Variant Impacts of Managers' Cognitive Propensities on Bullwhip Effect Mitigation

by

Ahmet Onur Agca

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Engineering

University of Warwick, Warwick Manufacturing Group

January 2022

Table of Contents

Table of Figures vii
List of Tables ix
Acknowledgement x
Declaration xi
Abstract xii
List of Abbreviations xiii
1. Introduction 14
1.1 Chapter Introduction 14
1.2 Problem Setting and Background 14
1.3 Research Rationale 15
1.4 Research Aim and Questions 16
1.5 Chapter Summary and Thesis Outline 17
2. Literature Review 18
2.1 Chapter Introduction 18
2.2 Inventory Management in Supply Chains 18
2.2.1 Downstream Side: Demand and Bullwhip Effect 19
2.2.2 Upstream Side: Supply, Inventory and Order Placement 25
2.3 Managers' Cognitive Heterogeneity 27
2.3.1 Behavioural Operations and Supply Chain Management 27
2.3.2 Dual-Process Theory 32
2.4 Decision-Making Environment 38
2.4.1 Informational Presence 39
2.4.2 Time Pressure 40
2.4.3 Problem Complexity 41
2.5 Chapter Summary 42
3. Conceptual Model and Hypotheses 43
3.1 Chapter Introduction 43
3.2 Hypotheses 43
3.2.1 Baseline Hypothesis: Dual-Process Theory on Bullwhip Effect Mitigation 43
3.2.2 Situational Hypotheses 44
3.3 Conceptual Model and Chapter Summary 48
4. Research Design 50

4.1 Chapter Introduction..50
4.2 Research Fundamentals..51
4.2.1 Research Philosophy..51
4.2.2 Research Approach..52
4.2.3 Research Method: Questionnaire-based Online Experiment 53
4.3 Experiment..56
4.3.1 Defining the Variables 56
4.3.2 Utilising Variables...57
4.3.3 Construct Validity I - Control of Nuisance Variables 69
4.3.4 Construct Validity II – Control of Operability 70
4.3.5 Experiment Process 71
4.4 Data Collection...76
4.4.1 Unit of Analysis..76
4.4.2 Sampling and Participant Recruitment 76
4.4.3 External Validity II – Screening..77
4.5 Data Analysis..77
4.5.1 Conclusion Validity – Analysis 80
4.6 Quality of the Research..80
4.7 Ethics..81
4.8 Chapter Summary 81
5. Results..82
5.1 Chapter Introduction..82
5.2 Descriptive Results...82
5.3 Quality Control Analysis...83
5.3.1 Control of Nuisance variables..84
5.3.2 Control of Operability...88
5.3.3 Control of Internal Validity: Randomisation..89
5.4 Pre-analysis Dataset Preparation...90
5.5 Main Study Analysis 91
5.5.1 Baseline Hypotheses 91
5.5.2 Time Pressure Hypotheses...93
5.5.3 Problem Complexity Hypothesises...94
5.6 Chapter Summary 96
6. Discussions..98
6.1 Chapter Introduction..98
6.2 Bullwhip Effect Mitigation through the Lens of Dual-Process Theory 98

6.3 Environmental Conditions for Bullwhip and Cognition........99
6.3.1 Information Presence in Bullwhip Effect and Behavioural Studies........99
6.3.2 Time pressure in Bullwhip Effect and Behavioural Studies........101
6.3.3 Problem complexity in Bullwhip Effect and Behavioural Studies........103
6.4 Chapter Summary........105
7. Conclusion........107
7.1 Chapter Introduction........107
7.2 Revisiting the Research Questions........107
7.3 Theoretical Contributions........109
7.4 Practical Implications........110
7.5 Limitations and Future Research Directions........111
7.6 Chapter Summary........113
References........114
Appendices........129
A – Pool of Potential CRT Questions........129
B – Pilot Studies........132
B-1 The First Version: Beer Game Experiment........133
B-2 Second Pilot Study and the Verification of the Amendments........144
C – Main Experiment Appendices........145
C-1 Experiment File........145
C-2 Instructions........146
C-4 Codes Utilised in the Experiment........149
C-5 Raffle Information Message and Raffle Video........153
D – Ethical Appendices........154
D-1 Participant Information Leaflet (PIL)........154
D-2 Ethical Approval Form........154
E – Additional Analysis Appendices........154
E-1 CRT Previous Exposure Analysis........154
E-2 OMG Learning Curve Analysis........155

Table of Figures

Figure 2.1 A framework of supply chain management 19

Figure 2.2 Value perception in the 'Prospect Theory' 29

Figure 2.3 (a) Taxonomy of potential variables of BSCM discipline (b) Psychological factors in behavioural antecedents 31

Figure 2.4 Taxonomy of concepts in managerial cognitive studies 32

Figure 2.5 Cognitive style matrix 37

Figure 3.1 Conceptual Model 48

Figure 4.1 Research Onion 50

Figure 4.2 Conceptual Model (Also Figure 3.1) 56

Figure 4.3 Distribution of academic works that utilised the CRT by 2015 58

Figure 4.4 Seasonal demand flow for the 15 periods 67

Figure 4.5 Separation of the periods 78

Figure 4.6 Rigour consistency of a research 80

Figure 5.1 Distribution of total participants 83

Figure 5.2 Boxplot of the completion time of the experiment 83

Figure 5.3 Gender – Thinking disposition distribution 84

Figure 5.4 Gender – Thinking disposition distribution 84

Figure 5.5 Previous exposure distribution for a) participants and b) CRT questions 85

Figure 5.6 Previous exposure distribution alongside CRT responses 85

Figure 5.7 Previous exposure-dependent CRT performance 86

Figure 5.8 Visualised comparison of beginning and end performances 87

Figure 5.9 CRT response success and time distribution 88

Figure 5.10 Participant performance distribution 89

Figure 5.11 Coding examples for hypothesis measurement samples 90

Figure 5.12 Average order comparison for baseline hypotheses (a) H1, (b) H2a, (c) H2b 92

Figure 5.13 Average order comparison for time pressure treatment hypotheses 93

Figure 5.14 Average order comparison for problem complexity treatment hypotheses 95

Figure 5.15 Hypothesis results for proposed conceptual model 97

Figure 7.1 Research balance framework 108

Figure A.1 Exemplary slide sets for the Beer Game version 138

Figure A.2 Drop rate throughout the experiment 139

Figure A.3 CRT response ratio analysis 141

Figure A.4 CRT response ratio – previous exposure interaction 141

Figure A.5 CRT response ratio – answer time interaction 141
Figure A.6 CRT response ratio – numeracy control interaction.. 142

List of Tables

Table 2.1 Operational causes of the Bullwhip Effect 22

Table 2 2 Behavioural causes of the Bullwhip Effect 23

Table 2.3 Solutions for Bullwhip Effect mitigation 23

Table 2.4 Comparison of main features of dual thinking systems 36

Table 4.1 Positivism-Interpretivism comparison in three aspects 52

Table 4.2 Epistemological comparison of selected research methodologies in management studies 53

Table 4.3 Comparison of experimental research methods 55

Table 4.4 Customisation summary of the order management game 65

Table 4.5 Comparison of experiment designs 72

Table 4.6 Treatment combinations 73

Table 5.1 Distribution of participants throughout the treatment combinations 89

Table 5.2 Normality tests for the measurement samples 91

Table 5.3 Mann-Whitney U-Test results for baseline hypotheses 92

Table 5.4 Mann-Whitney U-Test results for time pressure treatment hypotheses 94

Table 5.5 Mann-Whitney U-Test results for problem complexity treatment hypotheses 96

Table A.1 Feedback categories 140

Table A.2 Summary of CRT evaluation 143

Acknowledgement

All along the years of this doctoral research, the list of invaluable people to thank has grown extensively, but since day 1, I have already been thanking one person, my supervisor Prof Janet Godsell. It was not just a doctorate for me but was a total change in the direction of my life. Dear Jan, I want to express my deepest gratitude to you for believing in me, making me always feel your support, providing me countless opportunities, and patiently guiding me in my apprenticeship to be a researcher. I would not be here without all these and I want you to always know that you have literally changed a life.

Second, I want to thank the teaching fellows whom I have worked with: Dr Adrian Watt, Dr Johannes Habel, Dr Ian Evans and Dr Mujthaba Ahtamad. I indeed learned more than I taught throughout all modules and I owe you so much for providing these opportunities.

I also want to thank my dear friends and colleagues for their priceless companionship. Donato, Joshua, Taofeeq, Steven, Sumeer, Zakiah, Reem, Sha, Rajesh, Susanna, both Alex's, Prang, Naili, Thames, Anthea, Aitana, Francis, Miray, Fatima, and of course Vivienne. Muco, Ugur and Yesim, I want to note a special thanks for your great support in the lockdown times. We have been going through very strange times which I feel they should be noted down here for ever.

I thank my dear cats Sek, Pebbles, Tux, and especially Paulo. I have spent all the lockdown with Paulo and you little furry thing, I could not manage that well without your warm purr.

Next, I would like to express my greatest thanks to my dear wife, Umran. For the last four years, we have gone through so many things together, from the most precious memories to the hardest challenges. I want to thank you for every little thing you did, for being with me in this journey, never stopping believing in me, for your beautiful heart and endless love. Teşekkür ederim Kimyam.

Lastly, I also want to thank some people that are special to me, starting with Jules Verne and Alexandre Dumas, for bestowing me the vast lands of their imagination. Thanks to them, I have become a relentless person, just like their characters. Next, I want to thank my science heroes Michael Faraday and James Clerk Maxwell, who have been the starters of my love for science. Finally, I want to thank Mustafa Kemal Ataturk, who is my lifelong inspiration for the truth, progression and diligence.





Declaration

I confirm that this thesis is my own work. Including the practical tasks like data collection and analysis, all parts of the thesis were conducted and written by me.

I confirm that this thesis has not been submitted for a degree at another university.

Abstract

The Bullwhip effect (BWE) was identified as a problem for the supply chains more than half a century ago. Complicating the inventory management and order fulfilment, it caused higher cost throughout the supply chain. Although a great amount of effort was spent by the researchers to identify its reasons and to find solutions, BWE continues to appear. Previous studies investigated the issue from various operational and behavioural aspects, however, the impact of individuals' cognitive differences on the occurrence and mitigation of the bullwhip has been understudied. Adopting the dual-process theory from cognitive sciences as the theoretical lens, this study identified interpersonal cognitive differences and their associated bullwhip mitigation performance under three decision-making environments: information presence, time pressure and problem complexity.

To test the hypotheses that measure these performance differences concerning various configurations of managerial cognitive propensity and environmental conditions, this study applied an online experiment (n=623), which comprised of two sections. Participants were manipulated for their cognitive propensities for either the intuitive (System 1) or the rational (System 2) thinking system via a customised Cognitive Reflection Test (CRT). A devised Order Management Game was utilised to measure their bullwhip management performance under different environmental conditions.

The findings initially affirmed the previous findings with regards to the outperformance of System 2 managers to System 1 managers. It was observed that both types of managers track the customer demand better when they were provided with additional information, although this performance improvement was not expected for System 1 managers. However, this positive impact of information was eliminated when the time pressure or problem complexity was added to the decision-making environment. It was viewed that although the System 2 managers continued to outperform System 1 managers under time pressure, the performance gap between the managerial groups decreased. Time pressure boosted System 1 managers' bullwhip mitigation performance while reduced that of System 2 managers. Problem complexity, however, decreased the bullwhip mitigation performance of both managerial groups. But it was noteworthy that the performance decrease of System 2 managers was much lower than that of System 1 managers.

This study initially contributed to the field of the bullwhip effect by revealing the behavioural and context-dependent variations concerning its mitigation. These inferences can also be utilised for practical aims to efficiently utilise human resources. Secondly, the adopted dual-process theory was extended with the perspective of this study. Lastly, the prepared experiment tool brought novel aspects for the measurement of this study's variables. Regarding these expansions, this study also provides future research directions for the researchers.

Keywords: *Bullwhip Effect; Behavioural supply chain; Dual-process theory; Cognitive Reflection Test; Decision making.*

List of Abbreviations

BDG	Beer Distribution Game
BOM	Behavioural Operations Management
BSCM	Behavioural Supply Chain Management
BWE	Bullwhip Effect
CRT	Cognitive Reflection Test
NP	Newsvendor Problem
OM	Operations Management
OMG	Order Management Game
SC	Supply Chain
SCM	Supply Chain Management

1. Introduction

"Each problem that I solved became a rule, which served afterwards to solve other problems."

RENÉ DESCARTES

1.1 Chapter Introduction

This first chapter of the thesis will introduce you to the study in four sections.

Section 1.2 will describe the research problem that motivates this research and provide a summative background of the core concepts of this study.

Section 1.3 will highlight the gap in the extant literature and associated research rationale.

Research questions to fill this gap and the research objectives will be presented in Section 1.4.

The summary of this chapter is given in Section 1.5, which will address the thesis outline and provide an introduction to the following chapters as well.

1.2 Problem Setting and Background

Supply chain management (SCM) is based on the idea of managing the flow of inventory from raw material to customer, while receiving the flow of information and income through the opposite direction. In other words, orders moving upstream and materials downstream comprise the traditional understanding of supply chains (SCs) (Steckel, Gupta and Banerji, 2004). Throughout the SC, various decisions are made by the managers such as storing inventory, placing orders and tracing the shipments. Likewise, various problems are also faced when managing those activities and among them the bullwhip effect (BWE) (Forrester, 1958) is accepted as a common SC disruption. It relies on the amplification of variability in orders towards the upstream SC. This amplification becomes a problem for the SCs as the reactive inventory level fluctuates, resulting in excessive or insufficient inventory stock, which means extra cost when managing that inventory, or customer shortages (Croson and Donohue, 2006). The issue has created a great interest in the field for a long time (Lee, Padmanabhan and Whang, 1997a,b; Chen et al., 2000) and various operational solutions have been proposed so far. However, the effect continues to occur in SCs, despite some reductions in its amplitude (Steckel, Gupta and Banerji, 2004; Croson and Donohue, 2006). From a broad perspective, this study aims to contribute to the endeavour to alleviate bullwhip in SCs, as preceding studies have also done.

Focusing on the operational basics of BWE, researchers have found some reasons behind the phenomenon, such as inventory rationing and order batching (Lee et al., 1997a,b). However, the various solutions provided have not brought an ultimate solution to the issue. Results from Croson and Donohue (2006) assert that the bullwhip phenomenon is not caused only by operational triggers such as unpredictability or fluctuating demand profile. It continues to occur even in the optimal conditions where the mentioned operational complications are removed. Researchers have therefore changed their perspective to a new direction: behavioural sciences.

