0.07	-3.07	-3.00	-8.27***	-1.28	-1.80
	(0.00)	(0.04)	(0.00)	(0.00)	(0.01)	(0.00)	(0.08)	(0.00)	(0.00)
Inv. High	-8.10**	-3.68	-1.82	1.12	2.38	-3.22	-3.76	-3.89*	-0.01
	(0.17)	(0.12)	(0.09)	(0.09)	(0.10)	(0.10)	(0.11)	(0.13)	(0.08)
Spearman rank correlation coefficients									
Auction	0.23*	0.11	0.18	0.15	0.07	0.12	0.07	0.18	-0.12
Inv. Low	0.20	0.22*	0.18	0.06	0.07	0.11	0.37***	0.12	0.07
Inv High	0.30**	0.31**	0.02	0.09	0.05	0.15	0.29**	0.27**	0.11

Notes: Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

Table 8.1.25: Explanatory Power – IARA plus Probability Weighting

	SGp	SGhigh	SGlow	SGsure	SGall	PGp	PGhigh	PGlow	PGall
OLS coefficients									
Auction	12.16	6.79	8.00	0.21	-3.01	0.70	0.16	7.37	-2.03
	(0.04)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)
Inv. Low	-2.22	-11.20*	-6.84	0.52	-6.48	-6.00	-17.29**	-2.49	-1.94
	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.06)	(0.00)	(0.00)
Inv. High	-16.35	-8.00	-3.62	2.72	5.92	-7.00	-8.02	-7.85	1.09
	(0.15)	(0.11)	(0.09)	(0.09)	(0.11)	(0.10)	(0.10)	(0.11)	(0.09)
Spearman rank correlation coefficients									
Auction	0.22*	0.18	0.11	0.17	0.07	0.11	0.03	0.13	-0.14
Inv. Low	0.02	0.16	0.15	0.05	0.05	0.10	0.32***	0.13	0.03
Inv High	0.27**	0.24*	0.01	0.01	0.01	0.11	0.19	0.19	0.07

Notes: Stars are given as follows: *: p<0.05; **: p<0.01; ***: p<0.001.

8.2. Biased Probability Judgment and Other Bias in the Newsvendor Setting

8.2.1. General Instructions

Welcome to this experiment in decision making.

You will be asked to make a series of choices that will affect your payment at the end of the experiment. This experiment consists of 3 different parts. In each part, there will be one randomly chosen decision that will be relevant for your payoff. You will receive feedback on your performance after you completed all 3 parts. Please pay close attention to the instructions, and if you have any questions raise your hand and an employee of the lab will help you with any questions you might have. Throughout the experiment, the different payment options will be listed in so-called EXPERIMENTAL CURRENCY (EC). In the end, you will receive €1 for each 250 EC you have. For example, if you have 4683 EC at the end of the experiment, your payoff will be €18.73 in the end (as 4683 = 18.73 * 250). Thus, fractions of 250 EC can be converted to the exact Cent. Also, you will be asked to fill out a short questionnaire, where your answers are not relevant for your payoff. All your decisions and answers during the experiment will stay completely anonymous to everyone.

In 'Part 1', you will make decisions in six different tasks. One of the tasks will be randomly chosen to be relevant for your payoff. After you made these six decisions, 'Part 2' and 'Part 3'

will follow. Instructions for 'Part 2' and 'Part 3' will be provided later. Your final payoff will then be the sum of your payoffs from 'Part 1', 'Part 2' and 'Part 3'. For some of the tasks, it is possible to make losses. The experiment is designed in a way that your payments will be positive in the end. How high your payment is, depends on your choices throughout the experiment. If you have questions, please raise your hand at any time and an experimenter will provide assistance.

8.2.2. Elicitation Procedure for the Gambler's and Hot-hand Fallacies

Imagine a coin that will be flipped. The possible outcomes are either 'HEADS' or 'TAILS', and both outcomes are equally likely.

You will earn 2000 EC if you manage to guess the correct outcome of the coinflip, if the computer chooses Task I. to be relevant for your payoff. You will earn 0 EC if you do not manage to guess the correct outcome of the coinflip, if the computer chooses Task I. to be relevant for your payoff. How large do you think the probability in % is that the coinflip will result in 'TAILS'? The answer given to this question does not affect your payment.

Please, choose whether you bet on 'HEADS' or 'TAILS'! The coinflip will be simulated by the computer afterwards.

/In order to minimize experimenter demand effects, the following task was conducted 45-60 minutes after the task that first measured the perceptions the coinflip's outcomes. /

Imagine a coin that will be flipped. The possible outcomes are either 'HEADS' or 'TAILS', and both outcomes are equally likely.

You will earn 2000 EC if you manage to guess the correct outcome of the coinflip, if the computer chooses Task VI. to be relevant for your payoff. You will earn 0 EC if you do not manage to guess the correct outcome of the coinflip, if the computer chooses Task VI. to be relevant for your payoff. Additionally, you will see the results of the last 8 simulated coinflips on the next screen.

Suppose, the history of the coinflip was the following during the last 8 flips:

1st coinflip: 'TAILS'

2nd coinflip: 'TAILS'

3rd coinflip: 'TAILS'

4th coinflip: 'HEADS'

5th coinflip: 'TAILS'

6th coinflip: 'HEADS'

7th coinflip: 'HEADS'

Last coinflip: 'HEADS'

How large do you think the probability in % is that the 9th coinflip will result in 'TAILS'? The answer given to this question does not affect your payment.

Please, choose whether you bet on 'HEADS' or 'TAILS'! The coinflip will be simulated by the computer afterwards.

8.2.3. Elicitation Procedure for Regret Aversion

Please select which of the lotteries on the bottom of the screen ('LEFT' or 'RIGHT') you prefer.

50% WIN 2000 EC 50% WIN 400 EC OR 50% WIN 1400 EC 50% WIN 1000 EC

You chose the lottery on the LEFT (RIGHT) SIDE. Suppose that the computer plays it out for you and you win 400 (1400) EC. If you had chosen the lottery on the RIGHT (LEFT) SIDE, you would have won 1000 (2000) EC. On a scale from 1 to 10, how much do you regret that you chose the lottery on the LEFT (RIGHT) SIDE in this case? (1=”no regret”; “10=extreme regret”) YOUR ANSWER TO THIS QUESTION DOES NOT AFFECT THE OUTCOME OF THIS LOTTERY OR YOUR EARNINGS IN ANY MANNER.

8.2.4. Elicitation Procedure for Loss and Risk Aversion

Please choose where you prefer to switch from Option 'LEFT' to Option 'RIGHT'.

Option 'LEFT'			Option 'RIGHT'	
100%	0 EC	1	50% WIN 600 EC	50% LOSE 600 EC
100%	0 EC	2	50% WIN 600 EC	50% LOSE 599 EC
100%	0 EC	3	50% WIN 600 EC	50% LOSE 550 EC
100%	0 EC	4	50% WIN 600 EC	50% LOSE 500 EC
100%	0 EC	5	50% WIN 600 EC	50% LOSE 400 EC
100%	0 EC	6	50% WIN 600 EC	50% LOSE 300 EC
100%	0 EC	7	50% WIN 600 EC	50% LOSE 200 EC
100%	0 EC	8	50% WIN 600 EC	50% LOSE 100 EC
		9 – ALWAYS LEFT		

Please choose where you prefer to switch from Option 'LEFT' to Option 'RIGHT'.

Option 'LEFT'			Option 'RIGHT'	
50% WIN 1564 EC	50% WIN 400 EC	1	50% WIN 800 EC	50% WIN 600 EC
50% WIN 1067 EC	50% WIN 400 EC	2	50% WIN 800 EC	50% WIN 600 EC
50% WIN 1020 EC	50% WIN 400 EC	3	50% WIN 800 EC	50% WIN 600 EC
50% WIN 984 EC	50% WIN 400 EC	4	50% WIN 800 EC	50% WIN 600 EC
50% WIN 961 EC	50% WIN 400 EC	5	50% WIN 800 EC	50% WIN 600 EC
50% WIN 931 EC	50% WIN 400 EC	6	50% WIN 800 EC	50% WIN 600 EC
50% WIN 887 EC	50% WIN 400 EC	7	50% WIN 800 EC	50% WIN 600 EC
50% WIN 815 EC	50% WIN 400 EC	8	50% WIN 800 EC	50% WIN 600 EC
		9 – ALWAYS LEFT		

8.2.5. Elicitation Procedure for Anchoring and Overconfidence

In Task V., you will be asked 14 questions. On the first screen for Task V., you will receive 10 questions. You will have 5 minutes to answer these questions. If this task is chosen to be relevant for your payoff in 'Part 1', you will receive 200 EC for each question that you manage to answer correctly. For the questions with two choice options, please click on the provided answer that you think is correct. For the open questions, an answer is regarded as correct if your answer is within 10% range of the correct answer. So, if the computer chooses Task V. randomly to be relevant for your payoff and you answered 7 questions correctly, you will receive 1400 EC (as 1400 = 7 * 200).

In what year was George Washington elected President of the United States?

In what year did the second European explorer land in the New World?

What is the freezing point of vodka in degree Celsius?

What is the boiling point of water on top of the Mount Everest in degree Celsius?

Is the length of the Mississippi River greater or less than 3460 kilometers?

What is the length of the Mississippi River in kilometers?

Is the number of United Nations members greater or less than 29?

What is the number of United Nations members?

Was the telephone invented before or after 1920?

In which year was the telephone invented?

How many questions do you think you answered correctly among the previous 10 questions?

Consider all the other participants in this experiment (not only those here today). What percentage of them answers MORE from the previous 10 questions correctly than you did?

8.2.6. The Newsvendor Task

In 'Part 2', you are the owner of a store that sells 'WIDGETS', a fictional type of good. In each period, you will have to decide how many 'WIDGETS' you want to order from your supplier so that you can sell them to your customers. The number of customers coming into your store is

random in each period, and you do not know how many customers you will have in the upcoming period when you make your decision how many 'WIDGETS' you want to order. The number of customers coming into your store is a random number from 0 to 300 in each period. Every number is equally likely, and the numbers are independent from each other from period to period. Each customer buys one 'WIDGET' from you for 12 EC each. Each 'WIDGET' you order from the supplier costs you 3 EC. Remember, 250 EC convert to € 1. This means that if you order more 'WIDGETS' than there are customers, you will sell each customer one 'WIDGET' and the rest of them cannot be sold or transferred from period to period (i.e. they are lost).

Example: You order 200 'WIDGETS' and have 160 customers. You will earn 160*12= 1920 EC from the sold 'WIDGETS' and you have to pay your supplier 200*3= 600 EC. Your payoff from this period would therefore be 1320 EC (=1920-600). If you order less 'WIDGETS' than there are customers, you will sell all of your 'WIDGETS', but you will lose income since you could not satisfy some of your customers.

Example: You order 160 'WIDGETS' and there are 200 customers. You will earn 160*12= 1920 EC and have to pay 160*3= 480 EC. Your payoff will therefore be 1440 EC (=1920-480).

Example: You order 300 WIDGETS and have 5 customers. Your payoff would be -840 EC (=5*12-300*3).

So, to sum up, your payoff in a given period will be 12 * Amount Sold – 3 * Amount Ordered. You will decide how many units of 'WIDGETS' to order in each of the upcoming 35 periods. One of the 35 periods will be randomly chosen to be relevant for your payoff for 'Part 2'. You will receive feedback on your performance after every period including your potential payoff in EC which is based on the demand and your order quantity. This could help you to improve your performance. However, you will not know which period was chosen by the computer to be relevant for your payoff at this stage. This information will be provided after completing 'Part 2' and 'Part 3'. So, each and every decision could matter regarding your earnings.

Please, note that you will have to make your ordering decisions within a given time, which is 90 seconds for periods 1-5, 60 seconds for periods 6-10 and 30 seconds for the remaining periods. If you do not submit your order quantity within the given time, you will order 0 units of 'WIDGETS'.

Order Decision

Period: __

Please list how many items you would like to order during this round.

Order Feedback

Period: __

The number of customers during this period was: __

Your order quantity was: __

You sold the following amount: __

Your costs from ordering 'WIDGETS' in EC was: __

Your income from sales of 'WIDGETS' in EC was: __

If this period is chosen to be relevant for your payoff, your profit in EC during this round will be: __

8.2.7. Elicitation Procedure for Personality Traits

How well do the following statements describe your personality?

I see myself as someone who...

STRONGLY DISAGREE				STRONGLY AGREE
X	X	X	X	X

... is reserved

... is generally trusting

... tends to be lazy

... is relaxed, handles stress well

... has few artistic interests

... is outgoing, sociable

... tends to find fault with others

... does a thorough job

... gets nervous easily

... has an active imagination

8.2.8. Elicitation Procedure Cognitive Reflection

1. A bat and a ball cost 1.10 € in total. The bat costs 1.00 € more than the ball. How much does the ball cost? (in cents)
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (in minutes)
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? (in days)

8.2.9. The Correlation Matrix of Elicited Preferences

Table 8.2.1: Correlation Matrix of Elicited Individual Factors in the High Margin Setting

	1.	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**	**8.**	**9.**	**10.**	**11.**	**12.**	**13.**	**14.**	**15.**	**16.**	**17.**	**18.**	**19.**
1.	1.0	-0.2	0.0	-0.1	0.0	0.1	0.2	0.1	0.4	-0.2	0.1	-0.1	-0.1	0.1	-0.1	0.1	-0.1	0.0	-0.1
2.		1.0	-0.2	0.0	0.0	0.0	0.4	0.1	0.0	0.1	0.3	-0.2	0.0	0.0	0.1	-0.2	0.3	0.0	0.1
3.			1.0	0.1	0.0	0.0	-0.4	0.0	0.0	-0.2	0.0	0.2	0.2	-0.1	-0.1	-0.1	-0.1	0.0	0.1
4.				1.0	-0.2	-0.1	-0.2	-0.1	0.0	0.0	0.0	0.1	0.0	0.2	-0.3	0.0	-0.1	0.1	-0.2
5.					1.0	0.1	-0.1	0.1	-0.1	0.0	-0.1	-0.3	0.1	-0.1	-0.1	0.1	0.0	-0.2	-0.1
6.						1.0	0.1	0.2	0.0	0.1	0.1	0.1	0.3	0.1	0.0	0.1	0.2	0.1	0.3
7.							1.0	0.2	0.1	0.1	0.1	-0.2	-0.1	0.3	0.0	0.2	0.1	0.0	0.1
8.								1.0	0.1	0.2	0.0	-0.2	0.3	-0.1	0.0	0.0	0.4	-0.1	0.2
9.									1.0	0.1	-0.1	-0.1	0.1	0.1	0.0	0.1	-0.1	-0.1	0.0
10.										1.0	0.0	0.1	0.8	0.0	0.0	0.1	0.1	0.1	0.2
11.											1.0	-0.1	0.0	-0.1	0.2	-0.3	0.4	0.0	0.0
12.												1.0	0.1	0.2	-0.1	0.2	-0.2	0.3	0.0
13.													1.0	-0.1	0.0	0.0	0.2	0.0	0.1
14.														1.0	-0.3	0.4	-0.4	0.4	0.0
15.															1.0	-0.3	0.3	-0.3	-0.1
16.																1.0	-0.4	0.1	0.1
17.																	1.0	-0.2	0.1
18.																		1.0	0.0
19.																			1.0

Notes: 1: Risk attitude, 2: Regret aversion, 3: Loss aversion, 4: Overconfidence, 5: Externally provided anchoring, 6: Self-generated anchoring, 7: Cognitive ability, 8: Perceived complexity, 9: Satisficing, 10: Age, 11: Sex, 12: Income, 13: Years of university education, 14: Extraversion, 15: Agreeableness, 16: Conscientiousness, 17: Neuroticism, 18: Openness, 19: Biased probability judgment

Table 8.2.2: Correlation Matrix of Elicited Individual Factors in the Low Margin Setting