Even though behavioural sciences are adopted by many disciplines, in operations management (OM) they have been overlooked until recently (Knemeyer and Naylor, 2011). Traditional studies in the SCM and OM fields have suggested that managers behave rationally when they make decisions. However, this idea relies on decision makers having unlimited cognitive capabilities (Simon, 1955); in real life, however, this is not true. The impact of the human factor in OM and SCM studies has therefore gained importance (Maitland and Sammartino, 2015), in constructing the fields 'Behavioural Operations Management (BOM)' and 'Behavioural Supply Chain Management (BSCM)'. These fields, in contrast to their traditional counterparts, begin with accepting that a human being – as a decision-maker – has bounded rationality, resulting in some systematic errors in his/her decisions. Accordingly, researchers of behavioural fields believe that human intervention should be taken into consideration when searching for solutions to operational anomalies. It is believed that even though human psychology cannot be altered, various precautions and manipulations can be produced in order to reduce the number of systematic errors and obtain the best output from resources, including humans (Gino and Pisano, 2008).

In line with this, some researchers examined the behavioural causes of BWE in their studies (e.g., Wu and Katok, 2006; Niranjan, Wagner and Bode, 2011). The majority of them focused on two main biases: underweighting the supply line and anchoring (Zhao and Zhao, 2015). However, this focus created an anomaly in the topic. Most of the studies limited their behavioural aspect to intuition and more specifically to biases (Kaufmann, Wagner and Carter, 2017), but human behaviour and cognition are not composed of biases alone. Therefore, to bring a holistic behavioural perspective to the SCM field, to properly determine the complete boundaries of humans' *'bounded rationality'* (*see* Simon 1956; Kahneman 2003) and consequently to avoid this anomaly of single-direction research stream, it is required to construct a concrete basis which defines human cognition and its relative advantages and disadvantages.

Dual-process theory provides a suitable theoretical base from which to fulfil the aforementioned need for behavioural studies in the SCM discipline. It goes beyond limiting human cognition to bias-related aspects by defining the bias as only a part of the intuitive thinking system. It proposes that human cognition is comprised of two thinking systems: intuitive thinking system (namely System 1) with quick and autonomous decision-making capability and rational thinking system (System 2) with slow but detailed decisions (Campitelli and Labollita, 2010; Pennycook et al., 2016). These thinking systems have relative advantages and disadvantages, and they collaboratively process the problem for a solution which is considered acceptably close to the optimum (Evans, 2003). Dual-process theory has been abundantly utilised in various fields; however, in explaining the bullwhip occurrence, its utilisation has been limited to the invaluable research of Dr Brent Moritz. Over the last decade, Moritz and his colleagues have measured the individual cognitive differences in Newsvendor Problem (NP) performance via dual-process theory, and then similar differences but in a more complex inventory management simulation, namely the Beer Distribution Game (BDG) (e.g., Moritz, Hill and Donohue, 2013; Narayanan and Moritz, 2015; Moritz, Narayanan and Parker, 2020). In all these invaluable studies, it was overall observed that participants with higher cognitive reflection (System 2) perform better in managing inventory and demand flows in comparison to participants with lower cognitive reflection (System 1).

1.3 Research Rationale

Building on the accumulated literature on the BWE and cognitive decision-making, the works of Dr Moritz have brought a dual-perspective explanation for the occurrence of bullwhip in SCs. However, it is essential to note the immaturity of research in this direction. While the separation of cognitive thinking systems constructs the foundation of dual-process theory, theory embodies invaluable details in relation to the interactions between those thinking system and their comparative roles in decision-making process. The results of Moritz's studies have, overall, demonstrated the superiority of higher cognitive reflectivity over lower cognitive reflectivity. Leaving this research lead at that point, hence, will indirectly contribute to the aforementioned bias against the competences of the intuitive thinking system. Instead, future research is needed to increase the coverage of dual-process theory in explaining the bullwhip occurrence.

To reveal the further capabilities of both thinking systems in dealing with the BWE, this research aims to highlight their characteristics. It was mentioned that an intuitive thinking system produces fast and autonomous decisions, while a rational thinking system processes the decision slowly and elaborately with more effort. In line with these characteristics; Campitelli and Labollita (2010) relate thinking systems' effectivity to the context of the problem. This aspect was supported by Carter, Kaufmann and Wagner (2017) who describe the thinking system as context-dependent. Likewise, Chater et al. (2018) mentioned the performance difference of thinking systems, depending on the appropriateness or hostility of the decision-making environment. Conclusively, this study has also followed this research perspective and has aimed to enhance the dual-process theory's application to the BWE phenomenon through investigating the impacts of different environmental contexts.

This study has applied three, systematically chosen, environmental contexts: presence of additional information, exposure to time pressure and higher complexity of the problem. While the first situational context has been selected to explore the impacts of dual thinking systems on an operational solution, the latter two have been considered as suitable for the characteristics of respective thinking systems. To elaborate, while time pressure is suggested to be suitable for intuitive thinking system's decision-making performance which produces instant decisions, it may result in the opposite impact for the decisions of the rational thinking system, which requires further time and effort for decisions. Likewise, problem complexity is regarded as a benign environment for the rational thinking system which works in detail, but as a hostile environment for the processing of intuitive thinking system since it intrinsically simplifies the problem and may lose important aspects.

1.4 Research Aim and Questions

Summarising the problems and gaps in the field, this study mainly aims to contribute to the endeavour to solve the BWE phenomenon in SCs. To achieve this broad target, it will extend the research lead that looks for the impacts of cognitive propensity to preferably utilise either thinking system (System 1 or 2) in decision-making, by investigating the environmental context-dependent performance variations. In line with this goal, answers to the following main research question and sub questions will be sought:

RQ: How do environmental conditions impact on the bullwhip mitigation performance of managers with different cognitive propensities via their demand tracking capabilities?

- How does the cognitive propensity of supply chain managers impact on their demand tracking capability?
- How does additional information presence differ from the performance of supply chain managers with different cognitive propensities?
- How does time pressure in the decision-making environment differ from the performance of supply chain managers with different cognitive propensities?
- How does the increased complexity of the problem differ from the performance of supply chain managers with different cognitive propensities?

1.5 Chapter Summary and Thesis Outline

This chapter has provided the introductory information about the study to inform the reader about the main concepts of the study. It has included a summative background about the core points of the study, research problem and rationale that have driven it, research questions and objectives to achieve, and lastly the structure of the study.

The rest of the study will be presented in six main chapters and each chapter is briefly explained in the following paragraphs.

Chapter 2 – Literature Review: This chapter will provide the essential aspects of three fields of literature. First, it will start by explaining a common SC disruption, namely the BWE from the aspects of downstream flow of inventory and upstream flow of demand. Next, managers' cognitive propensity will be presented by expressing the basics of behavioural OM and SCM fields, cognitive sciences, dual-process theory and cognitive thinking systems. Finally, the determined decision-making contexts will be described, namely information presence, time pressure and problem complexity.

Chapter 3 – Conceptual Model and Hypotheses: This chapter will present the relationships between the concepts reviewed throughout the previous chapter. These relationships will be the building blocks of a conceptual model and the hypotheses that will be tested in this study.

Chapter 4 – Research Design: This chapter will explain all the research processes applied before or during this study. Starting with the research philosophy and approach of the researcher, the selected research method will be discussed. It will follow the application of the selected method -i.e. the experiment- in great detail. Later, the essentials of the data collection and analysis methods will be presented. While the chapter will conclude by addressing the ethical considerations, the quality of the thesis will be touched upon throughout the chapter.

Chapter 5 – Results: This chapter will demonstrate the application of the planned research processes and the findings reached. It will first present the descriptive results and quality control checks, then the findings related to the proposed hypotheses.

Chapter 6 – Discussions: This chapter will interpret the findings presented in the previous chapter regarding the hypotheses by integrating them with the extant literature.

Chapter 7 – Conclusions: This chapter will conclude the thesis by addressing the achievement of the questions set at the beginning. Next, the theoretical contributions and practical implications of this thesis will be highlighted. Finally, the limitations of this study will be mentioned, together with the future research directions.

Following the main chapters, the Appendices will present additional information that was excluded from the main chapters in order to preserve the flow and integration of the core arguments.

2. Literature Review

"All truly wise thoughts have been thought already thousands of times; but to make them truly ours, we must think them over again honestly, until they take root in our personal experience."

JOHANN WOLFGANG VON GOETHE

2.1 Chapter Introduction

This chapter will provide a review of the existing literature on three main topics that underpin this research:

- Inventory management in SCs,
- Managers' cognitive heterogeneity, and
- Decision-making environment.

In detail, Section 2.2 will reveal the inventory management from two aspects of SCs: downstream (2.2.1) and upstream (2.2.2). The former subsection will mention the occurrence of the bullwhip effect (BWE), its bond to demand management, the reasons behind it and potential solutions that have been proposed to date. The latter subsection, on the other hand, will explain the interconnection among the supply, inventory and order management, and the factors that impact on managers' ordering decisions.

Section 2.3 will analyse the managers' cognitive heterogeneities in two subsections: behavioural operations (BOM) and supply chain management (BSCM) (2.3.1), and dual-process theory (2.3.2). The first will give background information on the fundamentals of behavioural and decision-making sciences, their impacts on OM and SCM fields, and the emergence of BOM and BSCM fields. Then, Subsection 2.3.2 will discuss the working mechanisms of human cognition, its role in individuals' decision-making processes, dual-process theory's contribution to the topic, its dual components, namely intuitive and rational thinking systems, and the interrelationship between them.

Section 2.4 focuses on three environmental conditions that impact individuals' decision-making processes: information presence (2.4.1), time pressure (2.4.2) and problem complexity (2.4.3).

The chapter concludes with Section 2.5, which summarises the most prominent points discussed throughout the chapter.

2.2 Inventory Management in Supply Chains

Together with the changing dynamics of the industrial world, it has been understood that firms cannot behave as if they are isolated from others (Carter, Rogers and Choi, 2015). Successful in-house implementation of operations will be inadequate without an aligned set of partners (Gavirneni, Kapuscinski and Tayur, 1999). Therefore, the considered unit for competition has been upgraded from firm competition to SC (Baihaqi and Beaumont, 2006; Christopher, 2017), and these shifts have created the base of an SC idea. SCs are described as alliances of companies that cooperatively aim to meet the end customer demand (Klueber and O'Keefe, 2013). To achieve that aim, they deliver products or services to the market, while they also collect order information and financial flow back to the upstream SC partners (Moritz et al., 2013; Narayanan and Moritz, 2015). Accordingly, SCM can be defined as:

“Strategic and systematic coordination of a network of firms, in order to manage the downstream flow of products or services to the end customer, and upstream flow of information and finance to the network members.” (Bustinza, Parry and Vendrell-Herrerro, 2013; Klueber and O’Keefe, 2013).

Tan (2001) collates the SC literature into two main schools: purchasing and supply perspective, and transportation and logistics perspective. From both perspective, Figure 2.1 shows the benefits of adopting SC understanding in the ‘objective’ boxes. In line with those points, Fawcett, Magnan and McCarter (2008) highlight some expected benefits from an effective SCM, such as inventory reduction, enhanced delivery performance, shorter research and development (R&D) cycles, higher quality, better asset management, and superior channel relationships.

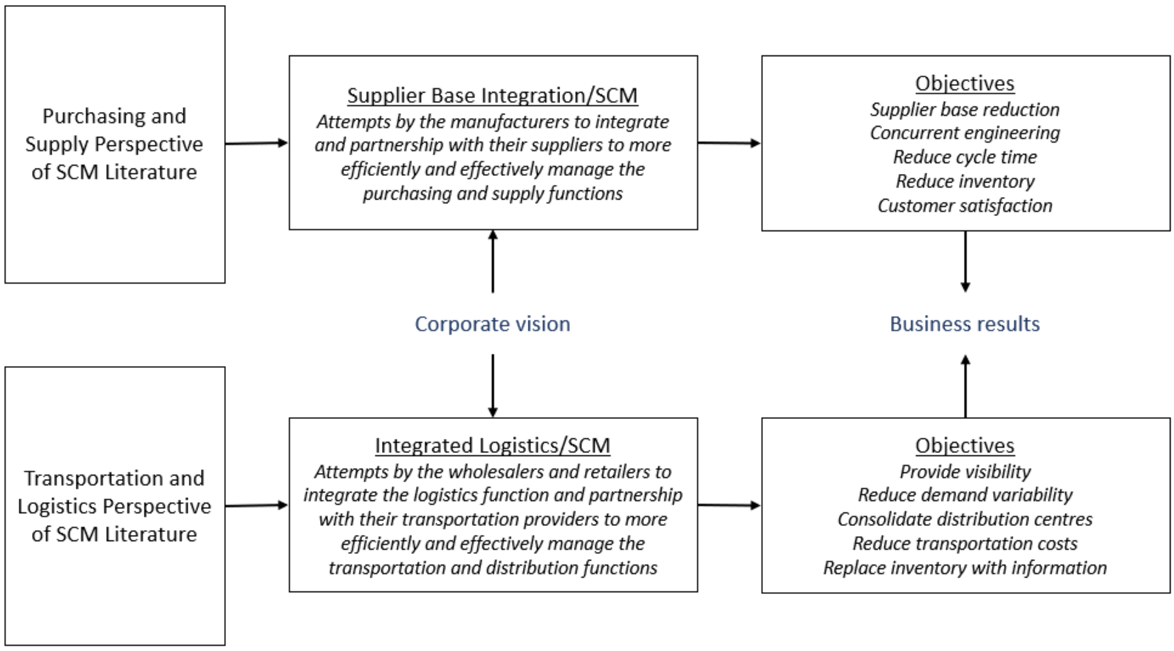


Figure 2.1 A framework of supply chain management (Tan, 2001)

For both perspectives, inventory and its management among the echelons are the core in SCM. Understanding of inventory varies in echelons. While downstream SC partners mostly handle finished goods for end customer, firms of downstream SCs – such as factories – focus more on creating raw material and manufacturing the products. Although it seems simple and straightforward, a rise in customer-led servitisation and globalisation increases the complexity of SCs and makes those tasks more difficult to manage (Narayanan and Moritz, 2015). However, no matter how complex and long they are, SCs are built on the successful balance between supply and demand (Drucker, 1973; Porter, 1985; Steckel et al., 2004; Mentzer and Estel, 2010).

2.2.1 Downstream Side: Demand and Bullwhip Effect

Although the term ‘supply chain’ is the most commonly accepted term to define the field, it has also been called the ‘demand chain’ by many (Heikkilä, 2002; Jüttner, Christopher and Baker, 2007; Christopher and Ryals, 2014). This fact alone may indicate the importance of demand in SCs; however, we will go beyond this and elaborate its further benefits.

Having information about the incoming demand is also essential for inventory management. It is the starting point of all inventory-related decisions and strategies such as order fulfilment, cost reduction and controlling the related inventory levels, stockouts and backlogs. For all these decisions, companies analyse their customers' demand profile and adapt their procurement and production schedules accordingly. Additionally, they detect the potential mitigable risks and produce solutions by negotiating either with suppliers or customers. All these activities are based on the forecasted or obtained demands of the customer.

Although this may seem very straightforward, practice is more complex than theory. Because of the varying order and shipment delays in different levels of SCs, and SC members' decision-making capabilities, SC members experience different periods of surplus and stockouts (Fisher et al., 1994). This leads them to act in a more conservative way, which results in orders that are larger than the amount of downstream sales. This behaviour amplifies through upstream and ends up with a propagated distortion in orders, stocks, production capacity and even in storage space (Steckel et al., 2004; Strozzi, Bosch and Zaldivar, 2007; Coppini et al., 2010; Narayanan and Moritz, 2015; Moritz et al., 2020).

In the SCM literature, this flow of events is defined as the 'Bullwhip (or Forrester) Effect (BWE)' (Steckel et al., 2004), which was first put forward by Forrester (1958) and coined by Lee et al. (1997a,b) (Sterman, 2000; Disney and Towill 2003; Chen and Disney, 2007).

BWE is accepted as a well-known problem for SCs as it causes many negative outcomes. Among the most crucial are:

- The additional requirement of managing inventory, such as stricter control or more frequent expedition,
- Relative overstocking due to unexpected demand variability,
- Potential shortages against customer demand,
- Periodical capacity inefficiencies in operations such as production or logistics, and
- Consequent cost increases and profit reductions (Lee et al. 1997b) (Croson and Donohue, 2006; Chen and Disney, 2007; Coppini et al., 2010; Narayanan and Moritz, 2015).

What makes the BWE an important concept in the field is its persistency. Though half a century has passed since the phenomena was first identified, the studies of Bray and Mendelson (2012) and Shan et al. (2014) assert that nearly two-thirds of companies still experience the problem (Narayanan and Moritz, 2015). Firms (e.g., Wal-Mart, P&G, GE's Appliance Division, Baxter) have invested in reforming their SCs to sort out the negative impacts of bullwhip (Steckel et al., 2004). The problem has been seen in many sectors (e.g., Croson et al., 2014; Moritz et al., 2020). In the electronics industry, for example, Lee et al. (1997b) mention Hewlett-Packard's BWE-related loss of millions of dollars which occurred through inefficient capacity and excessive stocking. This was triggered by a sudden demand increase for the LaserJet printers after a supply shortage. Likewise, Cisco Systems experienced a similar problem with a cost of more than US$ 2 billion, when the company ended up with excessive inventory after retailer demands reduced drastically (Oliva and Gonçalves, 2005). In their empirical study, Terwiesch et al. (2005) also detect the issue in the computer industry by comparing the volatility in semiconductor manufacturing and sales of personal computers (Moritz et al., 2020). Furthermore, Cachon, Randall and Schmidt (2007), express the prevalence of BWE even in the wholesale industries and industries with low seasonality (Croson et al., 2014).

Being an expensive industrial problem, over the years the BWE phenomenon has become a popular topic for academic studies (Kahn, 1987; Lee et al., 1997a, 2000; Metters, 1997; Baganha and Cohen, 1998; Chen et al., 2000; Chatfield et al., 2004; Geary, Disney and Towill, 2006; Coppini et al., 2010; Narayanan and Moritz, 2015). Prior research aimed to prove the existence of the phenomenon by searching for the amplified orders towards the upstream SCs (Moritz et al., 2020). This was followed by studies that elaborate the spectrum and search for the possible reasons behind it and for potential countermeasures (Coppini et al., 2010; Moritz et al., 2020), which were aggregated under two schools of thought: 1) operational inefficiencies in SCs such as design faults, long delays in transportation or production, or communication deficiencies, and 2) behavioural or cognitive reasons that cause suboptimal decisions related to inventory and order management (Steckel et al., 2004; Oliva and Gonçalves, 2005; Wan and Evers, 2011; Tokar, Aloysius and Waller, 2012). It is also essential here to highlight that these two paths interact as managers' behavioural reactions depend on the operational context (Steckel et al., 2004).

With regards to the operational aspects, the BWE can be caused in relation to all parts of the SC (supplier, internal or customer) and also to the market dynamics. The two main factors seem to be stochastic lead-times on the supplier side and a lack of information sharing from the customer side. This study has comprehensively reviewed the literature and grouped the operational causes of the BWE basically into two: internal and external causes. Among the internals are strategies like order batching, ration gaming and inventory management policy can be counted. Examples for the external causes are, on the other hand, lead time of delivery, demand signal processing, price variations, information sharing among SC partners, non-linear interactions, horizontal competition and product nature. Table 2.1 provides explanations for these operational causes.

Regarding the second group, behavioural research started with Sterman's (1989a,b) studies utilising the Beer Distribution Game (BDG). Then, they continued with many behavioural experiments (Oliva and Gonçalves, 2005; Castañeda, 2019) in two special aspects: cognitive limitations and biases. The former researched the impacts of managers' cognitive capabilities on the occurrence of the BWE, whereas the latter specifically focused on bullwhip occurrence because of managers' intuitive biases. The main biases that literature provides are the supply line underestimation, level bias (anchoring and pull-to-centre effect) which leads to misperception of inventory and demand tracking, and coordination risk perception. Table 2.2 provides the details regarding these behavioural causes as well as their resources.

In finding the reasons behind bullwhip inefficiencies, academics have also focused their studies on potential solutions. The proposed solutions can be grouped into three main categories. Firstly, the job can be simplified for the operator by reducing the cognitive load of the manager, training them or automating the decisions. Secondly, improving the internal operations by applying appropriate strategies can be another way of mitigating the bullwhip. Lastly, decreasing the external uncertainties by reducing the lead time, increasing visibility and collaboration are expressed in the literature as solutions for the BWE. Table 2.3 explains these solutions in detail, their impacts on the different BWE causes and resources for further information.

Table 2.1 Operational causes of the Bullwhip Effect

	Factors	Explanation	Resources
Internal	Order batching	Larger batches lead to higher production lead times and require companies to store more products. This is also called the *Burbidge effect.*	Lee et al. (1997a,b), Disney & Towill (2003), Lee, Padmanabhan and Whang (2004), Coppini et al. (2010), Narayanan & Moritz (2015)
	Ration gaming strategies	If managers choose to apply some ration gaming strategies due to shortages, it may distort the order flow towards the upstream SC. This is also called the *Houlihan effect.*	Lee et al. (1997a,b), Lee et al. (2000), Disney & Towill (2003), Lee et al. (2004), Oliva & Gonçalves (2005)
	Inventory management policy	In order to control their production and distribution, managers implement some ordering policies that suit their inventory and capacity requirements. However, these ordering policies may cause the BWE. Safety stock, for instance, is a well-known policy for unexpected demand peaks. Besides that, organisations may also hold extra stock when they have not built a strong coordination among the partners. Consequently, they may not trust the information acquired from partners and keep extra stock, also called coordination stock.	Lee et al. (1997a,b), Dejonckeere et al. (2004), Steckel et al. (2004), Chen & Disney (2007), Coppini et al. (2010), Croson et al. (2014)
External	Lead time ***(S)****	Non-zero lead times, especially in the inbound logistics, complicate the situation when meeting customer demands. Similarly, it becomes harder when the lead time has a stochastic nature. Being incapable of knowing the upcoming lead times, managers become more cautious about demand failure risks and tend to hold more safety stock. These lead time-related situations have long been accepted as a trigger of the Forrester effect and both situations result in an increase in BWE.	Lee et al. (1997a,b), Chen (1999), Chen et al. (2000), Lee, So & Tang (2000), Disney & Towill (2003), Chatfield et al. (2004), Steckel et al. (2004), Coppini et al. (2010), Narayanan & Moritz (2015)
	Demand signal processing ***(C)***	Demand signal processing has a critical place in inventory management; however, the difficulties in reaching the required customer demand information result in uncertainties and incorrect demand forecasting. Wrong forecasts will lead to faulty supply activities, and consequently to the BWE.	Lee et al. (1997a,b), Lee et al. (2000), Disney & Towill (2003), Lee et al. (2004), Coppini et al. (2010), Narayanan & Moritz (2015)
	Price variations ***(C)***	Price fluctuations for various reasons, such as promotional impact on the purchase behaviour of customers, change the demand profile, which creates uncertainty and BWE along the SCs.	Lee et al. (1997a,b), Lee et al. (2000), Lee et al. (2004), Disney & Towill (2003), Coppini et al. (2010)
	Information sharing ***(S, C)***	Together with the non-zero lead times, disruption in information sharing is regarded as another core enabler of the Forrester effect. Lacking the point of sale (POS) or market research information from the downstream side or supply status information from the upstream constrain the forecasting activities of the firms and leave them open to any unexpected demand fluctuations, in other words the BWE. Despite accessing information being key, its quality is also essential as this will define the quality of the decisions taken.	Forrester (1958; 1961), Chen et (1999), Cachon and Fisher (2000), Chen et al. (2000), Lee et al. (2000), Chatfield et al. (2004), Steckel et al. (2004), Coppini et al. (2010), Narayanan & Moritz (2015)
	Non-linear interactions ***(S, C)***	When the Forrester effect was first described, the non-linear interactions were shown as an explanation of the BWE. Along the SCs, firms interact with multiple suppliers or customers at different levels, as well as the internal departments also having their own horizontal hierarchies. This structure decreases the efficiency of information sharing and coordination among partners at any level.	Forrester (1958; 1961)
	Horizontal competition ***(M)***	Horizontal relationships also impact on the market competition. Activities of competing firms, such as ordering policies or their market share, are critical for a firm's demand and inventory management. The situation brings with it an ever-dynamic BWE state.	Oliva & Gonçalves (2005)
	Product nature ***(M)***	The nature of the product may also impact on the occurrence of the BWE. For example, the BWE is more common when SCs include products with seasonal demand, of low durability, or those which are more prone to be impacted by popular trends. Another factor related to product is its supply amount; when the supply of a product is critical, firms may order more than they need, as a precaution against any supply risk. This process will also result in the BWE through the upstream SC.	Oliva & Gonçalves (2005) Bloomfield and Kulp (2013)

***(S: Supplier related, C: Customer related, M: Market related)*

Table 2 2 Behavioural causes of the Bullwhip Effect

Factors			Explanation	Resources
Cognitive limitations			Much of the behavioural studies in BWE field is interested in the cognitive attributes of managers, such as intelligence, cognitive reflection, impulsiveness, locus of control or temporal foci.	Croson and Donohue (2006), Cantor and Macdonald (2009), Cantor and Katok (2012), Bloomfield and Kulp (2013), Narayanan & Moritz (2015)
Intuitive Biases	Underestimating the supply line		Having bounded rationality, managers often fail to adequately consider the feedback information and the time delays in the operations. These lead managers to underweight the supply line (the inventory that was previously ordered but has not arrived yet) and make hustle decisions only considering the present demand and in-stock inventory. Previous research has proved the common existence and strong impact of this bias in the deviation from optimal decisions and conclusively the occurence of the BWE.	Sterman (1989a,b, 1994, 2000), Oliva & Gonçalves (2005), Croson et al. (2014), Narayanan and Moritz (2015)
Intuitive Biases	<u>Level bias:</u> - Anchoring - Pull-to-centre effect	Misperception of the inventory	Misperception of the inventory means either underreaction to the excessive inventory or overreaction to the backlog. For instance, managers may end up with insufficient inventory as they postpone ordering when they have excessive inventory, but they may also overreact to the backlog when they face shortages. Though some studies assert that these two situations differ in their impact on the BWE, Oliva & Gonçalves (2005) refute that distinction in their empirical work. Nevertheless, both sides agree that misperception of inventory triggers the BWE.	Oliva & Gonçalves (2005), Narayanan & Moritz (2015), Zhao and Zhao (2015), Castañeda (2019)
Intuitive Biases	<u>Level bias:</u> - Anchoring - Pull-to-centre effect	Misperceptions in demand tracking	Similar misperceptions can also be seen in the interpretation of the demand information. For example, managers may anchor their order amount to a certain amount (anchoring), may adjust the optimal order suggestions provided by decision support systems (adjustment) or give unnecessarily greater importance to the most recent demand (recency effect).	Sterman (1989a,b), Narayanan & Moritz (2015), Zhao and Zhao (2015), Castañeda (2019)
Intuitive Biases	Coordination risk		Safety stocks are commonly held by companies as a precaution against unexpected peaks in customer demand. A similar type of stock is coordination stock; however, the aim of keeping this stock is related to the lack of trust towards the partners in coordination. This distrust occurs when firms have insufficient certainty or knowledge about their partners' behaviours or cognitive abilities. This concern results in anomalies in ordering behaviours, and consequently with the BWE.	Croson et al. (2014)

Table 2.3 Solutions for Bullwhip Effect mitigation

Factors		Explanation	BWE Cause Mitigation	Resources
Simplify the job	Reducing the cognitive load	As human beings, we have bounded rationality. This results in misinterpretations of the situation when managing inventory or demand-related problems. In these instances, simplifying the job may be realised by reducing the managers' cognitive workload. It frees up the working memory and enables giving more attention to the target activity.	- Cognitive limitations ***(B)**** - Misperception of the supply line, inventory and demand ***(B)***	Croson et al. (2014)
Simplify the job	Automating the decisions	Managers have applied various techniques to reduce the amplitude of the BWE. These have included some solution models proposed by Operations Research, such as genetic algorithms (O'Donnell et al., 2006), fuzzy inventory controller (Xiong & Helo, 2006), distributed intelligence (De La Fuente & Lozano, 2007), proportional control (Disney & Towill, 2003; Chen & Disney, 2007) and a chaos theory technique (Strozzi et al., 2007). Over the years and following technological advances, firms have also integrated enhanced predictive analytics, such as machine-learning and artificial intelligence, to mitigate the negative effects of the BWE. All these methods are aimed at reducing the computational effort and simplifying the job.	- Cognitive limitations ***(B)***	Coppini et al. (2010)