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.
1.	1.0	-0.1	-0.2	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.1	0.0	-0.1	0.3	-0.1	0.0	0.1	0.0	0.0
2.		1.0	0.0	-0.1	-0.1	-0.1	0.2	0.3	0.0	0.1	0.0	-0.3	0.2	0.0	0.2	-0.1	0.3	0.1	0.1
3.			1.0	-0.4	-0.2	0.1	-0.2	-0.1	0.0	0.0	-0.2	-0.1	0.1	-0.2	0.2	0.0	-0.2	0.1	-0.1
4.				1.0	0.0	-0.1	0.0	-0.1	0.2	0.2	0.1	0.1	0.1	0.1	-0.1	0.1	0.1	-0.2	-0.1
5.					1.0	0.2	-0.1	0.0	0.0	0.1	0.1	0.1	0.1	0.2	0.0	0.1	-0.1	0.2	-0.2
6.						1.0	0.2	0.1	0.0	0.1	0.2	0.1	0.0	0.0	0.1	0.2	0.0	0.0	0.5
7.							1.0	0.4	-0.1	0.0	0.4	0.0	0.0	0.2	0.0	0.1	0.1	0.0	0.0
8.								1.0	-0.1	0.0	0.3	-0.1	0.0	0.0	0.0	-0.2	0.1	0.1	0.1
9.									1.0	-0.1	-0.2	-0.2	-0.2	-0.1	0.1	-0.1	0.3	0.2	-0.1
10.										1.0	0.1	0.3	0.7	0.0	-0.1	0.3	-0.1	0.1	0.1
11.											1.0	0.2	0.2	0.3	-0.1	0.0	0.2	0.1	0.3
12.												1.0	0.3	0.1	-0.3	0.1	0.0	0.1	0.3
13.													1.0	-0.1	-0.1	0.2	0.0	0.0	0.1
14.														1.0	-0.2	0.0	0.0	0.1	-0.1
15.															1.0	0.0	0.0	0.0	0.1
16.																1.0	-0.3	0.1	0.1
17.																	1.0	0.2	0.2
18.																		1.0	0.0
19.																			1.0

Notes: 1: Risk attitude, 2: Regret aversion, 3: Loss aversion, 4: Overconfidence, 5: Externally provided anchoring, 6: Self-generated anchoring, 7: Cognitive ability, 8: Perceived complexity, 9: Satisficing, 10: Age, 11: Sex, 12: Income, 13: Years of university education, 14: Extraversion, 15: Agreeableness, 16: Conscientiousness, 17: Neuroticism, 18: Openness, 19: Biased probability judgment

8.3. Behavior in Inventory Rationing Games

8.3.1. Instructions

/The following texts apply to the setting with the wholesale price contract with equally sized retailers. For the revenue sharing setting, the text was modified accordingly. Similarly, both the text and the visual size of the retailers were modified accordingly for the treatment where the retailers had different size./

Warehouses

You are taking part in an economic experiment. You can earn a considerable amount of money by participating. The text below will tell you how the amount you earn will be determined. During the experiment there are no Euros but Experimental Currency (EC). After your earnings are calculated your earned points are converted at the following exchange rate:

800 EC = 1 Euro

For example, if you have 4680 EC at the end of the experiment, your payoff will be € 5.85 in the end (as 4680 = 5.85 * 800). Thus, fractions of 800 EC can be converted to the exact Cent.

Anonymity

You will never be asked to reveal your identity to anyone during the decision-making part of the experiment. Neither the experimenters nor the other subjects will be able to link you to any of your decisions. In order to keep your decisions private, please do not reveal your choices to any other participant.

No Talking / No Cell Phones Allowed

Please read this document carefully and do not talk to any other participant during the experiment. Similarly, do not use your mobile phone during the experiment. If there is anything that you don't understand, please raise your hand. An experimenter will approach you and clarify your questions in private.

Your Task During the Experiment

You are the owner of a warehouse that sells 'WIDGETS', a fictional type of good to two retailers. For the next 20 periods, you will see how many 'WIDGETS' these retailers order from your warehouse. The retailers sell these 'WIDGETS' to their customers for 15 EC each. Your task will be to allocate the number of 'WIDGETS' you have between the retailers in each period. You earn 4 EC for each 'WIDGET' you sell to any retailer independent whether the retailer manages to sell that 'WIDGET' to a customer or not. Your total earnings in the experiment will be equal to the total amount you earned during the 20 periods. Both of the other retailers are operated by fellow participants in this room.
A picture visualizes the situation below.
/A sample screenshot was provided here./
Click 'NEXT →' for further details.

Unfortunately, there may not be enough 'WIDGETS' to satisfy all the retailers' orders in the upcoming periods. You cannot send more than what a specific retailer ordered. Nevertheless, you know that each retailer will have between 0 and 300 customers in every period, the exact amount is not known to the retailers before they order and you also won't know that. Every number between 0 and 300 is equally likely, and they are independent from each other across periods. Each customer buys exactly one 'WIDGET' from the respective retailer. Each retailer

has 0 ‘WIDGETS’ at the beginning of every period. So, retailers cannot carry over any ‘WIDGETS’ they did not manage to sell in the previous period. Thus, any unsold ‘WIDGET’ becomes obsolete at the end of each period. Similarly, customers, who were not served in the previous period, cannot be satisfied in any latter period and their demand is lost for the retailer. Similarly, you cannot carry over ‘WIDGETS’ across periods and cannot satisfy the retailers’ prior orders in later periods.

Before you decide how to allocate the available amount of ‘WIDGETS’ between the retailers in each period, you will also know how many ‘WIDGETS’ the retailers ordered from your warehouse in that specific period. In addition, you will also see how much profit you and the retailers had in experimental currency (EC) during the previous period. A calculator will also be provided to you on each decision screen to assist you.

After each period, you will receive feedback on how you allocated your ‘WIDGETS’ as well as on your and both retailers’ profits in the current period.

Please, note that you will have to allocate the ‘WIDGETS’ between the retailers within a given time, which is 90 seconds for periods 1-5, 60 seconds for periods 6-20.

Click ‘←Back’ to go back.

Click ‘NEXT →’ to see how a decision screen will look like.

Below, you can see how a decision screen looks like.

/A sample decision screen was provided here./

Click ‘←Back’ to go back.

Click ‘NEXT →’ to see examples.

Example 1.

Suppose that you have 60 ‘WIDGETS’ at your disposal this turn. Retailer 1 ordered 60 ‘WIDGETS’, Retailer 2 ordered 30 ‘WIDGETS’. You decided to send 40 ‘WIDGETS’ to Retailer 1 and 20 ‘WIDGETS’ to Retailer 2. Therefore, you earn 240 EC this turn (60 ‘WIDGETS’ * 4 EC).

Click ‘←Back’ to go back.

Click ‘NEXT →’ to see another example.

Example 2.

Suppose that you have 100 ‘WIDGETS’ at your disposal this turn. Retailer 1 ordered 20 ‘WIDGETS’, Retailer 2 ordered 140 ‘WIDGETS’. You decided to send 20 ‘WIDGETS’ to

Retailer 1 and 80 ‘WIDGETS’ to Retailer 2. Therefore, you earn 400 EC this turn (100 ‘WIDGETS’ * 4 EC).
Click ‘←Back’ to go back.
Click ‘NEXT →’ to proceed.

Control Question 1.: You have 150 ‘WIDGETS’. Retailer 1 ordered 80 ‘WIDGETS’ from you, Retailer 2 ordered 120 ‘WIDGETS’ from you. How much do you earn if you send 50 ‘WIDGETS’ to Retailer 1 and 100 ‘WIDGETS’ to Retailer 2?
Control Question 2.: You have 120 ‘WIDGETS’. Retailer 1 ordered 80 ‘WIDGETS’ from you, Retailer 2 ordered 120 ‘WIDGETS’ from you. How much do you earn if you send 70 ‘WIDGETS’ to Retailer 1 and 50 ‘WIDGETS’ to Retailer 2?
Click ‘←Back’ to go back.
Click ‘NEXT →’ to answer and start the experiment.
Thank you for answering the control questions, the experiment will start soon!

Retailers

You are taking part in an economic experiment. You can earn a considerable amount of money by participating. The text below will tell you how the amount you earn will be determined. During the experiment there are no Euros but Experimental Currency (EC). After your earnings are calculated your earned points are converted at the following exchange rate:

800 EC = 1 Euro

For example, if you have 4680 EC at the end of the experiment, your payoff will be € 5.85 in the end (as 4680 = 5.85 * 800). Thus, fractions of 800 EC can be converted to the exact Cent.

Anonymity

You will never be asked to reveal your identity to anyone during the decision-making part of the experiment. Neither the experimenters nor the other subjects will be able to link you to any of your decisions. In order to keep your decisions private, please do not reveal your choices to any other participant.

No Talking / No Cell Phones Allowed

Please read this document carefully and do not talk to any other participant until the experiment is over. Similarly, do not use your mobile phone during the experiment. If there is anything that you don't understand, please raise your hand. An experimenter will approach you and clarify your questions in private.

Your Task During the Experiment

You are the owner of a retail store that sells 'WIDGETS', a fictional type of good to your customers. To be able to sell ‘WIDGETS’, you will have to buy them in advance from a warehouse. For the next 20 periods, you will specify how many ‘WIDGETS’ you order from this warehouse. Each ‘WIDGET’ you buy costs you 4 EC and you receive 15 EC for each ‘WIDGET’ you sell to your customers. Your total earnings in the experiment will be equal to the total amount you earned during the 20 periods.
There is another retailer who orders ‘WIDGETS’ from the same warehouse. This retailer serves a different market and has different customers than yours. For the next 20 periods, you will order from the same warehouse. This warehouse will serve you and the other retailer, which will also remain the same during the experiment. Both the warehouse and the other retailer are other fellow participants in this room.
A picture visualizes the situation below.
/A sample screenshot was provided here./
Click ‘NEXT →’ for further details.

You know that you will have between 0 and 300 customers in every period, the exact amount is not known before you order. Every number between 0 and 300 is equally likely, and they are independent from each other across periods. Each customer buys exactly one 'WIDGET' from you. Again, the other retailer does not compete with you and there is no connection between the number of his/her customers and yours.
You will have 0 ‘WIDGETS’ at the beginning of every period before you post your order to the warehouse. So, you cannot carry over any ‘WIDGETS’ you did not manage to sell in the previous period. Thus, any unsold ‘WIDGET’ becomes obsolete at the end of each period. Similarly, customers, who were not served in the previous period, cannot be satisfied in any latter period and their demand is lost.
After you posted your order to the warehouse, you and the other retailer will receive ‘WIDGETS’ from it, which cost 4 EC each. Then, you will receive information about how many customers you had in the current period. You will sell each customer 1 ‘WIDGET’

automatically and receive 15 EC for each 'WIDGET' you sold. In addition, you will receive feedback on your profit in the current period as well.
Please, note that you will have to order your 'WIDGETS' from the warehouse within a given time, which is 60 seconds for periods 1-5, 30 seconds for periods 6-20.
Click '←Back' to go back.
Click 'NEXT →' to see how a decision screen will look like.

Example 1.
Suppose that you ordered 120 'WIDGETS' this turn. The warehouse sent you 100 'WIDGETS'. Assume you had 50 customers in this period. Therefore, you earn 350 EC this turn (50 'WIDGETS' * 15 EC – 100 'WIDGETS' * 4 EC = 750 EC – 400 EC).
Click '←Back' to go back.
Click 'NEXT →' to see another example.

Example 2.
Suppose that you ordered 80 'WIDGETS' this turn. The warehouse sent you 80 'WIDGETS'. Assume you had 100 customers in this period. Therefore, you earn 880 EC this turn (80 'WIDGETS' * 15 EC – 80 'WIDGETS' * 4 EC = 1200 EC – 320 EC).
Click '←Back' to go back.
Click 'NEXT →' to proceed.

Control Question 1.: You ordered 150 'WIDGETS'. The warehouse sent you 50 'WIDGETS'. How much do you earn if you had 220 customers?
Control Question 2.: You ordered 200 'WIDGETS'. The warehouse sent you 200 'WIDGETS'. How much do you earn if you had 70 customers?
Click '←Back' to go back.
Click 'NEXT →' to answer and start the experiment.
Thank you for answering the control questions, the experiment will start soon!

8.3.2. Decision Screens

Figure 8.3.1: Decision Screen for Warehouses in ID Treatments

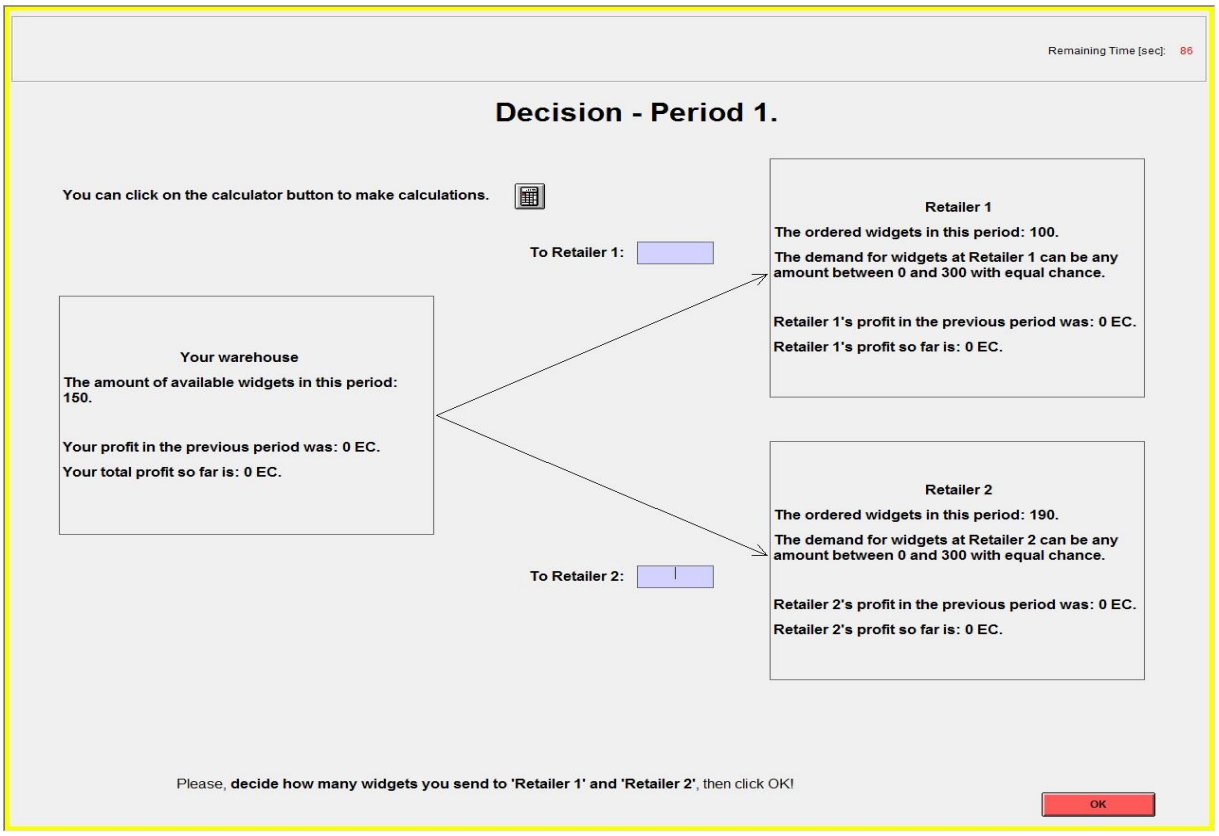
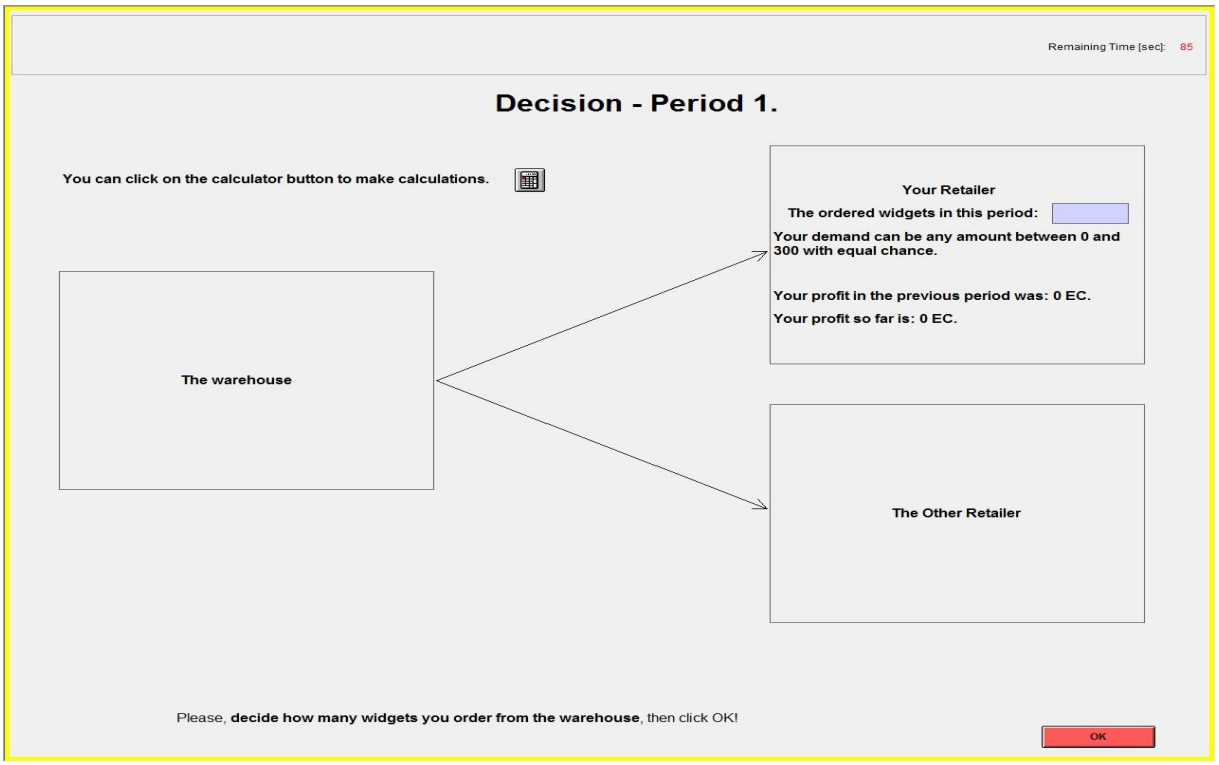