	Training	The importance of training to mitigate the BWE is mentioned by some resources (e.g., Wu & Katok, 2006; Tokar et al., 2012). Although they actually mean the mitigation of the BWE in the BDG context, educating managers about the occurrence of the BWE and corresponding solutions will also impact their inventory and order management performance.	- Misperceptions of the supply line, inventory and demand *(B)* - Better understanding of the required decisions and processes *(O)*	Narayanan & Moritz (2015)
Operational improvisations		As the BWE partly occurs through operational causes, some researchers believe that enhancements in those operations can alleviate the BWE in operational scope. Among the solutions suggested are: holding extra inventory (Chen and Disney, 2007), demand forecasting improvisation (Chen, 1998), capacity allocation schemes (Cachon and Lariviere, 1999), staggered order batching (Cachon, 1999), and everyday low pricing (Sogomonian and Tang, 1993). Another critical set of operational solutions is grouped under the multi-echelon inventory management title. Despite the studies in the area dating back to the half of the previous century, with the advancements in the SCM, it has gained further attention and it is still an active research domain. De Kok et al. (2018)'s typology research comprehensively explains the dimensions that must be considered for inventory management-related problems. The dimensions cover the number of echelons, SC structure, time, information, capacity, delay, demand, customer, policy, lot size, flexibility and performance indicator. While the required configurations of these dimensions may bring solutions to various problems, authors especially highlight solutions like setting centralised base-stock levels across the SC for the BWE. In line with this solution, Moinzadeh (2002) analyses inventory and replenishment policies in a SC when they are managed with information-supported installation stock policies. Lastly, Banerjee and Burton (1994) also provide an invaluable study by comparing the inventory replenishment performances under coordinated and independent inventory management policies.	- Demand signal processing *(O)* - Price variations *(O)* - Order-batching - Inventory *(O)*	Disney & Towill (2003) Croson & Donohue (2006) Chen & Disney (2007) de Kok et al. (2018) Moinzadeh (2002) Banerjee and Burton (1994)
Decreasing uncertainty and risks	Reducing lead time	Stochastic and non-zero lead times are two of the major triggers of the BWE, hence, shorter and known lead times will bring certainty to the operations and reduce the BWE.	- Lead time *(O)*	Lee et al. (1997, 2000), Kaminsky & Simchi-Levi (1998), Steckel et al. (2004), Oliva & Gonçalves (2005), Coppini et al. (2010), Narayanan & Moritz (2015)
	Increasing visibility	The other major antecedent of BWE occurrence is the issue related to information sharing. Providing better visibility among SC partners will definitely increase the certainty in the environment, and hence decrease the BWE. Despite having disputes, many authors specifically highlight the importance of demand (POS) information and assert that having known and stationary customer demand reduces the impacts of the BWE.	- Demand signal processing *(O)* - Information sharing *(O)* - Coordination Risk *(B)*	Lee et al. (1997, 2000), Croson & Donohue (2003), Coppini et al. (2010), Steckel et al. (2004), Narayanan & Moritz (2015)
	Increasing collaboration	As an extension to visibility, firms can increase the level of collaboration with their partners to mitigate the negative impacts of the BWE. Some example applications are vendor managed inventory (VMI) and collaborative planning, forecasting, and replenishment (CPFR).	- Information sharing *(O)* - Non-linear interactions *(O)* - Coordination Risk *(B)*	Narayanan & Moritz (2015)

**(B: Behavioural Factor, O: Operational Factor)*

Despite extensive research in this area, SCs still suffer from the BWE. Because of this, the phenomenon is still the focus of field researchers. Some experimental studies in particular (Sterman, 1989a,b; Kaminsky and Simchi-Levi, 1998; Croson et al., 2014; Croson and Donohue 2006) – that control for all the impacts and the environment – assert that the BWE persists even after applying all the operational solutions proposed (Castañeda, 2019). This assertion has led researchers to conduct further studies on the behavioural path in order to understand the human-related background of the phenomenon (Oliva and Gonçalves, 2005; Croson et al.; 2014; Narayanan and Moritz, 2015; Moritz et al., 2020).

2.2.2 Upstream Side: Supply, Inventory and Order Placement

When the company receives a demand from a customer, it starts preparing the demanded product in the allocated timeline. Even though production is the core action of manufacturing industries, timely and accurate provision of the necessary raw materials or components at the shop floor is also of the same importance, because these actions are closely connected through the SC. When the demand has come from the customer with a lead time, without advance supply preparations, the company may experience issues such as missing orders, delays, high inventory holding costs or quality problems. To eliminate these threats, it is therefore extremely important to have a balanced demand and supply strategy which will help in holding acceptable amounts of inventory with the lowest possible capacity utilisation (Chen and Disney, 2007).

Another aspect increasing the impacts of supply and demand in the bullwhip creation and management is their paradoxical nature (Chen and Disney, 2007). This paradox can be detected easier if we expand our vision from the company lens to the SC lens. The concepts of demand and order are, in fact, the same shared parameter of two matching equations. In other words, what we define as customer demand is actually the order of the downstream SC member. Vice versa, the order is a demand for the upstream SC member (Oliva and Gonçalves, 2005; Cui et al., 2015). Studies of Disney and Towill (2003) and Chen and Disney (2007) where they measure the amplitude of bullwhip by the variance of orders provide examples for this interaction between the demand and order. Moreover, order and demand management have also another dual role as they can be either a reason for BWE generation or a solution for its mitigation depending on the suitability of the applications. To sum, these highly interchangeable positions of order and demand increase the importance of their management and dynamics about how companies order, how the order placement decision is affected by the upcoming bullwhip, and how the order contributes to the BWE continuing through the upstream SC.

The working mechanisms of SCs were explained previously, depicting the upstream flow of orders and downstream flow of materials. It is important to determine the best ordering policy that can minimise the costs of both the inventory held and stockouts (missed orders) (Strozzi et al., 2007). Achieving this goal is easier said than done because of the uncertainties, such as unknown demand, shipping delays, and the dependence on other SC partners' decision-making and activity performances (Steckel et al., 2004).

Researchers in the field have mathematically modelled the ordering mechanism (Moritz et al., 2020). Although there are some alternative formula with different configurations according to their focal aspect, their base logic aims to accomplish three goals: (1) anchoring the order in the replacement of the stock dispatched (*DEMAND*), and then adjusting it with (2) the approximation of the desired and held inventory levels (*INVENTORY*), and lastly (3) the maintenance of a sufficient inventory flow in the supply line for the upcoming delivery periods (*SUPPLY LINE*) (Oliva and Gonçalves, 2005; Moritz et al., 2020).

Sterman (1989a,b) first modelled order placement in line with these three goals. According to Sterman's decision rule, given orders (O) are equal to deduction of in-stock inventory (S) and inventory in the supply line (SL) from the sum of expected demand for next period (CO) and the desired inventory (S') (*see* Equation 1). He enriched his formulae by adding human-related variations, such as interpretation of the demand information cues (θ) and the intuitive perception of inventory in-stock (α) and in the supply line (β).

$$O = CO+\alpha*(S'-S-\beta SL) \quad (1)$$

The first element of this formula – *Expected Order (CO)* – is calculated with a separate equation and forecasts the expectation of the decision maker, generally using the exponential smoothing technique:

$$CO = IO*\theta+CO'*(1-\theta) \quad (2)$$

where IO is the current incoming order, $CO_{(t-1)}$ is the expected order of the last period and θ is the smoothing factor (0≤θ≤1) (Sterman, 1989a,b; Oliva and Gonçalves, 2005).

This element is included in many studies similar to Sterman (1989a,b). Using this technique is critical because it determines how the decision maker interprets the demand information. For example, Gavirneni and Isen (2010) have shown that decision makers give higher attention to the more recent demands when they chase the demand. This is known as the recency effect bias and can be detected with higher empirical θ values.

Affirming the importance of demand in ordering, some researchers believe that it should be investigated more by highlighting the key roles of demand profile and the accessed information. Steckel et al. (2004), for example, searched for the impact of having more realistic demand patterns in the BDG and used three types of demand: a step-up demand function (Sterman's original), an S-shaped pattern (cyclic), and an S-shaped pattern with fixed disturbances. Likewise, Coppini et al. (2010) focused on different demand profiles (step, step with noise variance, cyclic and cyclic with noise variance) to analyse the change in the ordering performance of participants. This context-dependent nature of demand tracking led some researchers (e.g., Kleinmutz, 1993; Steckel et al., 2004; Oliva and Gonçalves, 2005; Moritz et al., 2020) to question the role of exponential smoothing in the decision rule, as it is not the sole forecasting technique. When the demand is known and stationary, researchers of this school assumed θ = 1 as the current and previous periods are the same. Starting with Chen (1999), this case has been named as base-stock policy where the given order equals the received demand of the period. Regarding cyclic demand – Coppini et al. (2010) proposed the Winters model to substitute the exponential smoothing.

Despite being context-dependent, demand and the interpretation of demand information have been regarded as key factors to analyse when placing orders to upstream SC partners.

Sterman's (1989a,b) original rule considered the other two elements in conjunction: the gap between the desired inventory level and in-stock and supply line inventory levels. For a known demand and level demand, firms can adopt leaner strategies and minimise the inventory holding cost. In other situations, however, filling the in-stock inventory up to a predefined target inventory level plays the role of buffer and absorbs the fluctuations in demand (Disney and Towill, 2003).

In addition to managing in-stock inventory, managers are also required to control the inventory in the supply line. When they order, they commonly experience a lag time between placing the orders and receiving the goods. This inbound replenishment lead time leads managers to carefully consider orders because of the delay in receipt. They need to analyse the potential situations of the following days and order accordingly.

In addition to analysing the inventory flow, Sterman's (1989a,b) decision rule also considers how this flow is interpreted by human decision makers with two intuition parameters: α as the fraction of inventory surplus (or backlog) and β as the fraction of supply line. Both factors get a value between 0 and 1, and the optimal value is 1 where the decision maker takes the entire inventory in-stock and in the supply line into consideration and leaves minimum room for bias-associated mistakes. However, various empirical studies, mostly using the BDG, (e.g., Oliva and Gonçalves, 2005) revealed that managers are generally far from being optimal in these parameters and they have a tendency to underestimate the supply line inventory.

To better understand the behavioural background to managers' inventory consideration when they place orders, some researchers further analysed the concept from two perspectives. The first is the strict separation of the in-stock inventory and supply line. Oliva and Gonçalves (2005) and Moritz et al. (2020), in their studies, asserted that the supply line should have a separate desired level parameter and smoothing parameter that have no connection to the desired level and smoothing parameter of in-stock inventory. Oliva and Gonçalves (2005), Strozzi et al. (2007), Coppini et al. (2010) and Croson et al. (2014) had an alternative other perspective, which is based on the backlog reaction. These studies, in their BDG experiments, used specific parameters for backlog conditions and tested if the participants reacted differently to the backlog than to the surplus inventory.

In sum, all these attempts, starting from the first work of Sterman (1989a,b), aimed to find the most comprehensive decision rules that can explain the order placement behaviour of managers by using different perspectives. These rules covered both the operational (demand and supply) and the behavioural constructs (intuition parameters). Since the behavioural constructs are the main determinants of the differentiation among decision makers (Moritz et al., 2020), it is essential to better understand the behavioural background of decision making in OM.

2.3 Managers' Cognitive Heterogeneity

As discussed in the previous section, decision makers' behavioural characteristics have a prominent role in the mitigation of the BWE, and in demand and order management. Therefore, it is critical to understand what the term "behavioural" covers. This section will first look at behavioural science and its adaptation to SC and OM literature, then cognition and dual-process theories will be examined in detail.

2.3.1 Behavioural Operations and Supply Chain Management

a. Behavioural Sciences and Decision Making

Behavioural science is the field that aims to explain the background processes of human decision making, including the motivating factors of reducing uncertainty and achieving the best results (Franken and Muris, 2005). It is a multidisciplinary field and consequently its findings have been adopted by various other disciplines, economics being in first place, and some others such as medicine, psychology and sociology (Knemeyer and Walker Naylor, 2011).

In economics, it is believed that individuals base their decisions simply on rational grounds and in two steps. First, they identify what is the best for their interest (epistemic rationality) and then choose it (instrumental rationality). This behaviour is explained by the expected utility theory which asserts that human beings consider the existing facts, then analyse and select the best option that brings the highest utility when making up their minds (Stanovich, 2016)**.** The theory was then utilised and developed by leading scholars, in particular from the Chicago School of Economics (e.g., Milton Friedman, Richard Thaler, Leonard Jimmie Savage), and highly rational decision makers were later named 'Econs' (Kahneman, 2011). The main idea was that the human decision can be mathematically modelled and is hence predictable. However, starting from the second half the 20th century with the famous Allais paradox (Allais, 1953), another school of thought developed and some researchers (Simon, 1955, 1956, 1957, 1987; Tversky and Kahneman, 1974, 1983) discovered that decision makers behave against the rationality which is supported by the rules of logic, statistics and normative decision-making theories, such as expected utility theory (Campitelli and Labollita, 2010; Nagarajan and Shechter, 2014). Opposing the fundamentals of the consensus in economics, this opinion has become the origin of the behavioural economics.

The limit in human cognition has been discussed within the concept of 'Bounded Rationality'. First used by Simon (1957), the phenomenon is based on the idea that human beings are not capable of learning, thinking or acting in a limitless rationality; instead they opt for a satisfying solution without a need to identify the best one (Huang et al., 2013; Vandendriessche, 2017). This suboptimality is thought to be caused by the limitations in people's cognitive capabilities (Gino and Pisano, 2008), which mainly result from some biases and heuristics (Tversky and Kahneman, 1974; Ancarani, Di Mauro and D'Urso, 2013). Heuristics works as the mental shortcut tools in our problem-solving process (Kahneman, 2011; Stanovich, 2016), especially when there is not enough information to make an optimal decision and we make snap judgements. But it cannot always manage to construct the right relationship between the past information (or emotion) and the existing context, and ends up with a suboptimal or even wrong decision (Schweitzer and Cachon, 2000). These deviations from rationality are known as biases (Stanovich, 2016). It is important to note that these suboptimalities or cognitive limitations are not just random decision-making noises, but are also systematic errors. This means that field research can diagnose these errors and provide sustainable solutions for these problems (Gino and Pisano, 2008). As a first step, three main heuristics come forward in the literature, which are availability, representativeness, and anchoring and adjustment (Gino and Pisano, 2008). Availability represents the likelihood or high frequency of occurrence that can ease the recall, and hence our mind may feel closer to connect the problem context with an easy idea to recall and match. Representativeness occurs when a concept is similar to another one. A similarity is constructed between them, and then we make a parallel decision-making process and ultimate decision at the end. Lastly, anchoring and adjustment refers to any preliminary expectation (i.e., anchor) in the mind of a decision maker, or the convergence (i.e., adjust) from the optimum to the anchor. These heuristics then lead to various biases. Carter et al. (2007) present 72 different types of biases of decision makers in SCM. Among these biases, we can count the halo effect (oversimplifying the situation with emotional coherence), confirmation bias (tendency to agree with the option that is related to previous experience), base-rate neglect, past imperfect (remembering is over experiencing), duration neglect, peak-end rule, repeated exposure, denominator effect, and a tendency to have false mental images.