Figure 8.3.2: Decision Screen for Retailers in ID Treatments



Note that the decision screen and the size of the boxes were adjusted accordingly for the treatment with differently sized retailers.

8.3.3. Shortage Gaming

Figure 8.3.3: Average Order Quantities Over Time in DIFF-WP

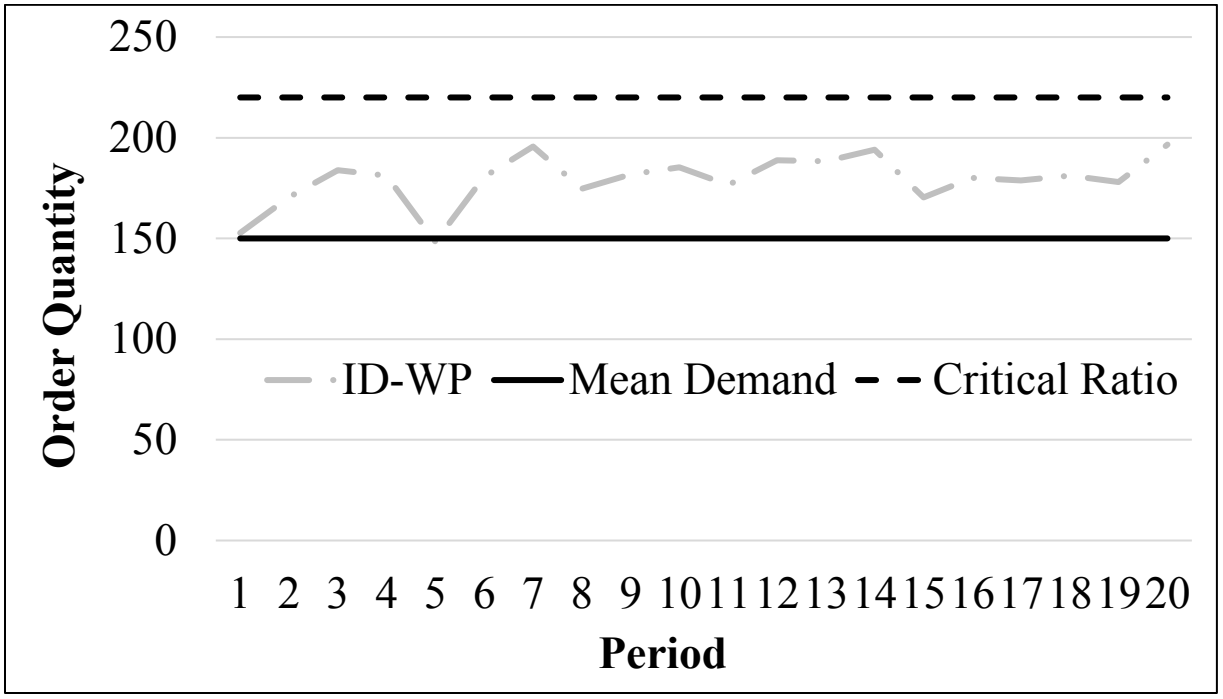


Figure 8.3.5: Average Order Quantities Over Time for the Small Retailer in DIFF-WP

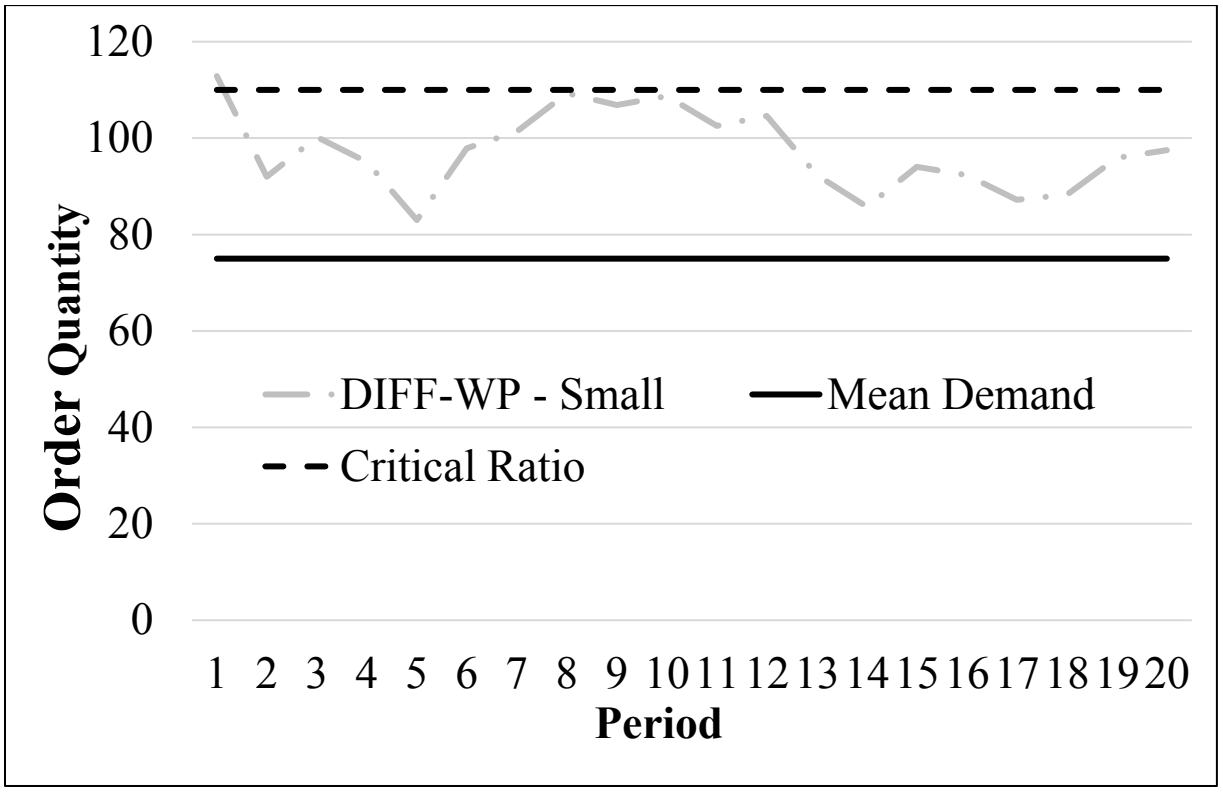


Figure 8.3.6: Average Order Quantities Over Time for the Large Retailer in DIFF-WP

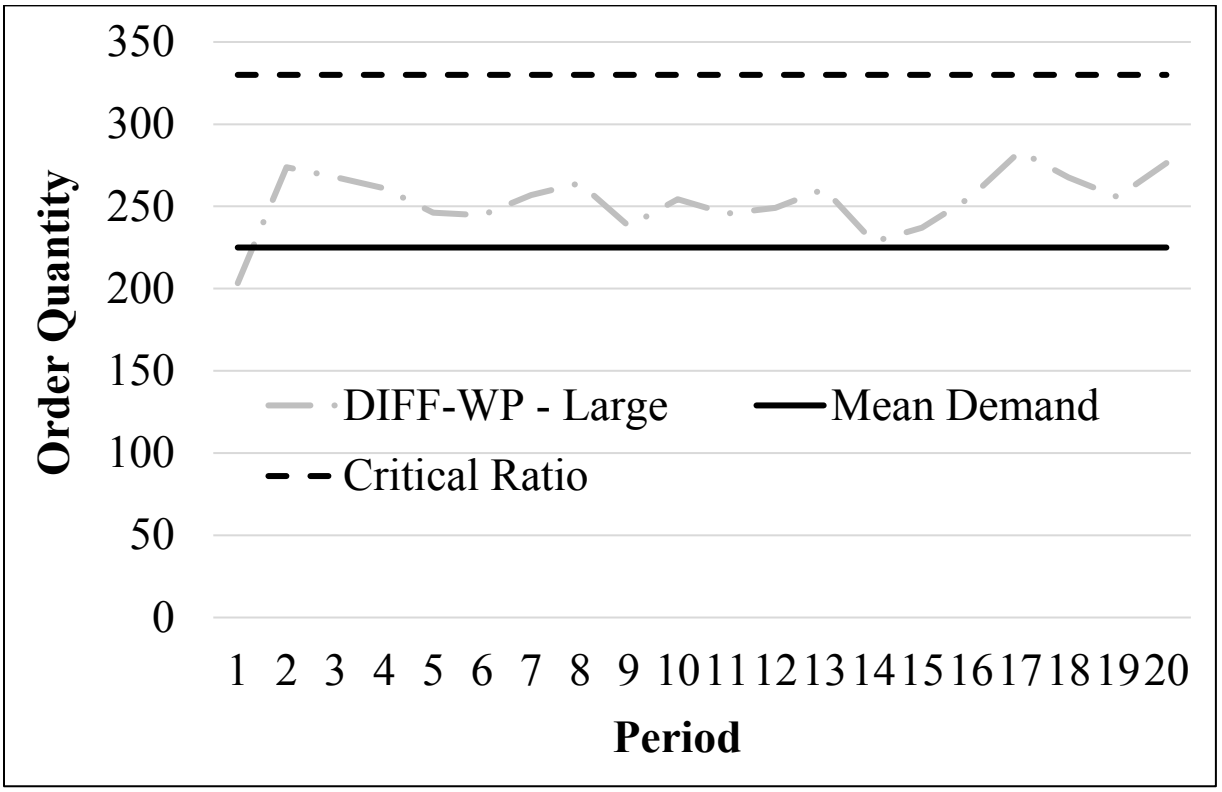
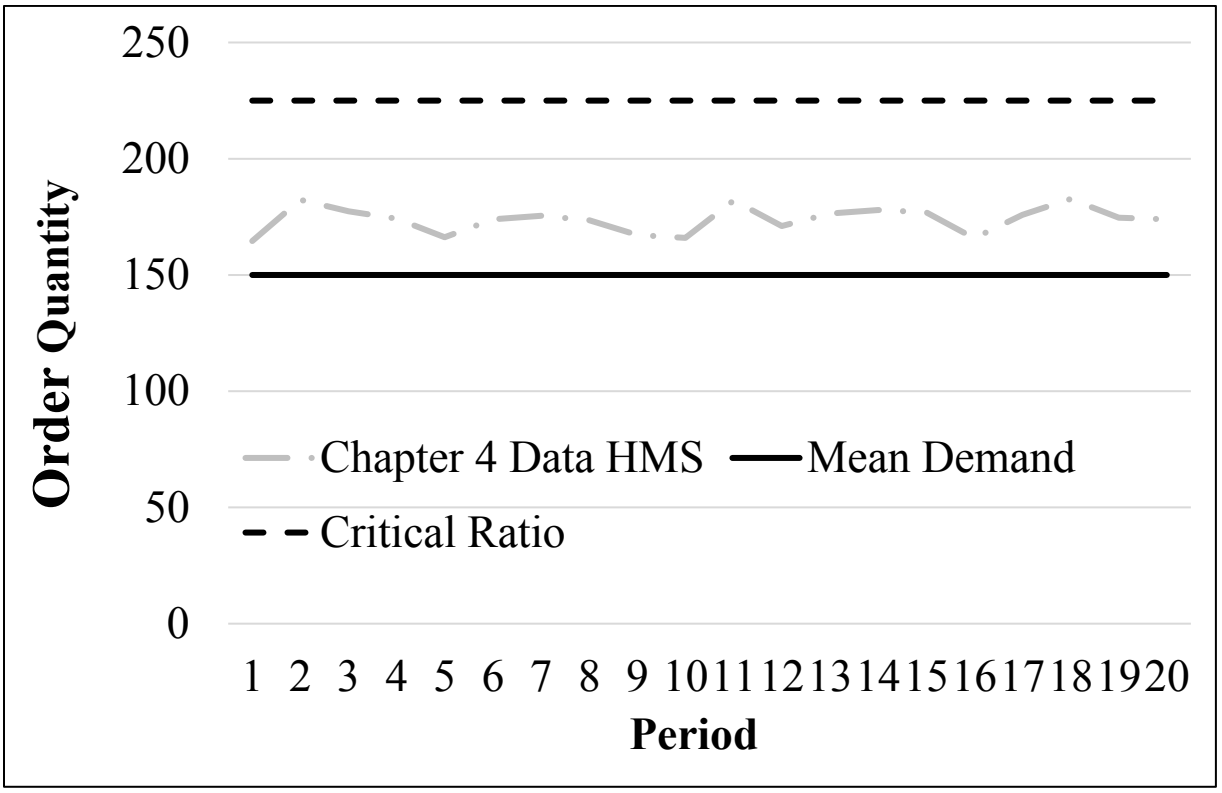


Figure 8.3.7: Average Order Quantities Over Time Without Rationing



The figures above show that the order quantities do not show shortage gaming in either of the treatments. The slightly increasing trend is similar in the data series from Chapter 4, which is explained by the imperfect learning other newsvendor studies documented. Consequently, the order quantities are between the mean demand and the optimal order quantity as in previous newsvendor studies.

8.4. Behavioral Decisions in Inventory Systems with Dual Sourcing Option

8.4.1. Instructions

/The following text was the instruction of treatment D7T. The instructions were adjusted for each treatment. Note that there was no Option 2 in treatment ST. In D10T, Option 2 cost 10 instead of 7.

Welcome to this experiment. You will be asked to make a series of choices that will affect your payment. Please pay close attention to the instructions, and do not hesitate to raise your hand in case you have any questions. Throughout the experiment, the different earnings options will be listed in € (EURO). In the end, you will receive the exact amount you earn in €. Also, you will be asked to fill out a short questionnaire, where your answers are not relevant for your earnings. All your decisions and answers during the experiment will stay completely anonymous to everyone. The experiment has two parts: 'Part 1' and 'Part 2'. Your final payment will then be the sum of your payments from 'Part 1' and 'Part 2'. Instructions for 'Part 1' will be provided on the next screens. Instructions for 'Part 2' will be provided after you finished 'Part 1'. For some of the tasks, it is possible to make losses. The experiment is designed in a way that your payments will be positive in the end. How high your payment is, depends on your choices throughout the experiment. If you have questions, please raise your hand at any time and an experimenter will provide assistance.

Part 1.

In 'Part 1', you will participate in an individual decision-making setting for 48 periods. You own a warehouse that will buy spare parts, a type of good your warehouse supplies to its customers. In each period, you will have to decide how many spare parts you want to order. The demand is random in each period, and you do not know how much it will be in advance. Thus, you will make your decision on how many spare parts to order before observing the exact demand. The demand will be a random number between 0 and 8 in each period. Every number is equally

likely, and the random numbers are independent from each other from period to period. Please, click NEXT → to continue reading further instructions.

On the top of the screen, you can see the visualization of the setting you will participate in. The big object in the middle of the screen symbolizes your warehouse. It shows how many units you have currently on stock or how many backorders you have. On the left side of the screen you can see your two suppliers you can order spare parts from and a box labelled as 'In Transit'. On the right side of the screen you can see your customers and a box labelled as 'Demand'. You have two options you can order from:
Option 1: The items you order from this supplier cost 6 each. This supplier can deliver with a delay of 1 period. So, the products you order from this supplier arrive in the next period. This supplier does not have a capacity constraint, so you can order as much as you want from it in each period.
Option 2: The items you order from this supplier cost 7 each. This supplier can deliver with a delay of 0 periods. So, the products you order from this supplier arrive in this current period. This supplier does not have a capacity constraint, so you can order as much as you want from it in each period. Please, click ← BACK to go back to the previous screen or NEXT → to continue reading further instructions.

In each of the 48 periods, the following happens:
1. You decide how much to order both from Option 1 and Option 2.
2. You receive the amount of spare parts that was 'In Transit'. So, these products are added to 'On stock'. You receive the amount of goods that you ordered from Option 2. So, these products are added to 'On stock'. In addition, Option 1 delivers the amount you decided to order from it. You will get this amount in the next period, so it moves to the 'In Transit' box.
3. You see how much the demand in that period (a randomly generated number between 0 and 8) was.
In each period, if you have equal or more spare parts than the demand was (after the spare parts 'In Transit' and the ones ordered from Option 2 had arrived), then the whole demand is satisfied. In this case, you might have remaining units under 'On stock' which is carried over to the next period. In each period, if you have less spare parts than the demand was (after the spare parts 'In Transit' and the ones ordered from Option 2 had arrived), then just part of the demand is satisfied. In this case, the demand was higher than what you could satisfy. Thus, you have the difference as 'Backorders', which are carried over to the next period and which will be satisfied

during the next period, if you have enough spare parts then. Note that it is not possible to have positive 'On stock' and positive 'Backorders' at the same time.
4. The computer records the costs you had at the end of the period: Remember, at the end of the period, each spare part 'On stock' costs you 2 and each 'Backorder' costs you 8. Moreover, each spare part ordered from Option 1 costs you 6 and each spare part ordered from Option 2 costs you 7.
Please, click ← BACK to go back to the previous screen or NEXT → to see an example.

The other decision-makers in this experiment participate in exactly the same individual decision-making setting as you do. Your and their costs will be recorded in each of the 48 periods. Your payment depends on the total cost you and they had in the 48 periods. The decision-maker with the lowest total cost will receive € 15.00. The participant with the highest total cost will earn € 0.00. The payment of the other decision-makers will be between € 0.00 and € 15.00 depending on their total cost relative to the decision-maker who has the lowest and highest total cost. Therefore, the less total cost you accumulate during the 48 periods, the higher your corresponding payment will be.
Suppose, the decision-maker is in the following initial situation:
There are currently 3 spare parts 'On stock' and 0 spare parts as 'Backorders'. In the next period 1 spare part arrives from Option 1, which you can see as 'In Transit' at the moment. The decision-maker knows that the demand during the next period will be a random number between 0 and 8. The decision-maker makes a decision and orders 4 spare parts from Option 1 and 1 spare part from Option 2. Please, click ← BACK to go back to the previous screen or NEXT → to see what happens in the next period.

The decision-maker observes the demand which is 7. The decision-maker had initially 3 spare parts 'On stock' and 0 'Backorders' (see previous screen). He received the 1 spare part which was 'In Transit' and the 1 spare part which he/she ordered from Option 2. So, the decision-maker has 5 spare parts (3+1+1), while the demand is 7. Thus, the decision-maker has 0 'On stock' and 2 'Backorders' (-2=5-7) after this period.
The decision-maker had a cost of 0 before this period.
The spare parts 'On stock' cost 0 this turn (0*2).
The 'Backorders' cost 16 this turn (2*8).
The spare parts ordered from Option 1 cost 24 this turn (4*6).
The spare parts ordered from Option 2 cost 7 this turn (1*7).

The decision-maker had a cost of 47 in this period (0+16+24+7).
The decision-maker has a total cost of 47 till now (0+47).
There are currently 0 spare parts 'On stock' and 2 spare parts as 'Backorders'. In the next period 4 spare parts arrive, which was ordered from Option 1 and which is 'In Transit' at the moment. The decision-maker knows that the demand during the next period will be a random number between 0 and 8. The decision-maker makes a decision and orders 7 spare parts from Option 1 and 2 spare parts from Option 2. Please, click ← BACK to go back to the previous screen or NEXT → to see what happens in the next period.

The decision-maker observes the demand which is 4. The decision-maker had initially 0 spare parts 'On stock' and 2 'Backorders' (see previous screen). He received the 4 spare parts which were 'In Transit' and the 2 spare parts which he/she ordered from Option 2. So, the decision-maker has 6 spare parts (4+2), while the demand is 4. In addition, there were 2 'Backorders' (4+2=6). Thus, the decision-maker has 0 'On stock' and 0 'Backorders' (0=6-6) after this period.
The decision-maker had a cost of 47 before this period.
The spare parts 'On stock' cost 0 this turn (0*2).
The 'Backorders' cost 0 this turn (0*8).
The spare parts ordered from Option 1 cost 42 this turn (7*6).
The spare parts ordered from Option 2 cost 14 this turn (2*7).
The decision-maker had a cost of 56 in this period (0+0+42+14).
The decision-maker has a total cost of 103 till now (47+56).
There are currently 0 spare parts 'On stock' and 0 spare parts as 'Backorders'. In the next period 7 spare parts arrive, which was ordered from Option 1 and which is 'In Transit' at the moment. The decision-maker knows that the demand during the next period will be a random number between 0 and 8. The decision-maker makes a decision and orders 5 spare parts from Option 1 and 1 spare part from Option 2. Please, click ← BACK to go back to the previous screen or NEXT → to see what happens in the next period.

The decision-maker observes the demand which is 0. The decision-maker had initially 0 spare parts 'On stock' and 0 'Backorders' (see previous screen). He received 7 spare parts which were 'In Transit' and 1 spare part which he/she ordered from Option 2. So, the decision-maker has 7 spare parts (0+7+1), while the demand is 0. Thus, the decision-maker has 8 'On stock' and 0 'Backorders' after this period.
The decision-maker had a cost of 103 before this period.

The spare parts 'On stock' cost 16 this turn (8*2).
The 'Backorders' cost 0 this turn (0*8).
The spare parts ordered from Option 1 cost 30 this turn (5*6).
The spare parts ordered from Option 2 cost 7 this turn (1*7).
The decision-maker had a cost of 53 in this period (16+0+30+7).
The decision-maker has a total cost of 156 till now (103+53).
There are currently 8 spare parts 'On stock' and 0 spare parts as 'Backorders'. In the next period 5 spare parts arrive, which was ordered from Option 1 and which is 'In Transit' at the moment. The decision-maker knows that the demand during the next period will be a random number between 0 and 8. The decision-maker makes a decision and orders 2 spare parts from Option 1 and 0 spare parts from Option 2 ... Please, click ← BACK to go back to the previous screen or OK to answer the control questions.

8.4.2. Control Questions

Consider the following situations at the beginning of a period:
There are 0 spare parts 'On stock', 7 spare parts as 'Backorders' and 8 'In Transit'. The decision-maker orders 5 spare parts from Option 1 and 2 spare parts from Option 2.
If the demand is 2, how many spare parts will be 'On stock' after this period?
What will be the cost of spare parts 'On stock' during this period?
If the demand is 2, how many spare parts will be as 'Backorders' after this period?
What will be the cost of 'Backorders' during this period?
What will be the cost of spare parts ordered from Option 1 in this period?
What will be the cost of spare parts ordered from Option 2 in this period?
What will be the total cost in this period?

8.4.3. Decision and Feedback

Please, decide how many spare parts to order from 'Option 1' and 'Option 2', then click OK!

Period 2
You ordered from Option 1:
Your cost of ordering from Option 1 in this period:
The amount that was In Transit you received:
You ordered from Option 2:
Your cost of ordering from Option 2 in this period:

The demand was:
You have spare parts 'On stock':
Your cost of having spare parts 'On stock' in this period:
You have 'Backorders':
Your cost of backordering spare parts in this period:
Your cost in this period is:
Your total cost till now is:

8.4.4. Cognitive Reflection Task

In 'Part 2', you will make decisions in four different tasks. One of the tasks will be randomly chosen to be relevant for your payment.

Part 2 Task I.

In Task I., you will answer three questions. You will see each question separately. Thus, you will see one question on each of the three upcoming screens. You will have 30 seconds for each question. If Task I. is chosen to be relevant for your payment, you will get €4.00 for each question answered correctly. Note that if you do not answer a question because you run out of time, then that will count as a wrong answer for that particular question.

A bat and a ball cost 1.10 € in total. The bat costs 1.00 € more than the ball. How much does the ball cost? (in cents)

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (in minutes)

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

8.4.5. Risk and Loss Attitude Elicitation Task

Part 2 Task II. & Task III.
A sample decision screen is provided above for Task II. and Task III. in 'Part 2'.

On each screen you see two columns ('LEFT' and 'RIGHT'). In each column you have eight rows with different payment possibilities. You have to decide which option you prefer in each row.

For example, on the screen provided above, in the first row you will have to choose between:

'LEFT': receive 12.00 EUROS with 100% probability

'RIGHT': receive 14.00 EUROS with 50.0% probability or 5.00 EUROS with 50.0% probability

In the sixth row, you would make a choice between:

'LEFT': receive 7.00 EUROS with 100% probability

'RIGHT': receive 14.00 EUROS with 50.0% probability or 5.00 EUROS with 50.0% probability

In the center of the screen you will find a number of radio buttons. You can only click one of those buttons on each screen. This button indicates at what row you want to switch from 'LEFT' to 'RIGHT'.

If Task II. or Task III. is selected to be relevant for your payment in 'Part 2', the computer will select one of the rows at random. In this row, your payment will depend on whether you have chosen 'LEFT' or 'RIGHT' in that specific row. Finally, the computer will randomize between the two possible outcomes based on the given probabilities if you chose a 50-50 situation.

Example I.

If you choose row 6 that means you prefer 'LEFT' for the first 5 rows, but you prefer 'RIGHT' for rows 6-8.

Please note that there is a "LOSE" in front of a number. That means that this result would be deducted from your payment, in case this was the decision that is relevant for your payment.

Example II.

If you choose row 1 - ALWAYS RIGHT that means you prefer 'RIGHT' in every row.

Please note that there is a "LOSE" in front of a number. That means that this result would be deducted from your payment, in case this was the decision that is relevant for your payment.

Example III.

If you choose row ALWAYS LEFT that means you prefer 'LEFT' in every row.

Please choose where you prefer to switch from Option 'LEFT' to Option 'RIGHT'.

8.4.6. Regret Aversion Elicitation Task

Part 2 Task IV.

Please select which of the lotteries on the bottom of the screen ('LEFT' or 'RIGHT') you prefer. If the computer selects this task to be relevant for your payment, it will simulate the lottery that you chose and the outcome will determine your payment.

You chose the lottery on the LEFT (RIGHT) SIDE. Suppose that the computer plays it out for you and you win € 4.00 (€ 12.00). If you had chosen the lottery on the RIGHT (LEFT) SIDE, you would have won € 8.00 (€16.00). On a scale from 1 to 10, how much do you regret that you chose the lottery on the LEFT SIDE in this case? YOUR ANSWER TO THIS QUESTION DOES NOT AFFECT THE OUTCOME OF THIS LOTTERY OR YOUR EARNINGS IN ANY MANNER.