As seen in the examples, biases and heuristics are closely related to human characteristics, such as experiences or emotions. While researchers were switching from purely rational theories to the boundaries of human rationality, they also searched for the impacts of these characteristics on human behaviour, because Expected Utility Theory was seen as over-descriptive for the issue (Nagarajan et al., 2014). This aspect was studied under 'Prospect Theory' – proposed by the leading scholars of the discipline, Amon Tversky and Daniel Kahneman (Tversky and Kahneman, 1974) – which modified the expected utility theory from two aspects. First, people's value perception –therefore their decision – may change depending on their risk management behaviour. Their preference to seek or avoid the risk depends on the value direction of the situation according to a reference point, in other words winning or losing (Frederick, 2005). Daniel Kahneman (2011) explains this revision with a clear example in his popular book 'Thinking Fast and Slow'. Utility theory presumes the marginal utility of wealth is equal for everyone. In other words, two people having £5 million pounds will have equal happiness. However, if these two people go to a casino for gambling, one with £1 million and the other with £9 million, and both possess £5 million at the end of the night, they would not have equal levels of happiness as one has quadrupled his/her wealth while the other nearly halved his/hers (Kahneman, 2011). The first person's case is obviously more desirable; however, utility theory only considers the amount, not the perception. The second distinction i.e. the Prospect Theory is related to the diminishing sensitivity principle. The decision maker's value perception depends on the risked amount's ratio to total wealth. As an example, discounting to £1,900 from £2,000 is more preferable than losing half of £200 capital, even though both amounts are equal (*see* Figure 2.2). In addition, Prospect Theory also explains the human behaviour towards the risk and loss perception. People fear loss. They see not losing as more preferable than gaining. Next, people refrain from taking risk when they gain, but prefer it when they are in a losing situation (Nagarajan et al., 2014).

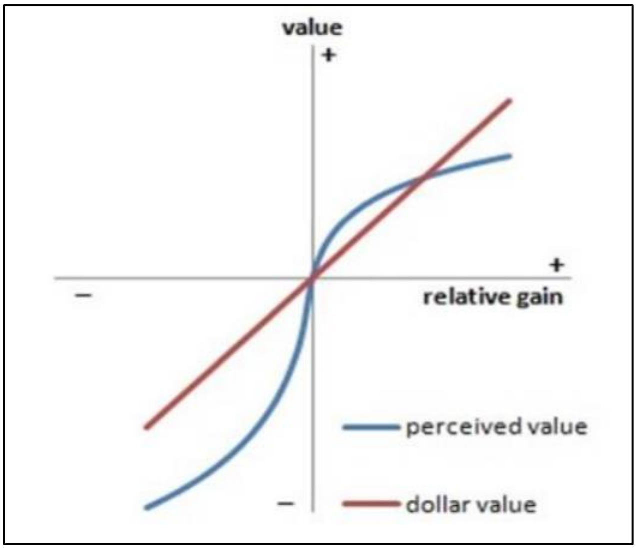


Figure 2.2 Value perception in the 'Prospect Theory' (Nagarajan et al., 2014)

b. Emergence of Behavioural Operations and Supply Chain Management

Alongside the evolution of behavioural sciences, OM was also nurtured in the 20th century. The history of operations field can be started from Frederick Taylor's time-motion studies in the beginning of the 20th century (Gino and Pisano, 2008). From that day forward, operations have evolved from various perspectives, such as environment, nature of operations, tools and infrastructure (Gino and Pisano, 2008). However, the only component that has remained constant is people, i.e. employees, operators, stakeholders, managers, investors and customers (Gino and Pisano, 2008). People have always been at the centre of the operational developments, as the practitioners of those innovations (Gino and Pisano, 2008). Their perceptions and impacts have been a moderating factor for the success of the application of these developments (Narayanan and Moritz, 2015). For example, the success of an SC has often been based on the decision quality of the managing individuals (Narayanan and Moritz, 2015). Hence, some researchers (*see* van Riel, Ouwersloot and Lemmink, 2003) in the field have mentioned the importance of humans by describing them as a resource that may lead organisations to success. Yet, that perspective did not turn into a trend in the OM discipline (Cantor, Blackhurst and Cortes, 2014). By utilising a saturation-importance evaluation, Wieland, Handfield and Durach (2016) showed that the '*people*' side of SCM had been the most understudied aspect of the discipline (Schorsch, Wallenburg and Wieland, 2017).

This deficiency in academic interest actually has an explanation coming from more than half a century ago: most of the traditional theories adopted in operational management (OM) studies have accepted that all human resources are fully rational and behave rationally. In other words, managers are supposed to possess unlimited decision-making ability and therefore become capable of evaluating and differentiating the signal and noise, to utilise information while detecting and discarding anything irrelevant, and to identify and consider alternatives and variables in decision-making processes in accordance with the suggestions of the theories (Gino and Pisano, 2008; Knemeyer and Walker Naylor, 2011). However, it is hard to implement these ideas in reality. Traditional decision models are often underutilised by the people in the system, either a decision maker or an operator (Steckel et al., 2004; Schorsch et al., 2017)**.** Knemeyer and Walker Naylor (2011) explain that human-related anomaly with a very simple example. They contemplate a situation where two SC partners gain more profit through the help of economies of scale and collaborating in transportation and warehousing assets. However, a manager of one side ends the contract because s/he considers that the other party gains more than his/her own firm. Although there was a triple advantage for the two companies and the SC, the biased fairness perception of that manager led to the failure of an optimal process. Considering the wide range of feelings that affect human decisions, such as risk perception, impulsiveness, arrogance, or confusion, we can easily extrapolate the impact of human into the management of operations and the

SC. This gap between the theory and its application puts forward the equal importance of the 'human' asset and the operational resources (Schorsch et al., 2017).

In summary, the need for the behavioural perspective can be expressed in two interconnected perspectives. First, more research on the discipline can explain regarding failures in the operational solutions (Bendoly, Donohue and Schultz, 2006) and hence create more comprehensive theories (Carter et al., 2015). So, enrichment of the behavioural side of the field is important to produce further opportunities and create better performing SCs (Schorsch et al., 2017). Second reason to raise the interest in the behavioural SCM is the scarcity of the studies, in contrast to its vitality for the progress of the discipline (Gavirneni and Isen, 2010; Cantor et al., 2014; Schorsch et al., 2017). More studies would result in unexpected opportunities by defining systematic solutions to detectable systematic human errors, hence industry standards can be carried forward (Gino and Pisano, 2008).

In line with these requirements for the human aspect**,** the field of Behavioural OM (BOM) has grown. BOM can be defined as a field that investigates the interaction of individuals' cognitive and behavioural attributes with the operational systems and processes, such as design, management and improvement (Gino and Pisano, 2008). Similarly, interest has also increased towards the behavioural studies in the SCM area (Bachrach and Bendoly, 2011), resulting in the occurrence of the Behavioural SCM (BSCM) field. BSCM is composed of two components: behavioural part – investigating individuals' behaviours in relation to incentives, their consciousness, and their performance in optimality – and SC, which is essentially on the coordination and integration of SC members (Schorsch et al., 2017).

In line with these definitions, various perspectives have grown in the body of BOM and BSCM (Huang et al., 2013). From a unit of analysis perspective, two main schools of thought have appeared in BOM and BSCM literature (Schorsch et al., 2017; Castañeda, 2019). Utilising cognitive psychology, researchers have observed the properties of individuals that can solve manager-centred decision-making problems, including indirect reciprocity (Kraft, Valdés and Zheng, 2018), and consumers' prosociality, altruism, reciprocity, informed rationality and uncertainty (Bendoly and Swink, 2007). But human behaviour is not limited by its own decisions but impacts on relationships as well, which comprise a great part of OM and SCM disciplines. Hence, the disciplines have also benefited from social psychology and analyses of the social systems that influence operations (Gino and Pisano, 2008; Schorsch et al., 2017). In this aspect, inter-personal, inter-group and inter-firm relationships have become the base of researches which have specifically observed the effects of social concepts such as trust (Özer, Zheng and Ren, 2014), power, persuasion, organisational culture, and commitment (Tsanos and Zografos, 2016).

In line with these opinions, Schorsch et al. (2017) provide a comprehensive study where they separate the BOM and BSCM fields with this unit of analysis perspective. This work creates a guiding framework for the BSCM studies (*see* Figure 2.3a). According to this framework, a BSCM study emerges from the behavioural context, which is essentially the intersection of an operational inefficiency and human intervention. Then, the context is investigated from the decision makers' psychological point of view. It is also critical to demonstrate these factors to highlight their coherence with the dominant two-view perspective of the field (*see* Figure 2.3b). Studies are also affected by moderators which can be defined as the elements that are not directly related to the decision makers but can impact on their decision. Finally, the outcomes are achieved, and this is where the BSCM differs from the psychological studies. The outcomes are combined with the starting behavioural context because the ultimate aim is not to define the situation, but instead to provide a continuous development by creating a new and better behavioural context.

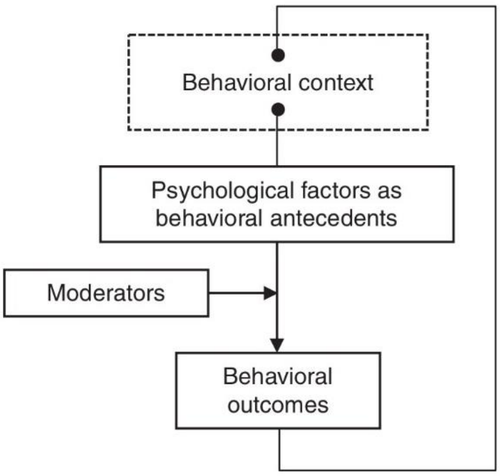


Figure 2.3 (a) Taxonomy of potential variables of BSCM discipline (Schorsch et al., 2017)

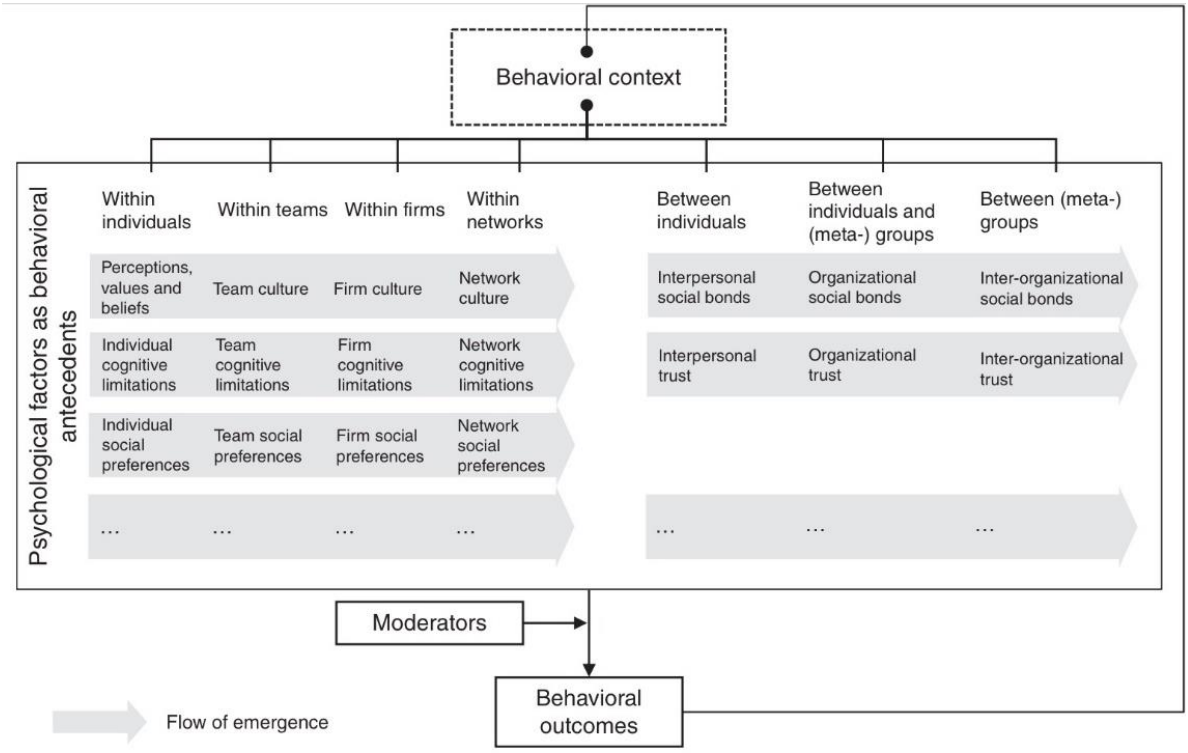


Figure 2.3 (b) Psychological factors in behavioural antecedents (Schorsch et al., 2017)

Studies of BOM and BSCM can be categorised from a problem-specific perspective as well. Although the behavioural lens is adopted to bring solutions for various problems (e.g., procurement and purchasing, contract design, cultural context, consumer behaviour), two distinct opinions appear in the BOM and BSCM literature (Gurnani, Ramanchandran and Ray, 2014; Schorsch et al., 2017). The first area studies inventory-related problems that are caused by stochastic demand. Behavioural investigations of the BWE are included in this group (Disney and Towill, 2003). Gino and Pisano (2008) demonstrate in their systematic literature review that inventory management (particularly the BWE phenomenon) occupies an important place in BSCM fields, composing a third of the studies. The second category, on the other hand, focuses on why firms fail in SC coordination (Gurnani et al., 2014; Schorsch et al., 2017).