8.4.7. Decision and Feedback Screenshots

Figure 8.4.1: Decision Screen for the Dual Sourcing Treatments

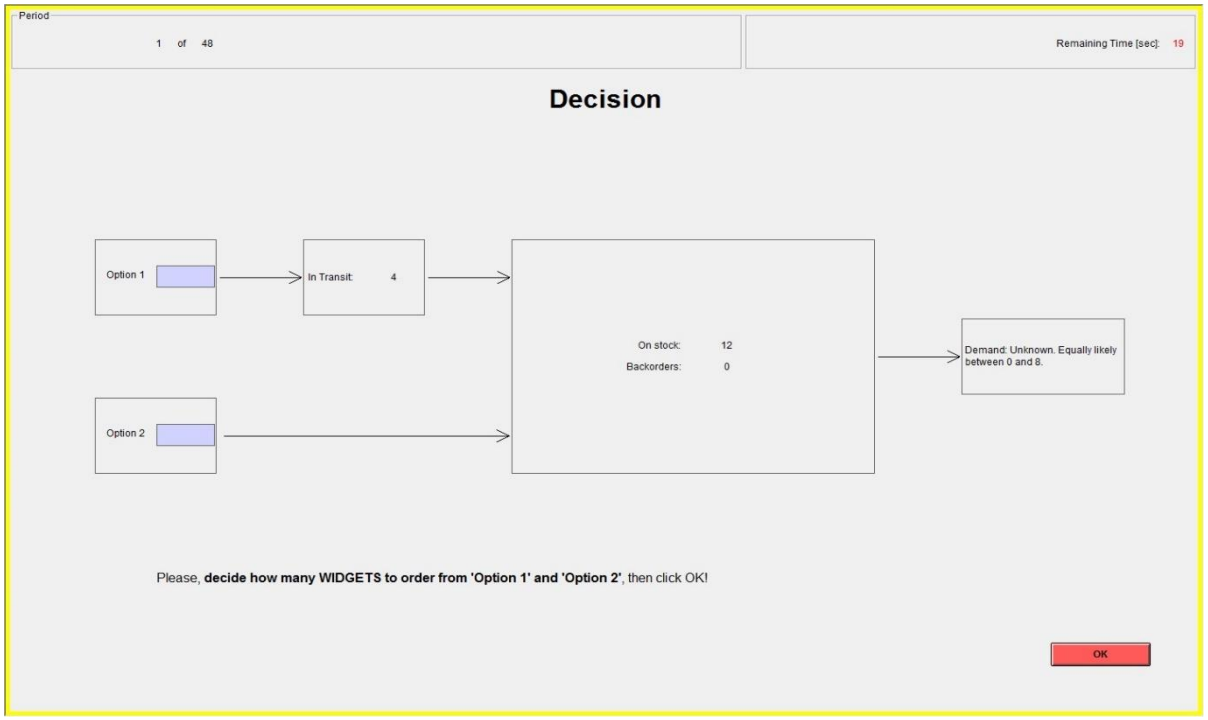


Figure 8.4.2: Feedback Screen for the Dual Sourcing Treatments

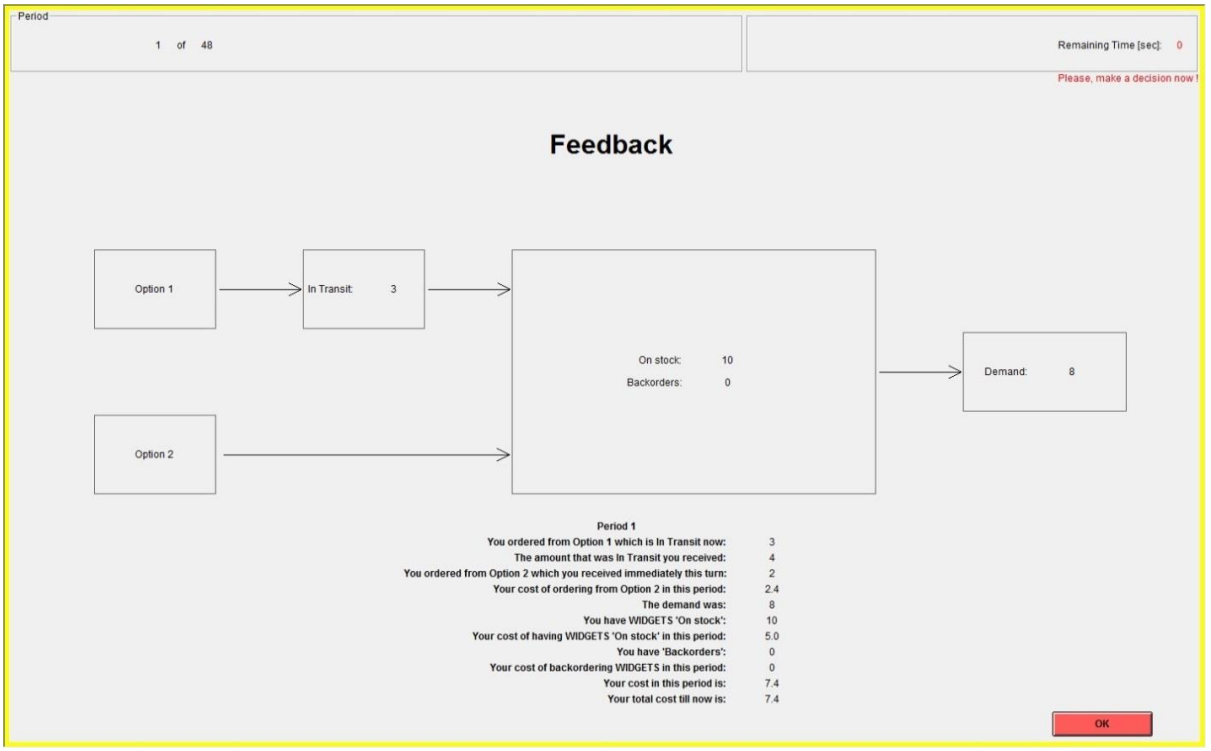


Figure 8.4.3: Decision Screen for the Single Sourcing Treatment

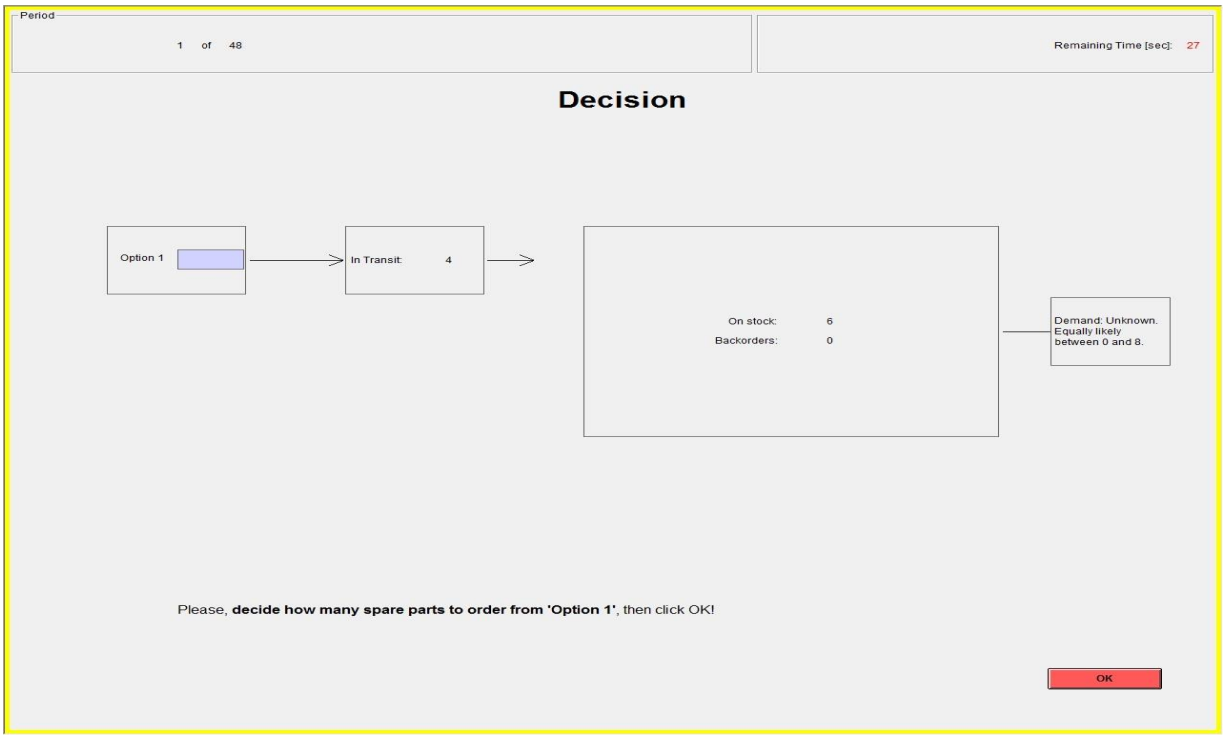


Figure 8.4.4: Feedback Screen for the Single Sourcing Treatment

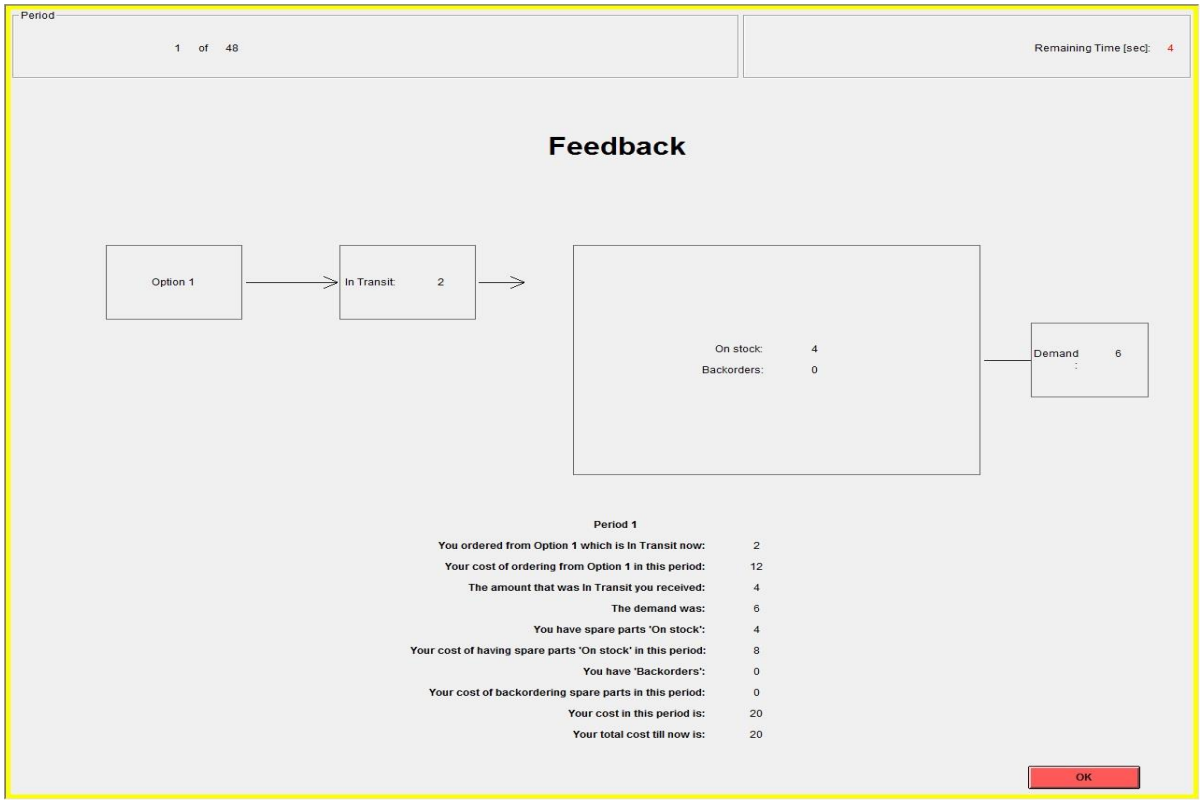


Figure 8.4.5: Elicitation of Risk Aversion

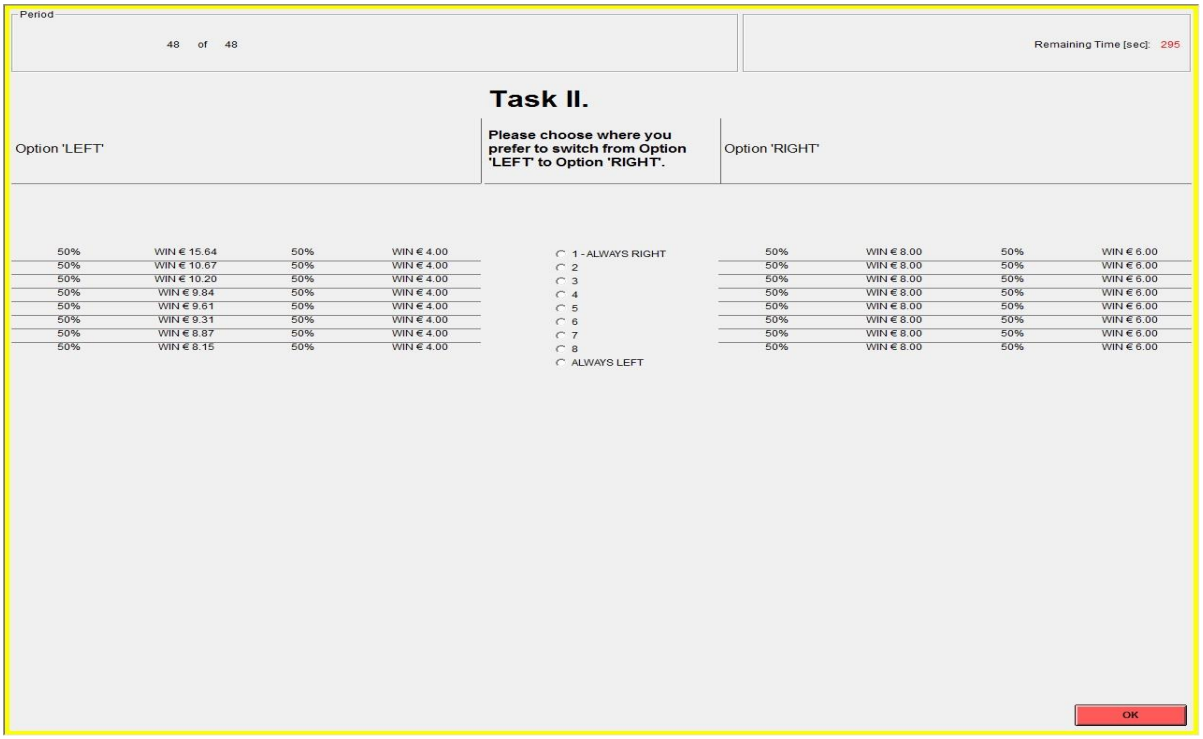


Figure 8.4.6: Elicitation of Loss Aversion

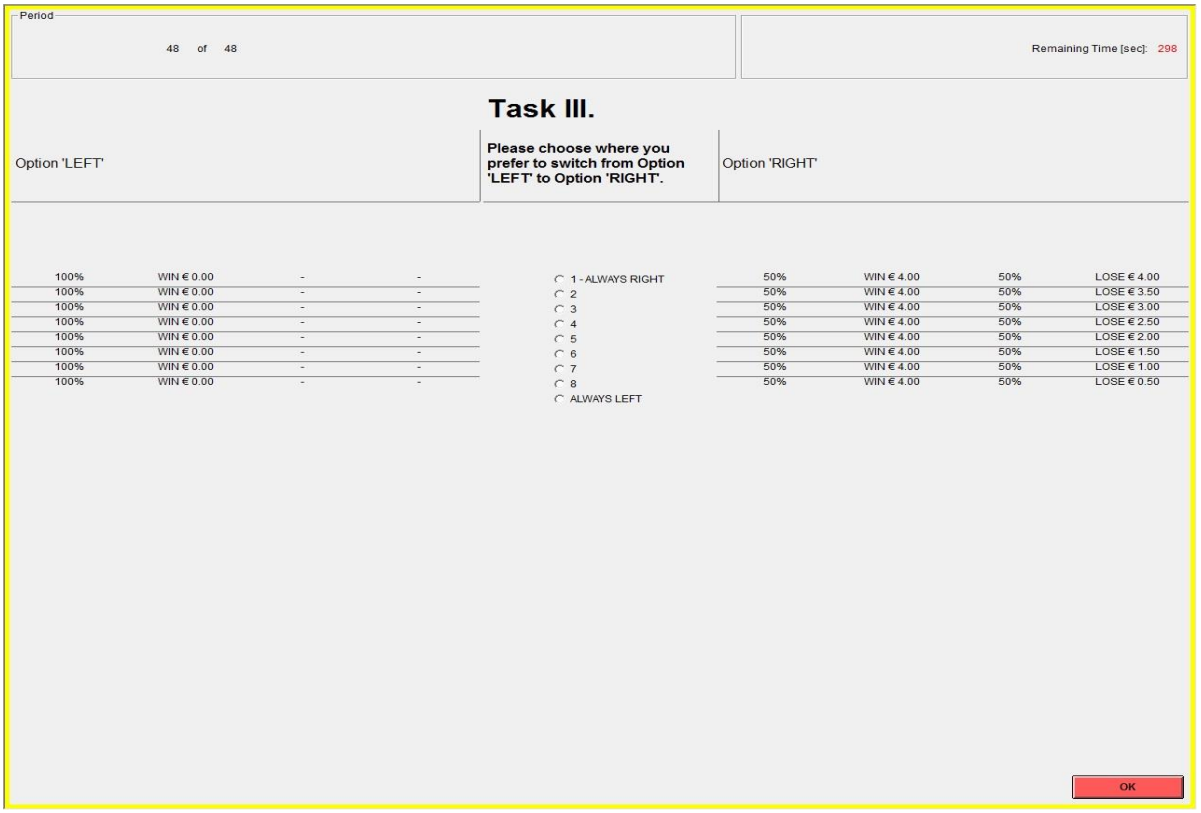
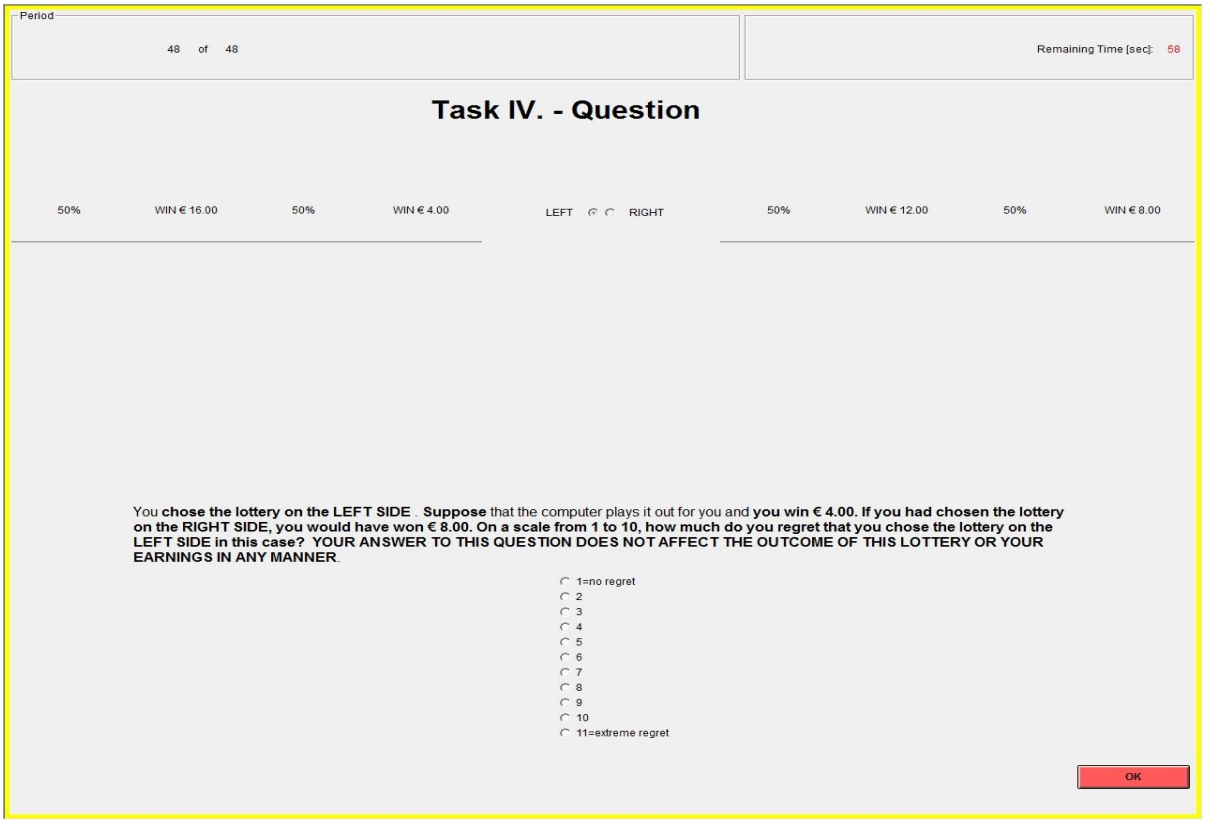


Figure 8.4.7: Lottery Choice

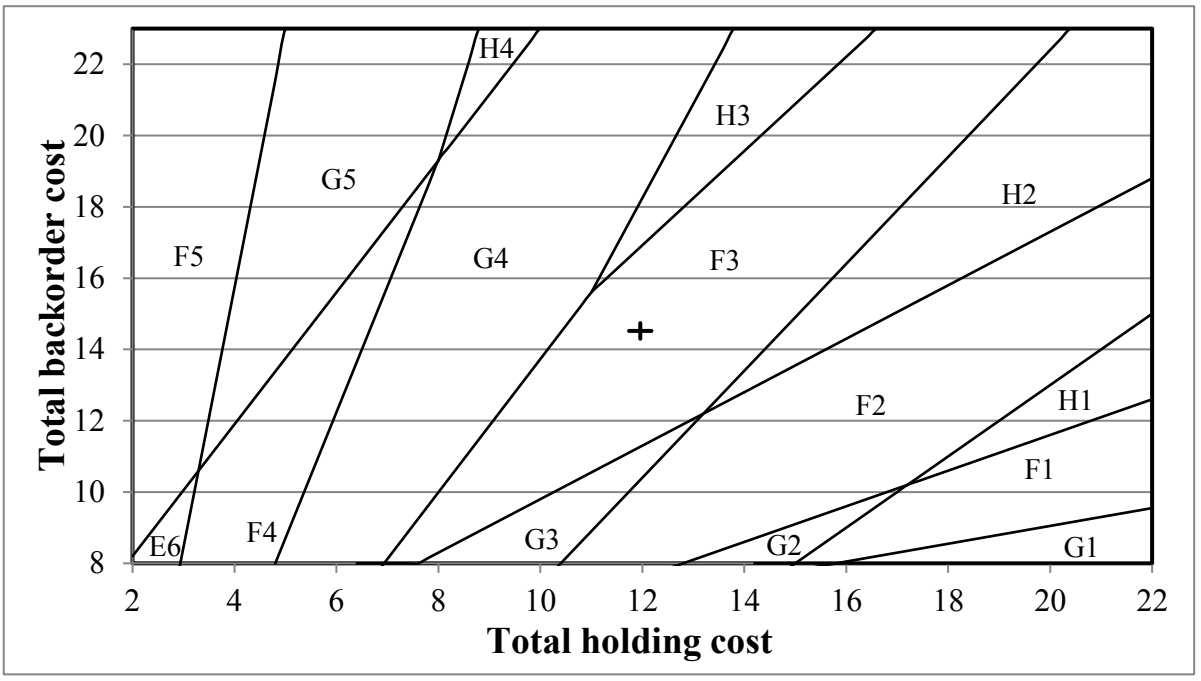
Figure 8.4.8: Elicitation of Regret Aversion



8.4.8. Validity of the Psychological Cost Model

Figure 8.4.9 shows the optimal usage ratios of the fast and slow supply mode depending on the total holding and backorder costs, taking various levels of psychological costs into account. This figure is consistent with H6.2 because it implies a higher usage ratio for the fast supply mode. We decided to vary the original h_{ps}=9.96 and b_{ps}=6.52 parameters reported by Ho et al. (2010) between 0% and 200%. Our objective in doing this was to see whether there are any major changes in the predicted behavior if we assume psychological costs that differ slightly from the original ones. We investigated whether a better fit could be achieved between Model 3 and the actual behavior if we relax our assumptions on the size of the psychological costs. We calibrated the same margin and cost parameters for our experimental setting as in Ho et al. (2010), but our dual sourcing environment was cognitively different from the newsvendor setting, which might imply that a better alignment could be achieved with different parameters.

Figure 8.4.9: Parametric Analysis of Psychological Costs for Treatment D7T

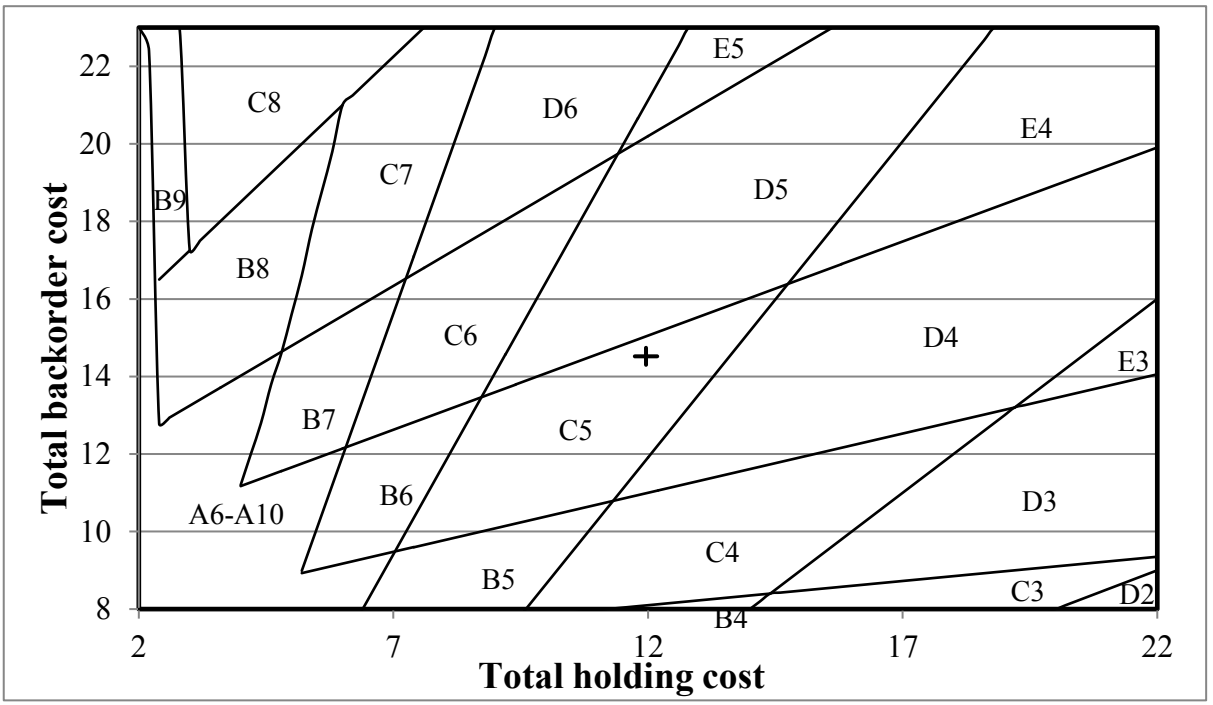


Notes: This figure shows the results of the parametric analysis of psychological costs. The cross shows the original values found in Ho et al. (2010).

The ratio of order quantities from the slow option divided by the overall orders and also the corresponding order-up-to parameters are relatively stable between wide ranges of psychological costs and increase/decrease significantly stepwise at particular thresholds. The letters differentiate regions where the following average usage proportion of the slow supply mode is optimal: E: 72.22%, F: 58.33%, G: 41.67%, H: 22.22%. The regular and emergency order-up-to levels and thus the optimal policies (z_S;z_F) are as follows in periods 1-44: E6: (10;6), F1: (4;1), F2: (5;2), F3: (6;3), F4: (7;4), F5: (8;5), G1: (4;2), G2: (5;3), G3: (6;4), G4: (7;5), G5: (8;6), H1: (4;3), H2: (5;4), H3: (6;5), H4: (7;6). As we already pointed out, policies are not stationary and differ somewhat at the very end – the last 4 periods – of our time horizon. Moreover, the same letter has different numbers, which means that, despite the different order-up-to parameters, these different policies result in the same split between the slow and the fast supply mode.

Figure 8.4.10 shows the extensive parametric analysis of the psychological costs for treatment D10T. We varied the psychological cost of leftovers and shortages similarly to Figure 8.4.9.

Figure 8.4.10: Parametric Analysis of Psychological Costs for Treatment D10T



Notes: This figure shows the results of the parametric analysis of psychological costs. The cross shows the original values found by Ho et al. (2010). Subjects' actual usage of the slow mode was 89.82%.

The letters differentiate regions where the following proportion of the slow supply mode is optimal: A: 100.00%, B: 97.22%, C: 91.67%, D: 83.33%, E: 72.22%. The regular and emergency order up to levels and thus the optimal policies (z_S;z_F) can be calculated with stochastic dynamic programming in line with Whittemore and Saunders (1977): A6: (9;1), A7: (10;2), A8: (11;3), A9: (12;4), A10: (13;5), B4: (7;0), B5: (8;1), B6: (9;2), B7: (10;3), B8: (11;4), B9: (12;5), C3: (6;0), C4: (7;1), C5: (8;2), C6: (9;3), C7: (10;4), C8: (11;5), D2: (5;0), D3: (6;1), D4: (7;2), D5: (8;3), D6: (9;4), E3: (6;2), E4: (7;3), E5: (8;4). As we have a finite horizon, policies are not stationary and are somewhat different at the very end – the last 5 periods – of our time horizon. Moreover, the same letter has different numbers, which means that, despite the different order up to parameters, these different policies result in the same split between the slow and the fast supply mode.

8.4.9. Further Robustness Checks

We also check if mean order quantities and their standard deviations decrease or increase over time: OLS regressions do not show any increasing or decreasing order pattern in any of the treatments at the 5% significance level. Our choice for the OLS is justified since there is no

significant (Ljung-Box Q-test, p>0.05) autocorrelation between the average order quantities in our time series with lags of 1 and 2.

We check whether our subjects base their decisions on their inventory position or on physical on-hand inventory. We minimize the sum of squared errors for each subject over 48 periods with the following formula:

$$\epsilon_{itF}^2 + \epsilon_{itS}^2 = (S_{iIP} - S_{it})^2 + (F_{iIP} - F_{it})^2 \quad (8.4.1)$$

where S_{iIP} (F_{iIP}) is the optimal inventory position for the slow (fast) supply mode. Notice that equation (8.4.1) quantifies the congruence to the particular policy that uses a specific inventory position which gives the lowest squared errors. A similar equation shows the fit to that specific policy based on the physical inventory level if we replace S_{iIP} by S_{iIL} and F_{iIP} by F_{iIL}. In general, the mean squared errors are significantly lower (t-test, p=0.02) for the equation with the inventory positions (MSE=6.94, R^2=0.71), implying that subjects base their decisions on the inventory position instead of the physical inventory level (MSE=10.89, R^2=0.45). Furthermore, 53 out of our 65 subjects have a lower squared error term for the model relying on the inventory position. Based on this evidence, we decide to look at the fit of models relying on the inventory position.

We report that Model 1 was not able to explain the actual inventory positions well at all. The actual inventory positions after ordering but before demand realization [ST: 6.96 (3.39), D7T: 6.26 (2.21) and D10T: 7.84 (1.83)] are significantly different from the predictions by Models 1, 2, 4 and 5 (W-tests: for all treatments: p<0.01). We see no significant difference to Model 3 (W-tests, ST: p>0.15, D7T: p>0.39, D10T: p>0.91), since the optimal inventory position after ordering but before demand realization changes from 11 (Model 1) to 8 (Model 3) in ST and D10T and from 10 to 6 in D7T. In addition, holding less inventory on purpose originates in the widely-documented ex post inventory error minimizing behavior (Kremer et al., 2014), which is regarded as a special case of the psychological cost model by Ho et al. (2010). This explanation seems even more convincing since robustness checks do not reveal any significant correlation (p>0.05) between the subjects' inventory level/position before demand realization and their elicited risk attitude, loss aversion parameter, regret preferences or any other socioeconomic characteristics. We note that there are no differences across groups with different cognitive ability when it comes to their inventory positions.

Next, we define the following loss functions for each of the above mentioned models (*M*) separately for each subject and period in line with Hansen et al. (2011):

$$L_{itM1} = |IP_{AitM} - IP_{PitM}| \tag{8.4.2}$$

$$L_{itM2} = |\frac{S_{AitM}}{F_{AitM}+S_{AitM}} - \frac{S_{PitM}}{F_{PitM}+S_{PitM}}| \tag{8.4.3}$$

Equation (8.4.2) quantifies the difference between the actual (IP_A) and predicted (IP_P) inventory positions. Equation (8.4.3) shows the difference between the actual and the predicted usage ratios of the slow option compared to all orders. In order to make comparisons between models, we calculate the differences between period- and subject-specific L_{itM1} and L_{itM2} values for each pair of models in line with Hansen et al. (2011). As we have 65 subjects altogether who participated in 48 periods, this means 3120 observations for each pair of models. A particular model is superior to another one if it has lower L_{itM1} and L_{itM2} values. Our results suggest that Model 3 has significantly (t-tests, $p<0.01$) lower values compared to Model 2, Model 4 and Model 5 This is summarized in Table 8.4.1. We conclude that Model 3 explains the observed behavior clearly better than the other models. Recall that, by definition, Model 1 provides predictions only on the average parameters. Therefore, it cannot be used for such period-specific comparisons. Yet, we note that the average predictions by Model 1 were very far away from the actual observations. As we explained before, Model 2 can be regarded as the period-specific approach relying on the same SDP approach as Model 1. All in all, Model 3 still gives the best fit if we define the loss function with quadratic terms instead of absolute differences or if we truncate the first ten and the last five periods from the analysis.