Despite all this research, it is important to mention that the field is still in its infancy (Gino and Pisano, 2008; Schorsch et al., 2017). Even though the interest in and attempts towards the behavioural studies started half a century ago, it has been regarded as a buzzword for a long time and only started to develop after the millennium (Gavirneni and Isen, 2010). Therefore, there is still a substantial need to carry OM and SCM fields further.

2.3.2 Dual-Process Theory

a. Cognition

Besides behavioural sciences, bounded rationality and subsequent emerging concepts around human behaviour have also impacted on another neighbouring field, namely the cognitive sciences (Chater et al., 2018). Cognition is accepted as a multidisciplinary field and is contributed to cognitive psychology, cognitive science, social psychology, cognitive neuroscience and behavioural sciences. This broadness brings richness to the cognition context, but at the same time a potential complexity due to the difficulty of coordination among the disciplines. Therefore, throughout the history of research related to cognition, various definitions and terms can be found. Beyond all the complexities, 'cognition' can be simply defined as the sum of mental activities and structures that are utilised in information acquisition and processing (Helfat and Peteraf, 2015). These activities may include all potential predecessors of human behaviour, such as attention, perception, interpreting, deciding, analysing, problem-solving, biases, incentives and sentiments (Gino and Pisano, 2008; Kosmol, Reimann and Kaufmann, 2018). Analysing the structures, however, is not straightforward because of the various terminologies used in the field, such as cognitive ability, mental processes, structures, cognitive maps, scheme, schemata and mental models. Some terms have been used to identify the same concept by the representatives of different disciplines. The most prominent example of this utilisation is the description of thinking systems. Various combinations of two sets of words are used to define the same concept by different authors: "cognitive, decision-making, thinking, mental" and "styles, systems, models, processes and structures" (*see* van Riel et al., 2003; Frederick, 2005; Toplak, West and Stanovich, 2011; Chater et al., 2018). On the other hand, sometimes a single term may also be used to define two separate phenomena. For example, Kosmol et al. (2018) and Kaufmann, Carter and Rauer (2016) utilise the *cognitive map* concept to describe the cognitive structures or lenses that filter individuals' perception of the external environment, whereas the same term is used to define the spatial regions of our brain in neuroscience (Epstein et al., 2017).

This study will focus on the second category – thinking systems; however, to bring internal clarity and be consistent with the managerial cognitive studies, it is useful to highlight the taxonomy of the terms discussed so far, as is illustrated in Figure 2.4:

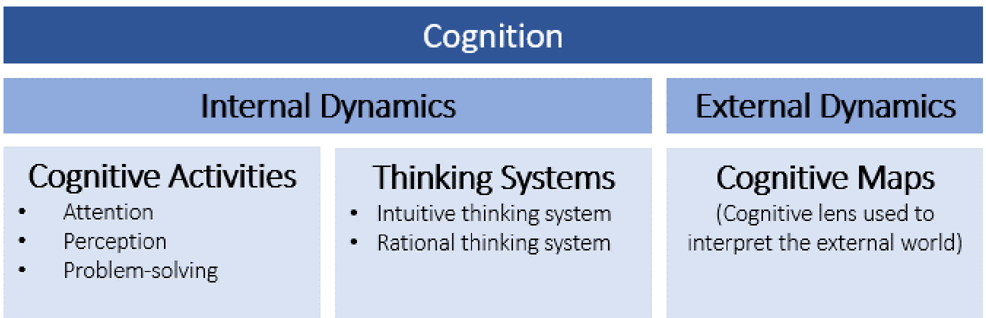


Figure 2.4 Taxonomy of concepts in managerial cognitive studies

b. Dual Thinking Systems

When we face a problem, we utilise various computational mechanisms in our brains (Toplak, West and Stanovich, 2014). We use the term 'thinking systems (or styles)' to identify the cognitive means that are utilised to understand and solve these problems (Ettlie et al., 2014). Cognitive and decision-making literature has identified several thinking styles over time (van Riel et al., 2003); however, simultaneous with the 'Great Rationality Debate', a consensus has emerged on the distinction of two distinct thinking systems (Chater et al., 2018), which was first asserted by Herbert Simon (1987) (Kaufmann et al., 2017). Though it is the start of this idea for the cognitive field, the debate on conflicting but also complementary forces of the human mind has actually been long discussed in social sciences. Nisula (2016) greatly exemplifies the depth of this idea with Socrates' quote from Plato's 'Phaedrus' (370 BC): *"In every one of us there are two ruling and directing principles, whose guidance we follow wherever they may lead; the one being an innate desire of pleasure; the other, an acquired judgment which aspires after excellence."* and Friedrich Nietzsche's Dionysian-Apollonian dichotomy from 'The Birth of Tragedy'. Therefore, as an interesting topic to pursue since ancient times, this two-system formation of human cognition has been in the focus of cognitive and behavioural studies. Moreover, it has even raised public interest with some recent popular press such as the "Thinking Fast and Slow" of Daniel Kahneman (2011) and "Blink" of Malcolm Gladwell (2005) (Evans, Handley and Bacon, 2009; Moritz, Siemsen and Kremer, 2014). However, as a social construct, the idea has always been open to interpretations, and various analysis models have emerged over time to analyse the dynamics of these two systems (*see* Wason and Evans, 1975; Epstein, 1994; Sloman, 1996; Chaiken and Trope, 1999; Stanovich, 1999; Stanovich and West, 2000; Kahneman and Frederick, 2002; Dane and Pratt, 2007; Evans, 2007, 2008; Stanovich, 2009; Evans, 2010; Evans and Over, 2013; Healey, Vuori and Hodgkinson, 2015; De Neys, 2017; Grayot, 2020), which have been collected under the title of '*dual process (systems) theories*'.

Dual process theories are fundamentally based on the separation of two main processes in decision making. Although this separation might have been under various titles, such as Ettlie et al.'s (2014) linear and non-linear thinking styles, the most common separation to describe these thinking systems has been the terms of intuitive and rational thinking systems, which was first distinguished by again Herbert Simon (1957) (van Riel et al., 2003; Stieger and Reips, 2016). The former is referred to being the 'automatic' response system and regarded as capable of giving quick reaction and response to external circumstances. The latter, on the other hand, is known as the 'controlled' mode or 'executive function' and approaches with a more detailed and slower examination to the problem (Kahneman and Egan, 2011; Kaufmann et al., 2017; Stupple et al., 2017). These thinking systems are also commonly called System (Type) 1 and 2 respectively after Stanovich and West's (2000) labelling (Frederick, 2005). This thesis will also utilise both intuitive-rational dual and associated labels interchangeably.

It is also important to emphasise the ongoing development of the thinking systems topic with the help of constructive discussions in the field (*see* Evans and Stanovich, 2013; Damnjanović et al., 2019). One school of thought, including researchers who had also built the fundamentals of the dual-process theory itself, mentions that the term 'dual-process' may be misleading as there may be more than two systems that engage in our decision-making processes (Kaufmann et al., 2017). Evans (2010) also acknowledges that these opinions have support from neurosciences via brain scan imaging studies. Among those opinions, Stanovich (2009, 2011) asserts the division of rational thinking system into algorithmic and reflective systems. While the former is associated with intelligence, the latter refers to the capability of reaction to take over the decision-making process. Some research (*see* Kaufmann et al., 2017) proclaims a change of aspect in the intuitive thinking system that diversifies the experienced and emotional intuitive processes. The same authors, in another publication, also conceptualise the common-sense-like automatic response as a separate system from intuition. However, because of the unsettled state of these attempts, this study has stuck with the standard diversification of dual-process theories.

I. INTUITIVE DECISION MAKING

The term 'intuition' is used to define the hunch-based judging mechanism of our minds, which emerges from the synthesis of the context information and prior experiential knowledge (Ettlie et al., 2014). In the course of daily life, people's decision may depend frequently on their intuitions, and may end up better or worse. The source of this intuition may be a temporal psychological situation, established beliefs such as religion or paranormal things, moral values, or familiarities such as technology use (Bialek and Pennycook, 2018). Moritz et al. (2014) refer to, from the studies of Bonabeau (2003) and Mauboussin (2009), a survey conducted by the CTF Group among business executives indicates that intuition is preferred more in decision making than facts and figures by nearly half the respondents. This is indeed normal, but it is only understandable after learning what the norms are for both thinking systems.

The literature highlights six interrelated characteristics of intuition: unconsciousness, effort miserliness, speed, holism, self-induction, and subjectivity (van Riel et al., 2003; Ettlie et al., 2014; Moritz et al., 2014; Stanovich, 2016). First, the intuitive thinking process frequently occurs **automatically** – in other words unconsciously or subconsciously. Individuals cannot control or amend the depth or direction of the thinking process, although they are capable to suppress the intuitive decision *a posteriori.* Requiring low or no level of consciousness, intuitive decision becomes almost **effortless** for our minds. This effort term covers the cognitive power spent and also the attention given. Because of the absence or scarcity of the consciousness and effort, the thinking process happens **instantaneously**. However, these cannot make intuitive decisions deficient, because these are not weaknesses but strengths of human mind. It can quickly and subconsciously scan the cognitive memory to **holistically** review the previously obtained and stored information. Of course, it is not possible to review all the information we have for each decision and it is apparently in contrast to the speed of intuition. This review of the intuitive decision-making process, instead, covers a specific storage area that has a simplified and quickly retrievable version of the complex and chaotic knowledge network of our mind – cues such as the frequently asked questions or hotkeys in computer usage. Existing research conceptualises two main components related to that storage, which shape our decisions: experiences (justified intuition) and gut feelings, hunches or emotions (creative intuition) (Kaufmann et al., 2017). During the browsing process, our mind searches for a past event in that store that shows similar pattern and then tries to provide an applicable solution. Occasionally, our intuitive decisions may also be based on our feelings about the situation (Carter et al., 2017). The outcome of the decision coming from either experience-based or emotion-based intuition can be positive or negative. However, our intuitive thinking system – thanks to its **self-induction** characteristic – continuously replenishes itself using these outcomes. Every single experience updates our storage – either empowering the existing cue or revising it after a negative outcome. The intuitive thinking system can relate different experiences from seemingly unrelated domains and produce a better and unexpected solution for the existing problem. The last aspect of intuition is its **subjectivity**. Since its database specifically depends on an individual’s personal experience and characteristics, intuitive decisions cannot be subject to any validity or justification discussions in contrast to the rational thinking system, which will be explained in the next section. Moreover, because of the intuitive thinking system’s subconsciousness, even describing the decision-making process and making deliberate alterations to it become troublesome for the individual.

II. RATIONAL DECISION MAKING

As well as individual decision makers using the intuitive thinking system, they can also use the rational thinking system (Bialek and Pennycook, 2018). Simon (1978) identifies rationality as the collection of processes that are related to defining relevant decision criteria, collecting information, and analysing context to reach the best possible decision (Kaufmann et al., 2017). This definition is in line with the mainstream aspect of rationality in the cognitive science, which is based on two rationality types: epistemic and instrumental. While epistemic rationality is interested in what we define as ‘rational’, instrumental rationality is focused on the means of rational decision making (Stanovich, 2016). Subsequent implications of the decision indicate the rationality of a decision from the epistemic aspect. In other words, it may be straightforward to define the epistemic rationality of a decision by benchmarking the results of a decision with the initial goal of the decision maker. However, the steps of the rational decision-making process lack clarity and raise more interest in the field (Helfat and Peteraf, 2015). Therefore, researchers are still interested in and working to define the exact procedure of our decision-making process; however there are some cognitive functions that have been agreed upon to be related to the rational thinking system. Among those are logic and reasoning, abstract thinking, problem solving and purposeful planning (Helfat and Peteraf, 2015).

Another aspect on which the field researchers have a consensus is the common characteristics of rational thinking system, which are conscious control of decisions, constructive, slow, computationally expensive, objective and intentional (*see* Stanovich and West, 2000; van Riel et al., 2003; Moritz et al., 2014; Helfat and Peteraf, 2015). To briefly explain these concepts: first, individuals consciously have total cognitive control in the decision-making process and can decide on its depth in detail and extent in effort and time. Secondly, the decision-making process may be constructed by a number of sequential steps. These steps can be analysed, the rules of logic applied in the decision can be deduced and, following the same logic, the decision can be repeated by another individual. This conscious control and constructivist nature lead to the next two characteristics: slowness and need for high effort (Stanovich, 2016). Next, besides this transparency and systematic procedure, this thinking system gains objectivity by filtering or neutralising the potential effects of subjective factors such as emotions and experiences. Lastly, rational decision making is an intentional activity. As well as including a deliberate and active control on the decision-making process, this term also refers to setting the decision on a goal. Individuals possess a goal when rationally thinking, plan their decisions accordingly and define the most suitable outcome for the target by comparing the alternatives.

It is also important to highlight the difference between a rational thinking system and other cognitive abilities such as intelligence, as the former is a cognitive disposition and thinking style while the latter is the indicator of a cognitive ability (Stanovich, 2016). Extant literature (Evans, 2007; De Neys and Glumicic, 2008; Pennycook, Fugelsang and Koehler, 2015) exemplifies this difference by showing that individuals with both high and low cognitive abilities are equally susceptible to relying on the biases. However, due to the requirements of the functions, such as abstract thinking or problem-solving, cognitive abilities are utilised more effectively when individuals are using their rational thinking systems (Bialek and Armurat, 2018).

III. INTERRELATIONSHIPS BETWEEN SYSTEMS IN THE THEORY OF DUAL PROCESS

For a long time, intuitive thinking systems were identified with lower features such as obscurity and subjectivity, and even considered as more primitive compared to the rational thinking system (van Riel et al., 2003). This prejudice was mainly caused by the role of intuition in evoking biases in comparison to the rational thinking system's role to suppress them (Helfat and Peteraf, 2015; Bialek and Armurat, 2018). This has resulted with an imbalance in the topic as much of the research focused on the negative impacts of intuition, leaving the positive effects neglected (Helfat and Peteraf, 2015; Kaufmann et al., 2017).

However, increasing interest in the field and the amount of research have remunerated the intuitive thinking system. As the opinions that see it as a primitive and inferior side of the duo were eliminated, many studies have put forward its positive impacts for decision making (*see* Kahneman and Klein, 2009; van Riel et al., 2003; Kaufmann et al., 2017). The critical point to remember here is that human beings aim to produce the best decisions for every situation that suits the norms of rationality, no matter what thinking system they utilise consciously or unconsciously (Campitelli and Labollita, 2010; Alós-Ferrer, Garagnani and Hügelschäfer, 2016; Chater et al. 2018).