Table 8.4.1: Fit of Models Compared to Model 3

	M2	M4	M5		M2	M4	M5
Inventory position	1.36	1.76	1.67	Usage ratio	0.36	0.22	0.38
(ΔL_{itM1})	(5.01)	(2.41)	(3.97)	(ΔL_{itM2})	(0.42)	(0.11)	(0.45)

Notes: This table provides data on the averages (and the standard deviations) of differences between the individual- and period-specific loss function values that correspond to the different models compared to Model 3. M2 denotes the reoptimization approach. M4 (M5) is the myopic approach taking only the current (and the next) period(s) into account. All differences are significant at $p<0.01$.

We conduct a final test to see whether subjects hold less inventory since they do not recognize that it is optimal to order more units similarly to the myopic logic (Models 4 and 5) and therefore they keep a larger inventory position than the potential maximum demand which

is 8 in each period. There is not a single person who does not have an inventory position above 8 at least for multiple instances before the demand is revealed. We suggest that this is sufficient evidence to rule out the naive explanation relying on lack of knowledge or myopic optimization.

Chapter 9

9. Bibliography

Abdellaoui, M., Driouchi, A., & L'Haridon, O. (2011). Risk aversion elicitation: Reconciling tractability and bias minimization. *Theory and Decision*, *71*(1), 63-80.

Abdi, H. (2007). Bonferroni and Sidak corrections for multiple comparisons. In N. J. Salkind (Ed.), *Encyclopedia of Measurement and Statistics*. Thousand Oaks, CA, USA: Sage, 103-107.

Allon, G., & Van Mieghem, J. A. (2010a). Global dual sourcing: Tailored base-surge allocation to near- and offshore production. *Management Science*, *56*(1), 110-124.

Allon, G., & Van Mieghem, J. A. (2010b). The Mexico-China sourcing game: Teaching global dual sourcing. *INFORMS Transactions on Education*, *10*(3), 105-112.

Almlund, M., Duckworth, A. L., Heckman, J. J., & Kautz, T. (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, & L. Wößmann (Eds.), *Handbook of the Economics of Education*. Amsterdam, The Netherlands: Elsevier, 1-181.

Andersen, S., Fountain, J., Harrison, G. W., Hole, A. R., & Rutström, E. E. (2012). Inferring beliefs as subjectively imprecise probabilities. *Theory and Decision*, *73*(1), 161-184.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2006). Elicitation using multiple price list formats. *Experimental Economics*, *9*(4), 383-405.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2008a). Eliciting risk and time preferences. *Econometrica*, *76*(3), 583-618.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutström, E. E. (2008b). Lost in state space: Are preferences stable? *International Economic Review*, *49*(3), 1091-1112.

Anderson, L. R., & Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics*, *27*(5), 1260-1274.

Anderson, L. R., & Mellor, J. M. (2009). Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure. *Journal of Risk and Uncertainty*, *39*(2), 137-160.

Andersson, O., Holm, H. J., Tyran, J.-R., & Wengström, E. (2016). Risk aversion relates to cognitive ability: Preference or noise? *Journal of the European Economic Association*, *14*(5), 1129-1154.

Andreoni, J., & Harbaugh, W. T. (2010). Unexpected utility: Experimental tests of five key questions about preferences over risk. *Working paper, University of Oregon, Eugene, OR, USA*.

Andreoni, J., Kuhn, M. A., & Sprenger, C. (2015). Measuring time preferences: A comparison of experimental methods. *Journal of Economic Behavior & Organization*, *116*(1), 451-464.

Attema, A., & Brouwer, W. (2013). In search of a preferred preference elicitation method: A test of the internal consistency of choice and matching tasks. *Journal of Economic Psychology*, *39*(1), 126-140.

Bansal, S., & Moritz, B. B. (2015). Perceived versus actual value of production substitution flexibility: An experimental investigation. *Journal of Operations Management*, *38*(1), 56-70.

Bardsley, N. (2010). *Experimental Economics: Rethinking the Rules*. Princeton, NJ, USA: Princeton University Press.

Becker-Peth, M., Thonemann, U. W., & Gully, T. (2016). A note on the risk aversion of informed newsvendors. *Working paper, University of Cologne, Cologne, Germany*.

Bendoly, E., Croson, R., Goncalves, P., & Schultz, K. (2010). Bodies of knowledge for research in behavioral operations. *Production and Operations Management*, *19*(4), 434-452.

Bendoly, E., Donohue, K., & 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.

Berg, J., Dickhaut, J., & McCabe, K. (2005). Risk preference instability across institutions: A dilemma. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(11), 4209-4214.

Binswanger, H. P. (1980). Attitudes toward risk: Experimental measurement in rural India. *American Journal of Agricultural Economics*, *62*(3), 395-407.

Blais, A. R., & Weber, E. U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, *1*(1), 33-47.

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H. T. (2010). Belief elicitation in experiments: Is there a hedging problem? *Experimental Economics*, *13*(4), 412-438.

Bleichrodt, H. (2002). A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Economics*, *11*(5), 447-456.

Bleichrodt, H., Pinto, J. L., & Wakker, P. P. (2001). Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Management Science*, *47*(11), 1498-1514.

Bocqueho, G., Jacquet, F., & Reynaud, A. (2014). Expected utility or prospect theory maximisers? Assessing farmers' risk behaviour from field-experiment data. *European Review of Agricultural Economics*, *41*(1), 135-172.

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. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, *90*(1), 166-193.

Bolton, G. E., Ockenfels, A., & Thonemann, U. W. (2012). Managers and students as newsvendors. *Management Science*, *58*(12), 2225-2233.

Bordley, R., & LiCalzi, M. (2000). Decision analysis using targets instead of utility functions. *Decisions in Economics and Finance*, *23*(1), 53-74.

Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, *43*(4), 972-1059.

Bostian, A. J. 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.

Boute, R. N., & Van Mieghem, J. A. (2015). Global dual sourcing and order smoothing: The impact of capacity and lead times. *Management Science*, *61*(9), 2080-2099.

Brock, J. M., Lange, A., & Ozbay, E. Y. (2013). Dictating the risk: Experimental evidence on giving in risky environments. *American Economic Review*, *103*(1), 415-437.

Bruner, D. M. (2009). Changing the probability versus changing the reward. *Experimental Economics*, *12*(4), 367-385.

Cachon, G. P., & Fisher, M. (2000). Supply chain inventory management and the value of shared information. *Management Science*, *46*(8), 1032-1048.

Cachon, G. P., & Lariviere, M. A. (1999a). Capacity allocation using past sales: When to turn-and-earn. *Management Science*, *45*(5), 685-703.

Cachon, G. P., & Lariviere, M. A. (1999b). An equilibrium analysis of linear, proportional and uniform allocation of scarce capacity. *IIE Transactions*, *31*(9), 835-849.

Cachon, G. P., & Lariviere, M. A. (2005). Supply chain coordination with revenue-sharing contracts: Strengths and limitations. *Management Science*, *51*(1), 30-44.

Cachon, G. P., Randall, T., & Schmidt, G. M. (2007). In search of the bullwhip effect. *Manufacturing & Service Operations Management*, *9*(4), 457-479.

Camerer, C. F., & Ho, T.-H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, *8*(2), 167-196.

Cappelen, A. W., Nielsen, U. H., Sørensen, E. O., Tungodden, B., & Tyran, J.-R. (2013). Give and take in dictator games. *Economics Letters*, *118*(2), 280-283.

Chamberlain, E. H. (1948). An experimental imperfect market. *Journal of the Political Economy*, *56*(2), 95-108.

Charness, G., & Gneezy, U. (2010). Portfolio choice and risk attitudes: An experiment. *Economic Inquiry*, *48*(1), 133-146.

Charness, G., Gneezy, U., & Imas, A. (2013). Experiential methods: Eliciting risk preferences. *Journal of Economic Behavior and Organization*, *87*(1), 43-51.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, *117*(3), 817-869.

Charness, G., & Viceisza, A. (2016). Three risk-elicitation methods in the field: Evidence from rural Senegal. *Review of Behavioral Economics*, *3*(2), 145-171.

Chen, F. R., Li, J. B., & Zhang, H. Q. (2013). Managing downstream competition via capacity allocation. *Production and Operations Management*, *22*(2), 426-446.

Chen, L., & Davis, A. M. (2014). Validating models for bounded rationality and decision biases in a newsvendor experiment. *Working paper, The Fuqua School of Business, Duke University, Durham, NC, USA*.

Chen, Y., Su, X., & Zhao, X. (2012). Modeling bounded rationality in capacity allocation games with the quantal response equilibrium. *Management Science*, *58*(10), 1952-1962.

Chen, Y., & Zhao, X. (2015). Decision bias in capacity allocation games with uncertain demand. *Production and Operations Management*, *24*(4), 634-646.

Chiappori, P. A., & Paiella, M. (2011). Relative risk aversion is constant: Evidence from panel data. *Journal of the European Economic Association*, *9*(6), 1021-1052.

Cho, S.-H., & Tang, C. S. (2014). Technical note - Capacity allocation under retail competition: Uniform and competitive allocations. *Operations Research*, *62*(1), 72-80.

Choi, S., Kariv, S., Müller, W., & Silverman, D. (2014). Who is (more) rational? *American Economic Review*, *104*(6), 1518-1550.

Choi, T.-M. (2012). *Handbook of Newsvendor Problems: Models, Extensions and Applications*. New York, NY, USA: Springer.

Cohen, M., Jaffray, J. Y., & Said, T. (1987). Experimental comparison of individual behavior under risk and under uncertainty for gains and for losses. *Organizational Behavior and Human Decision Processes*, *39*(1), 1-22.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual*. Odessa, FL, USA: Psychological Assessment Resources, Inc.

Cox, J. C., Sadiraj, V., & Schmidt, U. (2015). Paradoxes and mechanisms for choices under risk. *Experimental Economics*, *18*(2), 215-250.

Crosetto, P., & Filippin, A. (2013). The "bomb" risk elicitation task. *Journal of Risk and Uncertainty*, *47*(1), 31-65.

Crosetto, P., & Filippin, A. (2016). A theoretical and experimental appraisal of four risk elicitation methods. *Experimental Economics*, *19*(3), 613-641.

Croson, R., Donohue, K., Katok, E., & Sterman, J. (2014). Order stability in supply chains: Coordination risk and the role of coordination stock. *Production and Operations Management*, *23*(2), 176-196.

Croson, R., & Sundali, J. (2005). The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of Risk and Uncertainty*, *30*(3), 195-209.

Cubitt, R. P., Starmer, C., & Sugden, R. (1998). On the validity of the random lottery incentive system. *Experimental Economics*, *1*(2), 115-131.

Cui, T. H., Raju, J. S., & Zhang, Z. J. (2007). Fairness and channel coordination. *Management Science*, *53*(8), 1303-1314.

Cui, T. H., & Zhang, Y. (2016). Cognitive hierarchy in capacity allocation games. *Working paper, University of Minnesota, Minneapolis, MN, USA*.

Csermely, T., & Minner, S. (2015). Behavioral decisions in inventory systems with dual sourcing option: An experiment. *Working paper, University of Vienna, Vienna, Austria*.

Csermely, T., & Minner, S. (2016). Newsvendor demand chasing and the gambler's and hot hand fallacies. *Working paper, University of Vienna, Vienna, Austria*.

Csermely, T., & Rabas, A. (2016). How to reveal people's preferences? A comprehensive comparison of multiple price list risk elicitation methods. *Journal of Risk and Uncertainty, 53*(2), in press.

Dave, C., Eckel, C. C., Johnson, C. A., & Rojas, C. (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty*, *41*(3), 219-243.

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.

Dohmen, T., Falk, A., Huffman, D., Marklein, F., & Sunde, U. (2009). Biased probability judgment: Evidence of incidence and relationship to economic outcomes from a representative sample. *Journal of Economic Behavior & Organization*, *72*(3), 903-915.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, *100*(3), 1238-1260.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522-550.

Drichoutis, A., & Lusk, J. (2012). Risk preference elicitation without the confounding effect of probability weighting. *Working paper, Munich Personal RePEc Archive, Munich, Germany*.

Dulleck, U., Fell, J., & Fooken, J. (2015). Within-subject intra- and inter-method consistency of two experimental risk attitude elicitation methods. *German Economic Review*, *16*(1), 104-121.

Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, *23*(4), 281-295.

Eckel, C. C., & Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, *68*(1), 1-17.

Eeckhoudt, L., Gollier, C., & Schlesinger, H. (1995). The risk-averse (and prudent) newsboy. *Management Science*, *41*(5), 786-794.

Eichberger, J., Güth, W., & Müller, W. (2003). Attitudes towards risk: An experiment. *Metroeconomica*, *54*(1), 89-124.

Engelbrecht-Wiggans, R., & Katok, E. (2008). Regret and feedback information in first-price sealed-bid auctions. *Management Science*, *54*(4), 808-819.

Engelbrecht-Wiggans, R., & Katok, E. (2009). A direct test of risk aversion and regret in first price sealed-bid auctions. *Decision Analysis*, *6*(2), 75-86.

Epley, N., & Gilovich, T. (2005). When effortful thinking influences judgmental anchoring: Differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making*, *18*(3), 199-212.

Eppen, G., & Schrage, L. (1981). Centralized ordering policies in a multi-warehouse system with lead times and random demand. *Multi-level Production/Inventory Control Systems: Theory and Practice*. Amsterdam, The Netherlands: North-Holland, 51-67.

Farquhar, P. H. (1984). State of the art - utility assessment methods. *Management Science*, *30*(11), 1283-1300.

Fausti, S., & Gillespie, J. (2000). A comparative analysis of risk preference elicitation procedures using mail survey results. *2000 Annual Meeting of Western Agricultural Economics Association*. Vancouver, BC, Canada.

Fehr, E., Klein, A., & Schmidt, K. M. (2007). Fairness and contract design. *Econometrica*, *75*(1), 121-154.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*(3), 817-868.

Fellner, G., & Maciejovsky, B. (2007). Risk attitude and market behavior: Evidence from experimental asset markets. *Journal of Economic Psychology*, *28*(3), 338-350.

Fershtman, C., & Gneezy, U. (2001). Discrimination in a segmented society: An experimental approach. *Quarterly Journal of Economics*, *116*(1), 351-377.

Fischbacher, U. (2007). Z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171-178.

Francis, G. (2013). *Multivariate Statistics.* Frenchs Forest, Australia: Pearson.

Fréchette, G. (2016). Experimental economics across subject populations. In J. H. Kagel, & A. E. Roth (Eds.), *The Handbook of Experimental Economics*. Princeton, NJ, USA: Princeton University Press, 435-480.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25-42.

Friedman, D., Cassar, A., & Selten, R. (2004). *Economics Lab: An Intensive Course in Experimental Economics*. London, UK; New York, NY, USA: Routledge.

Gächter, S., Johnson, E. J., & Herrmann, A. (2010). Individual-level loss aversion in riskless and risky choices. *Working paper, Institute for the Study of Labor (IZA), Nottingham, UK*.

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.

Gavirneni, S., & Robinson, L. W. (2015). Risk aversion and implicit shortage cost explain the anchoring and insufficient adjustment bias in human newsvendors. *Working paper, Cornell University, Samuel Curtis Johnson Graduate School of Management, Ithaca, NY, USA*.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, *17*(3), 295-314.

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, *112*(2), 631-645.

Gneezy, U., Rockenbach, B., & Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization*, *93*(1), 293-300.

Goeree, J. K., Holt, C. A., & Palfrey, T. R. (2003). Risk averse behavior in generalized matching pennies games. *Games and Economic Behavior*, *45*(1), 97-113.

Goldschmidt, K., Kremer, M., Thomas, D. J., & Craighead, C. W. (2014). Supply base diversification in the presence of high impact, low probability supply disruptions. *Working paper, University of St. Thomas, Minneapolis, MN, USA*.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, *38*(1), 129-66.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114-125.

Gurnani, H., Ramachandran, K., Ray, S., & Xia, Y. (2014). Ordering behavior under supply risk: An experimental investigation. *Manufacturing & Service Operations Management*, *16*(1), 61-75.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367-388.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453-497.