Two thinking systems are inevitable pieces of our cognition because of the trade-offs between them (Evans and Stanovich, 2013). In the Subsections 2.3.2 (b-I,II) (*p. 31*), two thinking systems were explained through the lens of five main characteristic aspects: consciousness, control, effort, approach and validity. However, there are further differences between the features of these systems, and they can be better demonstrated in a comparative display (*see* Table 2.4).

Table 2.4 Comparison of main features of dual thinking systems (*adapted from* van Riel et al., 2003)

	Intuitive Thinking System	Rational Thinking System
Consciousness	Subconscious/Unconscious	Conscious
Control	Automatically triggered	Deliberate
Effort	Low	High
Approach	Holistic and inductive	Focused and deductive
Validity	Subjective	Objective
Orientation	Affective: Pleasure-pain oriented (What feels Good)	Logical: Reason oriented (What is Sensible)
Processing	Parallel	Serial
Structure of knowledge	Associationist connections	Logical connections
Effect on behaviour	Behaviour mediated by 'vibes' from past events	Behaviour mediated by conscious appraisal of events
Form of information encoding	Encodes reality in concrete images, metaphors, narratives	Encodes reality in abstract symbols: words and numbers
Adaptivity	Resistant to change	Changes easily with the strength of argument
Detail level of result	Broad generalisation	Fine detail
Justification	Self-evident valid: 'Experiencing is believing!'	Requires to be justified via logic and empirical evidence
Errors	Normally distributed	Few, but large
Self-confidence in answer	High	Low
Confidence in method	Low	High

Among all these differences, the biggest trade-off in the interaction of dual thinking systems is the choice between cognitive power and cognitive expense. While System 2 has the capability to solve various problems with a high ratio of success, it requires more effort and attention, and processes slowly. On the other hand, System 1 requires much less attention and effort, solves the issue instantaneously; however, the result is not always as dependable as that of System 2 (Toplak et al., 2014).

To reap the most benefit from this trade-off between the thinking systems – which has been previously described as the ultimate aim of our collective mind – our mind applies a reciprocal complementarity between these two systems. Similar to the unbalanced approach to the intuition explained so far, many studies also incorrectly focused on the systems separately or looked at them as the two poles of a spectrum. However, a mere intuitive decision making may be regarded as too risky, whereas utilising only rational thinking system in decision making would not be fast enough

(Carter et al., 2017). In line with this example, the accumulation of research in the dual-process theories for the last half century refuted the standalone evaluation aspect by gradually discovering further interactions between the two systems (Kahneman, 2011; Moritz et al., 2014; Narayanan and Moritz, 2015; Alós-Ferrer et al., 2016; Stanovich, 2016). So, we can say that the two thinking systems work in harmony without any conflict (Chater et al., 2018).

At this point, van Riel et al. (2003)'s comprehensive explanation can be utilised to understand this interaction. Authors extend Hammond and Brehmer (1973)'s Cognitive Continuum Theory which basically puts the cognitive thinking systems into a continuous scale with the help of a two-dimensional matrix instead of regarding them as opposite poles (*see* Figure 2.5).

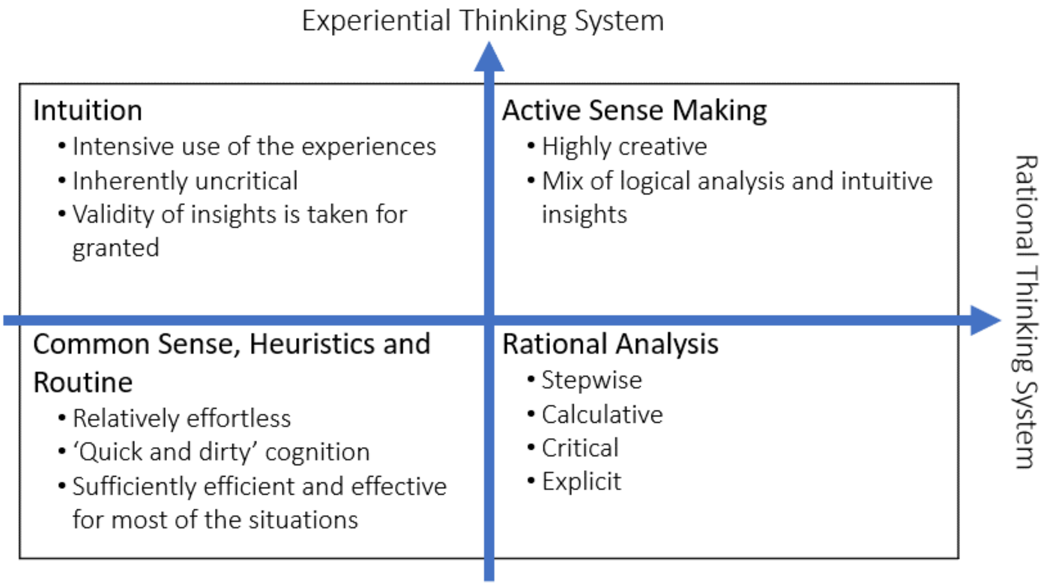


Figure 2.5 Cognitive style matrix (*adapted from* van Riel et al., 2003)

To follow the suitable cognitive style, dual thinking systems interact with each other in a sequence. This starts with the automatic triggering of the intuitive system. Being a low-effort system, it stays in the idle mode and makes numerous swift decisions from mundane everyday tasks to one-off, more complex problems (Carter et al., 2017). The decisions at this part can be matched with 'Common Sense' decisions: effortless, quick and efficient enough most of the time (Chater et al., 2018). System 1's parallel process scheme, in comparison to System 2's serial process, also plays an important role here, as the former can work on more than one cognitive job, while the latter focuses on a single job and stays on it till it is finished before passing to the next one (Toplak et al., 2014; Stanovich, 2016). Briefly returning to the survey mentioned above, we can now understand why half the respondents utilise their intuitive thinking system more, especially considering that our mind prioritises idleness and leans towards spending the least effort possible (Kahneman, 2011).

Sometimes, the quick common-sense system may not give a satisfying result. In this instance, our mind prefers to use more resources of the intuitive thinking system and searches for an association in the experiences database. Sometimes, however, building our decisions may be an outcome of compulsory external factors. For example, although our mind faces a more complex problem that requires more attention, effort and time, we may lack any of these conditions because of either tiredness, lack of information or time constraints. In this kind of occasion, System 1 takes over the responsibility from System 2, which falls into the left-top 'Intuition' box in the cognitive style matrix (Helfat and Peteraf, 2015).

However, once again, this does not mean the output decision will be poorer and there are times that it works the other way around, i.e. where System 1 is more advantageous (Helfat and Peteraf, 2015).

Van Riel et al.'s (2003) adaptation of Goldberg's two studies (1983, 1990) provides a clear framework for the functional benefits of intuition (*see* Table 2.5).

Table 2.5 Functions of intuition (van Riel et al., 2003)

	Explanation	Key Feature
Intuitive Discovery	Goes beyond simply providing answers to a certain problem but is an insight into the real nature of the dilemma. It is likely to suddenly occur when the mind is actually occupied with something other than the problem to which the intuition is related.	Inductivity
Creative Intuition	Generates new ideas and involves alternatives and possibilities rather than facts and provable information.	Creativity, Non-linearity
Intuitive Evaluation	Leads to a feeling of preferring one alternative over the other and to a feeling of certitude that tentative conclusions are correct.	Integration of Affect
Operative Intuition	Influences and prompts individuals' actions without entering consciousness and gives a particular direction of action that is unexplainable.	Tacit Knowledge, Experience
Intuitive Prediction	Deals with the unknown, under which circumstances rational analysis is ineffective.	Synthesis, Tacitness
Intuitive Illumination	Transcends the other five functions of intuition and is associated with a higher form of knowing and resembles the more spiritual side of intuition.	Holistic Input

Besides tapping into the further benefits of experiential intuition when our common sense's output does not satisfy us, there are times when we actively switch to our rational thinking system as well. On these occasions, System 2 automatically overrides System 1 (Campitelli and Labollita, 2010; Toplak et al., 2014; Stanovich, 2016) and the literature provides three major reasons for this. First, rational thinking system can be required to override due to the complexity of a problem, where the intuitive thinking system cannot produce a quick and pleasing solution (Kaufmann et al., 2017). Second, it can be caused by a previous experience that has resulted in a suboptimal decision because of the biases of System 1. Here, System 2 interferes to counteract the previously experienced suboptimalities and either filter them out or neutralise them (van Riel et al., 2003; Carter et al., 2017; Kaufmann et al., 2017). Lastly, it can even just be, because of the essentiality of the decision, that the decision maker may prefer to assert more attention (Chater et al., 2018). As mentioned in the previous section, unlike System 1, System 2 works deliberately, and the depth and intensity can be controlled by the individual intentionally. All these situations can be located in the bottom right box of the cognitive style matrix (Figure 2.5), namely 'Rational Analysis'.

The last and the most effective cognitive thinking process, 'Active Sense Making' is realised by the full cooperation of both thinking systems. Stanovich (2016, p.27) define System 1's role here as 'bringing the response into the right ballpark'. This prevents losing time when making a decision, especially urgent ones. System 2, on the other hand, takes responsibility for a rationality filter and quickly reviews the outputs of System 1, even where sole intuitive decisions may outperform (Ettlie et al., 2014).

2.4 Decision-Making Environment

In SCM and OM, as well as in our daily lives, various decisions are taken on an ongoing basis (van Riel et al., 2003). These decisions are mostly unique (Moritz and Parker, 2018), challenging and highly dynamic as they are impacted by the personal characteristics of the decision maker; however, there is another essential element that impacts on our decision: namely the external environment (Carter et al., 2017). These forces work in cooperation (van Riel et al., 2003; Shoda, 2004; Ayal et al., 2015; Kaufmann et al., 2017). That combination was first identified by Kurt Lewin's famous equation: B =

f(P,E) (*see* Lewin, 1936) where the parameters refer to behaviour, person and environment respectively. Although Lewin's equation study is nearly a century old, its statements still have validity. Therefore, it is necessary to learn the fundamentals of both elements. A perspective of personal characteristics has been discussed extensively in the previous section within the scope of cognitive thinking styles. This section will now focus on the environmental factors that influence our decisions, especially in the context of SCM and OM.

External environment, situation, nature, or context of the problem in other words, are at least as important as personal characteristics in terms of their impact on the decision. To highlight the importance of the external environment on a decision-making process, Ettlie et al. (2014) refer to Rokeach and Kliejunas (1972), who assert that two-thirds of our decisions are majorly influenced by the environment, while personal characteristics account for only one-third. From a psychological perspective, understanding the problem context is one of the main tasks because we, as people, reconfigure our decisions according to our changing perceptions of the environment (Shoda, 2004). This importance also stands for the decisions in SC operations. Managers are expected to be equipped with an adequate judgemental capability as these decisions influence various parts of the business (Jankelová, Jankurová and Rašovec, 2018). Especially with the development of more complex and global SCs, the extent of these decisions' impact has also increased (Cantor and Macdonald, 2009).

In their comprehensive work, Schorsch et al. (2017) explore the possible external decision contexts in the SC environment. They describe their findings as the moderators in mainly three categories: environmental, structural and procedural. While the former refers to the external factors that cannot be changed by the decision maker, which are the homophily (business similarities) and the market conditions; the latter two groups stand for the external but improvable factors. Among those factors are the SC design and technology adoption maturity level for the first, and formalization, information sharing and timing for the second. Another study on the topic, by van Riel et al. (2003), looks at the cognitive aspect and proposes a conceptual model to test six external factors that may affect our decisions: the availability of information, problem complexity, problem structuredness, importance of tacit knowledge maker, time pressure and the requirement for justification. Despite the fact that all are of critical importance for the decision output, information presence, problem complexity and time pressure are selected for a deeper investigation in this study. The main reason for this selection is that these three categories closely pertain to each field and enable effective measurement in both perspectives. Besides, information presence can be a preceding construct for information sharing and the tacit information categories of van Riel et al. (2003), and problem complexity can cover the task structuredness for an initial investigation. Future studies can extend the findings of this study on the remaining contextual factors.

2.4.1 Informational Presence

Information holds an essential place in the decision-making process of individuals (Bendoly and Swink, 2007). The core activity of decision making is accepted as the transformation process of the cognitive input into the decision output, where the cognitive input represents the information intake (Simon, 1956; van Riel et al., 2003). Stanovich (2016) and Chater et al. (2018) underline the importance of information for decision making by defining the environment as benign or hostile in accordance with the availability of the cues present in the environment.

Information sharing activities and capabilities also play a key role in their performance as well as the SCs to which they pertain (Cantor and Macdonald, 2009). This is because the existence or lack of required information has a direct impact on the quality of managers' perceptions, decisions and actions (Bendoly and Swink, 2007), especially with the advent of advanced technologies that can dramatically evolve the information producing, sharing and processing capabilities. The impact of information in SCs is researched in the frame of the acquired benefits, additional risks of increased information sharing and its mechanics, as well as the sharing preferences of stakeholders, e.g. reciprocal sharing (Bendoly and Swink, 2007).

In the SC context, these aspects are also examined in subfields, such as 'Supply Chain Visibility' and 'Transparency' (*see* Barratt and Oke, 2007; Caridi et al., 2010; 2014; Fan et al., 2013; Carter et al., 2015; Nooraie and Mellat Parast, 2015; Kraft et al., 2018). These studies have aimed at extending the extant information-sharing boundaries to receive the greater benefits of having a strong SC platform. Likewise, the impact of local information on the global SC has also been discussed (Oliva and Gonçalves, 2005), especially the risk and uncertainty concepts such as changing customer trends and corporate social responsibility risks. Besides this spatial extension, the information's impact on managers' decisions can also be temporally extensive. For example, negative information about a supplier influences the implementation of the current task; however, it may also have an impact on the future relationship with that supplier.

Going beyond the sole activity of sharing and receiving information, research on the topic has also been interested in the quality attributes of information. Among these attributes, the following aspects can be counted: information's content, type and nature (Steckel et al., 2004; Kraft et al., 2018), its amount, reliability and unanimity (Ancarani et al., 2013), and its timeliness (Bendoly and Swink, 2007).