Harrison, G. W., Humphrey, S. J., & Verschoor, A. (2010). Choice under uncertainty: Evidence from Ethiopia, India and Uganda. *Economic Journal*, *120*(543), 80-104.

Harrison, G. W., Johnson, E., McInnes, M. M., & Rutström, E. E. (2005). Temporal stability of estimates of risk aversion. *Applied Financial Economics Letters*, *1*(1), 31-35.

Harrison, G. W., Lau, M. I., & Rutström, E. E. (2009). Risk attitudes, randomization to treatment, and self-selection into experiments. *Journal of Economic Behavior & Organization*, *70*(3), 498-507.

Harrison, G. W., Lau, M. I., & Williams, M. B. (2002). Estimating individual discount rates in Denmark: A field experiment. *American Economic Review*, *92*(5), 1606-1617.

Harrison, G. W., List, J. A., & Towe, C. (2007). Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion. *Econometrica*, *75*(2), 433-458.

Harrison, G. W., & Rutström, E. E. (2008). Risk aversion in the laboratory. In J. C. Cox, & G. W. Harrison (Eds.), *Risk Aversion in Experiments*. Bingley, UK: Emerald Group Publishing Ltd., 41-196.

Hershey, J. C., Kunreuther, H. C., & Schoemaker, P. J. H. (1982). Sources of bias in assessment procedures for utility functions. *Management Science*, *28*(8), 936-954.

Herweg, F. (2013). The expectation-based loss-averse newsvendor. *Economics Letters*, *120*(3), 429-432.

Hey, J. D., Morone, A., & Schmidt, U. (2009). Noise and bias in eliciting preferences. *Journal of Risk and Uncertainty*, *39*(3), 213-235.

Ho, T.-H., Lim, N., & Cui, T. H. (2010). Reference dependence in multilocation newsvendor models: A structural analysis. *Management Science*, *56*(11), 1891-1910.

Ho, T.-H., & Su, X. (2009). Peer-induced fairness in games. *American Economic Review*, *99*(5), 2022-2049.

Ho, T.-H., Su, X., & Wu, Y. (2014). Distributional and peer-induced fairness in supply chain contract design. *Production and Operations Management*, *23*(2), 161-175.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644-1655.

Holt, C. A., & Laury, S. K. (2005). Risk aversion and incentive effects: New data without order effects. *American Economic Review*, *95*(3), 902-904.

Houser, D., Vetter, S., & Winter, J. (2012). Fairness and cheating. *European Economic Review*, *56*(8), 1645-1655.

Huck, S., & Müller, W. (2012). Allais for all: Revisiting the paradox in a large representative sample. *Journal of Risk and Uncertainty*, *44*(3), 261-293.

Jacobson, S., & Petrie, R. (2009). Learning from mistakes: What do inconsistent choices over risk tell us? *Journal of Risk and Uncertainty*, *38*(2), 143-158.

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, *21*(11), 1161-1166.

Jain, K., Girotra, K., & Netessine, S. (2014). Managing global sourcing: Inventory performance. *Management Science*, *60*(5), 1202-1222.

Janakiraman, G., Seshadri, S., & Sheopuri, A. (2015). Analysis of tailored base-surge policies in dual sourcing inventory systems. *Management Science*, *61*(7), 1547-1561.

Kagel, J. H., & Levin, D. (1986). The winner's curse and public information in common value auctions. *American Economic Review*, *76*(5), 894-920.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business*, *59*(4), 285-300.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, *98*(6), 1325-1348.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263-291.

Katok, E. (2011). *Using Laboratory Experiments to Build Better Operations Management Models.* Hanover, MA, USA: Now Publishers.

Katok, E., Olsen, T., & Pavlov, V. (2014). Wholesale pricing under mild and privately known concerns for fairness. *Production and Operations Management*, *23*(2), 285-302.

Katok, E., & Pavlov, V. (2013). Fairness in supply chain contracts: A laboratory study. *Journal of Operations Management*, *31*(3), 129-137.

Klosterhalfen, S., Kiesmüller, G., & Minner, S. (2011). A comparison of the constant-order and dual-index policy for dual sourcing. *International Journal of Production Economics*, *133*(1), 302-311.

Klosterhalfen, S., Minner, S., & Willems, S. P. (2014). Strategic safety stock placement in supply networks with static dual supply. *Manufacturing & Service Operations Management*, *16*(2), 204-219.

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

Kremer, M., Minner, S., & Van Wassenhove, L. N. (2014). On the preference to avoid ex-post inventory errors. *Production and Operations Management*, *23*(5), 773-787.

Kremer, M., Moritz, B. B., & Siemsen, E. (2011). Demand forecasting behavior: System neglect and change detection. *Management Science*, *57*(10), 1827-1843.

Kremer, M., & Van Wassenhove, L. N. (2014). Willingness to pay for shifting inventory risk: The role of contractual form. *Production and Operations Management*, *23*(2), 239-252.

Kunreuther, H. (2006). Disaster mitigation and insurance: Learning from Katrina. *Annals of the American Academy of Political and Social Science*, *604*(1), 208-227.

Lagodimos, A. G. (1992). Multi-echelon service models for systems under different rationing policies. *International Journal of Production Research*, *30*(4), 939-958.

Lagodimos, A. G., & Koukoumialos, S. (2008). Service performance of two-echelon supply chains under linear rationing. *International Journal of Production Economics*, *112*(2), 869-884.

Lankton, N., & Luft, J. (2008). Uncertainty and industry structure effects on managerial intuition about information technology real options. *Journal of Management Information Systems*, *25*(2), 203-240.

Laplace, P. (1814). *A Philosophical Essay on Probabilities.* Paris, France: M^{me} V^{e} Courcier.

Lariviere, M. A., & Porteus, E. L. (2001). Selling to the newsvendor: An analysis of price only contracts. *Manufacturing & Service Operations Management*, *3*(4), 293-305.

Lau, N., & Bearden, J. N. (2013). Newsvendor demand chasing revisited. *Management Science*, *59*(5), 1245-1249.

Lau, N., Hasija, S., & Bearden, J. N. (2014). Newsvendor pull-to-center reconsidered. *Decision Support Systems*, *58*(1), 68-73.

Lee, Y. S., & Siemsen, E. (2016). Task decomposition in the newsvendor experiment. *Management Science, forthcoming*.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The balloon analogue risk task (BART). *Journal of Experimental Psychology: Applied*, *8*(2), 75-84.

Levy, H. (1994). Absolute and relative risk-aversion: An experimental study. *Journal of Risk and Uncertainty*, *8*(3), 289-307.

Loch, C. H., & Wu, Y. (2007). *Behavioral Operations Management.* Hanover, MA, USA: Now Publishers.

Loch, C. H., & Wu, Y. (2008). Social preferences and supply chain performance: An experimental study. *Management Science*, *54*(11), 1835-1849.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, *92*(368), 805-824.

Lönnqvist, J.-E., Verkasalo, M., Walkowitz, G., & Wichardt, P. C. (2015). Measuring individual risk attitudes in the lab: Task or ask? An empirical comparison. *Journal of Economic Behavior & Organization*, *119*(1), 254-266.

Lu, L. X., & Lariviere, M. A. (2012). Capacity allocation over a long horizon: The return on turn-and-earn. *Manufacturing & Service Operations Management*, *14*(1), 24-41.

Lusardi, A., & Mitchell, O. S. (2008). Planning and financial literacy: How do women fare? *American Economic Review*, *98*(2), 413-417.

Lusk, J. L., & Coble, K. H. (2005). Risk perceptions, risk preference, and acceptance of risky food. *American Journal of Agricultural Economics*, *87*(2), 393-405.

Lyon, T. P. (2006). Does dual sourcing lower procurement costs? *Journal of Industrial Economics*, *54*(2), 223-252.

Mador, G., Sonsino, D., & Benzion, U. (2000). On complexity and lotteries' evaluation - three experimental observations. *Journal of Economic Psychology*, *21*(6), 625-637.

Mengel, F., & Peeters, R. (2011). Strategic behavior in repeated voluntary contribution experiments. *Journal of Public Economics*, *95*(1-2), 143-148.

Minner, S. (2003). Multi-supplier inventory models in supply chains: A review. *International Journal of Production Economics*, *81-82*(1), 265-279.

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.

Moritz, B. B., Siemsen, E., & Kremer, M. (2014). Deliberate or intuitive: Decision speed and cognitive reflection in demand forecasting behavior. *Production and Operations Management*, *23*(7), 1146-1160.

Murnighan, J. K., Roth, A. E., & Schoumaker, F. (1988). Risk aversion in bargaining: An experimental study. *Journal of Risk and Uncertainty*, *1*(1), 101-124.

Müller, W. (2001). Strategies, heuristics, and the relevance of risk aversion in a dynamic decision problem. *Journal of Economic Psychology*, *22*(4), 493-522.

Narayanan, A., & Moritz, B. B. (2015). Decision making and cognition in multi-echelon supply chains: An experimental study. *Production and Operations Management*, *24*(8), 1216-1234.

Ovchinnikov, A., Moritz, B. B., & Quiroga, B. F. (2015). How to compete against a behavioral newsvendor. *Production and Operations Management*, *24*(11), 1783-1793.

Park, M., Pekgün, P., & Keskinocak, P. (2015). Does a forecast-accuracy based allocation scheme induce customers to share truthful forecasts? *Working paper, University of South Carolina, Moore School of Business, Columbia, SC, USA*.

Persson, T., & Tabellini, G. (1994). Is inequality harmful for growth? *American Economic Review*, *84*(3), 600-621.

Ploner, M., & Regner, T. (2013). Self-image and moral balancing: An experimental analysis. *Journal of Economic Behavior & Organization*, *93*(1), 374-383.

Poulton, E. C. (1989). *Bias in Quantifying Judgments*. Hove, UK: Erlbaum.

Pratt, J. W. (1964). Risk-aversion in the small and in the large. *Econometrica*, *32*(1-2), 122-136.

Prelec, D. (1998). The probability weighting function. *Econometrica*, *66*(3), 497-527.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, *3*(4), 323-343.

Rabin, M., & Vayanos, D. (2010). The gambler's and hot-hand fallacies: Theory and applications. *Review of Economic Studies*, *77*(2), 730-778.

Raeva, D., Mittone, L., & Schwarzbach, J. (2010). Regret now, take it now: On the role of experienced regret on intertemporal choice. *Journal of Economic Psychology*, *31*(4), 634-642.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in English and German. *Journal of Research in Personality*, *41*(1), 203-212.

Ren, Y., & Croson, R. (2013). Overconfidence in newsvendor orders: An experimental study. *Management Science*, *59*(11), 2502-2517.

Reynaud, A., & Couture, S. (2012). Stability of risk preference measures: Results from a field experiment on French farmers. *Theory and Decision*, *73*(2), 203-221.

Rosenshine, M., & Obee, D. (1976). Analysis of a standing order inventory system with emergency orders. *Operations Research*, *24*(6), 1143-1155.

Russo, J. E., & Schoemaker, P. J. H. (1990). *Decision Traps: Ten Barriers to Brilliant Decision-making and How to Overcome Them*. New York, NY, USA: Simon & Schuster.

Sabater-Grande, G., & Georgantzis, N. (2002). Accounting for risk aversion in repeated prisoners' dilemma games: An experimental test. *Journal of Economic Behavior & Organization*, *48*(1), 37-50.

Saha, A. (1993). Expo-power utility: A flexible form for absolute and relative risk-aversion. *American Journal of Agricultural Economics*, *75*(4), 905-913.

Scheller-Wolf, A., Veeraraghavan, S., & Van Houtum, G. J. (2006). Effective dual sourcing with a single index policy. *Working paper, Carnegie Mellon University, Pittsburgh, PA, USA*.

Schultz, K. L., Juran, D. C., Boudreau, J. W., McClain, J. O., & Thomas, L. J. (1998). Modeling and worker motivation in JIT production systems. *Management Science*, *44*(12), 1595-1607.

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.

Segal, C. (2012). Working when no one is watching: Motivation, test scores, and economic success. *Management Science*, *58*(8), 1438-1457.

Sheopuri, A., Janakiraman, G., & Seshadri, S. (2010). New policies for the stochastic inventory control problem with two supply sources. *Operations Research*, *58*(3), 734-745.

Smith, V. L. (1962). An experimental study of competitive market behavior. *Journal of Political Economy*, *70*(2), 111-137.

Smith, V. L. (1976). Experimental economics: Induced value theory. *American Economic Review*, *66*(2), 274-279.

Stephens, T. A., & Tyran, J.-R. (2015). Money illusion, loss aversion, and asset allocation. *Working paper, Vienna University of Economics and Business, Vienna, Austria*.

Sterman, J. D. (1989). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision-making experiment. *Management Science*, *35*(3), 321-339.

Strohhecker, J., & Größler, A. (2013). Do personal traits influence inventory management performance? - The case of intelligence, personality, interest and knowledge. *International Journal of Production Economics*, *142*(1), 37-50.

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

Sundali, J., & Croson, R. (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgment and Decision Making*, *1*(1), 1-12.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Boston, MA, USA: Pearson.

Tanaka, T., Camerer, C. F., & Nguyen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *American Economic Review*, *100*(1), 557-571.

Thaler, R. H., Tversky, A., Kahneman, D., & Schwartz, A. (1997). The effect of myopia and loss aversion on risk taking: An experimental test. *Quarterly Journal of Economics*, *112*(2), 647-661.

Tomlin, B. (2006). On the value of mitigation and contingency strategies for managing supply chain disruption risks. *Management Science*, *52*(5), 639-657.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), 1275-1289.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297-323.

Veeraraghavan, S., & Scheller-Wolf, A. (2008). Now or later: A simple policy for effective dual sourcing in capacitated systems. *Operations Research*, *56*(4), 850-864.

Visschers, V. H. M., Meertens, R. M., Passchier, W. W. F., & de Vries, N. N. K. (2009). Probability information in risk communication: A review of the research literature. *Risk Analysis*, *29*(2), 267-287.

Von Gaudecker, H.-M., Van Soest, A., & Wengström, E. (2008). Selection and mode effects in risk preference elicitation experiments. *Working paper, Institute for the Study of Labor (IZA), Bonn, Germany*.

Von Gaudecker, H. M., Van Soest, A., & Wengström, E. (2011). Heterogeneity in risky choice behavior in a broad population. *American Economic Review*, *101*(2), 664-694.

Wakker, P. P. (2008). Explaining the characteristics of the power (CRRA) utility family. *Health Economics*, *17*(12), 1329-44.

Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity*. Cambridge, UK; New York, NY, USA: Cambridge University Press.

Wakker, P. P., & Deneffe, D. (1996). Eliciting Von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management Science*, *42*(8), 1131-1150.

Walker, J. M., Smith, V. L., & Cox, J. C. (1987). Bidding behavior in 1st price sealed bid auctions: Use of computerized Nash competitors. *Economics Letters*, *23*(3), 239-244.

Wang, C. X., & Webster, S. (2009). The loss-averse newsvendor problem. *Omega*, *37*(1), 93-105.

Weber, E. U., Blais, A. R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *15*(4), 263-290.

Whittemore, A. S., & Saunders, S. C. (1977). Optimal inventory under stochastic demand with 2 supply options. *SIAM Journal on Applied Mathematics*, *32*(2), 293-305.

Wilkinson, L., & Wills, G. (2005). *The Grammar of Graphics*. New York, NY, USA: Springer.

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, *37*(1), 60-82.

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, *42*(12), 1676-1690.

Wu, X., & Niederhoff, J. A. (2014). Fairness in selling to the newsvendor. *Production and Operations Management*, *23*(11), 2002-2022.

Zhao, Y., Zhao, X., Wang, L., Chen, Y., & Zhang, X. (2016). Does elicitation method matter? Behavioral and neuroimaging evidence from capacity allocation game. *Production and Operations Management*, *25*(5)*,* 919-934.