2.4.2 Time Pressure

Time pressure is identified as an essential situational factor for decision making, as the quality of decisions can be influenced when they are made relatively quickly (Hwang, 1994; Maule, Hockey and Bdzola, 2000; Thomas, Esper and Stank, 2010) and because we deal with many daily problems on which we are required to make a decision in a short time. Some examples are: an air traffic controller who needs to quickly direct the air traffic to avoid an accident and a police officer who needs to decide the right level of force used to control a violent situation; a more daily example is that of a driver on the road who approaches a junction with speed and sees that the amber light has replaced the green, within a short distance he/she needs to instantaneously decide to either stop or keep going (Ordonez and Benson, 1997).

The origins of time pressure research are based in the psychology field; however, researchers of many other fields have also noticed the importance of time management and put forward related studies. As it is closely interrelated with psychology, decision-making science has been one of those fields (Thomas et al., 2014). However, despite this interdisciplinary interest, there is no consensus on a single definition of time pressure (Hwang, 1994). In one aspect, it has been identified as the time constraint, or the imminent deadline (Thomas et al., 2010). However, this thought, which was based on time constraint or limit, has been criticised as being incomplete, asserting that time constraint alone is not the only element of time pressure (Thomas et al., 2010). Instead, it is better identified as the scarce time that is deficiently allocated for the completion of a specific task and characterised as a stressful condition (Thomas et al., 2010; 2014; Thomas, Esper and Stank, 2011b; Fraser-Mackenzie and Dror, 2011).

The literature demonstrates some positive outcomes of decisions under time pressure. For example, when individuals are under time pressure, their risk-taking attitudes attenuates and only increases when the expected value of the risk is high (Maule et al., 2000). Time pressure also increases the selectivity and process of information intake in terms of type and volume (Fraser-Mackenzie and Dror, 2011). However, as Thomas et al. (2014) identified, these are seen as 'temporary energising effects', comparing to the higher amount of negative impacts (Thomas et al., 2010). The most fundamental problem of time pressure comes with the feeling of stress and fear of not fulfilling the job in the allocated time (Fugate, Thomas and Golicic, 2012). Research specifically identifies that the quality of decisions is decreased in situations, such as negotiations, buyer behaviour, small group dynamics, and auditing performance (Fugate et al., 2012; Thomas et al., 2014). Time pressure is also associated with creating negative emotional outcomes, such as frustration, anger and low level of confidence (Thomas, Esper and Stank, 2011a). The reasons for these negative outcomes can relate to the additional cognitive load that comes with the increasing time pressure (Thomas et al., 2014). As well as the outcomes of decision making under time pressure, research on the topic has also looked into the coping strategies associated with it (Thomas et al., 2014). Among the most common are increasing the throughput by working faster (Fugate et al., 2012) and selectivity filtering only the required information intake into the decision-making process (Thomas et al., 2014).

Various subfields of business management have also had an interest in eliciting the potential impacts of time pressure, such as consumer behaviour, negotiations, retailing, and accounting (Thomas et al., 2014). SCM is also one of the fields that has close relationships to time management and definitely to the pressure of time deficiency (Fugate et al., 2012). Although the research on the impact of time pressure on SC performance is not seen as sufficient in the literature (Thomas et al., 2014) and there are calls for more studies on this aspect (*see* Kaufmann, Michel, and Carter, 2009), the extant research highlights the importance of time by referring to it as being 'of paramount concern' for SCM (Thomas et al., 2014). End-to-end management of retail SCs (Thomas et al., 2010), supplier evaluation and relationship situations where quick decisions are made by supply managers (Carter et al., 2017), are some examples of where time pressure effects are felt in SCs. However, as the SCM is tightly connected with OM, some other impact areas can be included, such as the need for quicker product innovation in competitive industries (Stalk and Hout, 1990), and time-related issues in behavioural operations (Thomas et al., 2014).

2.4.3 Problem Complexity

Similarly to time pressure, the definition of complexity also lacks a consensus among researchers because the notion of complexity differs across disciplines (Kalkancı, Chen and Erhun, 2014). Complexity is often perceived incorrectly as task difficulty. Although a complex problem may be difficult, the opposite assertion cannot be totally right because the difficulty of the task can be caused by other contextual situations, even by time pressure (Hwang, 1994). As a well-researched topic in psychology, there have been different insights and research aspects regarding what is identified as complexity. In line with this, three main perspectives are determined in relation to the complexity's impact on our decisions. Firstly, it is evaluated as a psychological challenge and investigated as an internal cognitive process. The second aspect is interested in the conformity between the task and the individual. Lastly, complexity is investigated as a characteristic of the task (Kalkancı et al., 2014).

SCM is also another field with a nature that is very prone to complexity and consequently suffers from it. Moreover, as the importance of SCM grows, it turns into a highly complex network of entities and relationships. Therefore, managing the complexity is a highly critical issue. For instance, a very fundamental step of SCs, the supplier selection process, has various steps from the analytical processing of the applicant suppliers to the assessment of relevant facts related to them. However, the only way to end up with the best supplier selection decision is done via the completion of all these complex steps, otherwise, it can result in various supply disruptions and risks (van Riel et al., 2003). Besides, the increasing interest in, and public and societal pressure of responsible and sustainable SCs, compels companies to be thorough and transparent in these processes. In another side of the SC, the management of the information and inventory flows is another set of complex processes. To maintain the right level of competitiveness, SCs must be prepared for any unexpected incidents from both their downstream and upstream sides. However, this requires extensive amounts of consideration and preparation that are all interrelated (Steckel et al., 2004).

2.5 Chapter Summary

This chapter provided an elaborate analysis of three main blocks of literature: Bullwhip and order management in SCs, behavioural backgrounds of operations and the external situations that can impact on managers' decisions. The following are the striking points of this literature review.

Being caused by the sudden oscillation of upstream customer demand and downstream inventory, the BWE is a long-standing issue in SCs and stands as a prominent problem depending on the industrial context. It is. Therefore, it still requires more research for improvement, although the literature does provide various operational and behavioural causes and solutions.

From the behavioural perspective, cognitive studies look for the mechanism of individuals' decision-making processes, and dual-process theory proposes that our decisions are produced by two cognitive thinking systems, namely intuitive and rational. The intuitive thinking system is the one that we automatically use in our quick gut-based decisions. On the other hand, the rational thinking system is intentionally used for more analytical problems that needs justification. Despite there being a flawed prejudice that accepts the intuitive system as inferior to the rational one, both systems have their own advantages. Moreover, they work in collaboration when producing the optimal decision.

Besides our cognitive characteristics and the utilisation of cognitive thinking systems, the situation of the decision-making environment also impacts on our decisions. Among those situations are information presence, time pressure and problem complexity. Information presence plays a key role in the quality of our decisions. This importance is also valid for the SC decisions as the information flow in both directions along the SC brings many opportunities for SC success and competitiveness, whereas its deficiency will have adverse impacts and risks for the SC. Individuals' decisions are impacted when they are under time pressure. Although the literature asserts some positive and mostly negative effects of reducing time, there is no consensus for it. The uncertain nature of many SC operations also requires managers to decide quickly in a short time. Likewise, problem complexity is a significant factor for a decision-making process. As SCs grow in geography and relationships, they turn into large and complex networks. Hence, managers are required to make increasingly more complex decisions.

3. Conceptual Model and Hypotheses

> *"In all affairs, it's a healthy thing now and then to hang a question mark on the things you have long taken for granted."*
>
> BERTRAND RUSSELL

3.1 Chapter Introduction

In the previous chapter, the literature has provided background information on three main fields that relate to this study: Bullwhip and order management in SCs, behavioural backgrounds of operations and the external situations that can impact on managers' decisions. This chapter will now present the hypotheses that address the gaps in the provided literature.

First, in the Section 3.2 the hypotheses will be presented in three groups: Baseline hypotheses, time pressure hypotheses and problem complexity hypotheses. Then, in Section 3.3, these hypotheses will be connected to the conceptual model that will be tested in this study, together with a summary of the hypotheses.

3.2 Hypotheses

3.2.1 Baseline Hypothesis: Dual-Process Theory on Bullwhip Effect Mitigation

In the previous chapter two key findings from the literature on cognition and behavioural SC management fields were presented. Firstly, human cognition is a crucial factor for the decision-making process. Secondly, those human factors are of critical importance for the success of a SC. There have been studies that focused on the impacts of managerial cognition on SC activities from various aspects, such as deviations from rational behaviours and cognition-related variations in purchasing behaviour (Kosmol et al., 2018). Another approach is related to the relationship between the cognitive reflection and the quality of operational decisions (Pan, Shachat and Wei, 2019), including ordering behaviour and hence the BWE management (Narayanan and Moritz, 2015).

Cognitive reflection is defined as the capability of filtering the intuitive decisions with the critical power of rational thinking system. So, it is a means of differentiating the propensity of a manager to either thinking system. Section 2.3.2 (b) (*p. 30*) has identified the distinctive characteristics of both thinking systems. Accordingly, they comprise different characteristics which will create relative advantages or disadvantages for them in relation to the features of the decision. Brent Moritz undertook a series of consecutive studies (e.g. Moritz et al., 2013; 2014; 2020 Narayanan and Moritz, 2015; Ovchinnikov, Moritz and Quiroga, 2015). These studies investigated how the cognitive differences of managers impact their order management decisions and bullwhip mitigation performance over various SC and business settings. The output of all these studies in different research contexts showed that managers' cognitive reflection levels and their deviation from optimal order management have negative correlation. In other words, the lower the cognitive reflection is the higher performance in BWE mitigation. This finding constructs the base of this study and will be analysed in different research contexts following the research path of Dr Moritz. However, to ensure the internal validity of the study, it will be adopted as the first hypothesis and will be retested:

H1: Managers with a cognitive propensity towards their rational thinking system (System 2) perform better than managers who have a propensity to use their intuitive thinking system (System 1) in tracking the customer demand to mitigate the bullwhip effect.

3.2.2 Situational Hypotheses

In the same direction as H1, the literature review chapter has shown a strong bias against the negative outcomes of the intuitive thinking system and its inferiority to its counterpart, the rational thinking system (Kaufmann et al., 2017) (*see* Subsection 2.3.2 [b-III], *p. 33*). However, there have also been some studies refuting this prejudice and highlighting the essential role of intuitive decision making in a collaborative cognitive decision-making process to produce the most optimal decision.

This is also valid in the SC management context. For example, Katok (2011b) explains how intuition can end up with a positive result in supplier selection and relationships. Likewise, as Ancarani et al. (2013) have quoted from Gigerenzer (2002), SC managers may utilise intuitive thinking system in their decisions as a means to 'live with risk'. Moreover, it is actually neither wise nor possible for SC managers to consistently use their rational thinking system for all decisions, due to the changing dynamics and increasing complexity of the SC operations and problems (Carter et al., 2017). There is still ambiguity in the role and importance of intuition in SC decision making which brings requirements for further research on the topic.

One aspect that can bring clarity to the role and process of cognitive decision making in SC operations is the investigation of situational factors. Campitelli and Labollita (2010) and Ayal et al. (2015) assert that the quality of a decision or the level of its deviation from optimality are not only related to the extent of either thinking system's utilisation, but also to the characteristics of the task and environment. Likewise, for the SC context, Schorsch et al. (2017) provide a highly beneficial work in the area and they highlight the importance and role of situational moderators in BSCM studies. The main situational factors that are associated with cognitive activities in SCM have been collated in Section 2.4, which later have been discussed regarding three main factors: information presence, time pressure and problem complexity.

This study aims to identify the gaps in the managerial cognition research by investigating the potential impacts of these situational factors on the bullwhip mitigation performance of SC managers with distinct cognitive propensities. The rest of this section will elaborately discuss these relationships for each situational factor.

a. Information Presence in Bullwhip Effect and Behavioural Studies

In the literature relating to cognitive sciences and SC management, information is an invaluable asset (*see* Section 2.4.1, *p. 37*). It has also been in the centre of BWE studies since Forrester's (1958) seminal work. As explained in Section 2.1, the lack of information and the uncertainty accompanying it were considered one of the major causes of the bullwhip phenomenon, while more information access and increased transparency are counted among the major bullwhip mitigation strategies (Lee et al., 1997b; Coppini et al., 2010). Companies' access to downstream information can improve company performance by increasing forecast accuracy, decreasing safety stock and SC costs, and becoming more dependable for customers (Cui et al., 2015; Narayanan and Moritz, 2015). Likewise, information is placed at the heart of the decision-making process, which is frequently identified as the activity of transforming the knowledge and information into managerial action. Therefore, the presence or lack of information directly impacts on the characteristic of the problem and hence the decisions of individuals (van Riel et al., 2003).

The provision of increasing amounts of information may not always be beneficial to decision-making processes. For example, Steckel et al. (2004) studied the impact of POS information on SC performance, and discovered that the impact depends on the nature of the demand. This was later affirmed by Cui et al. (2015) who demonstrate the variation of information benefits according to the demand characteristics. Moreover, the presence of information may not always be helpful at all as it may divert the individual from an optimal decision due to the latency of information access, or missing, inaccurate or misleading information (Bendoly and Swink, 2007). Trust issues may also limit the beneficial impact of information sharing. Due to coordination risk, managers may ignore the information or be cautious about the accessed information (Croson et al., 2014). Lastly, this sceptical look is utilised in the bullwhip context. In the experiment where Zhao and Zhao (2015) compared the impact of information in a multi-echelon inventory system, they found that complete information sharing is not a guarantee of better performance and that the BWE continues to appear even when there is access to more information. Therefore, instead of taking its positive influence for granted, it is better to assess the impact of information presence in varying SC dynamics.

The SC configuration context that this study aims to identify is based on the cognitive propensity of SC managers. Although previous literature has demonstrated the performance differences of managers with different cognitive reflection levels, this study asserts that that impact should also be investigated, together with the presence of information. In their conceptual work, van Riel et al. (2003) defined information as the building block of the rational thinking system and proposed a hypothesis to find out how additional information enhances the quality of decisions. Therefore, the second hypothesis of this study believes that this impact should also be valid in the SC context, specifically in the endeavour to mitigate the BWE.

> *H2a:* Managers with a cognitive propensity for their rational thinking system (System 2) perform better in tracking the customer demand to mitigate the bullwhip effect when they are provided with additional information.

Research that is interested in the impact of information on intuitive thinking system, on the other hand, mainly focuses on the idea that the intuition system is utilised more in uncertain environments which can be caused by the lack of complete, accurate and timely information (van Riel et al., 2003; Carter et al., 2017). The intuitive system's reaction to additional information in the environment, however, has not been the focus of any discussion. This is understandable as intuitive and rational thinking systems have overall opposite characteristics, as explained in Section 2.3.2 (*p. 30*). While the rational system's input is the available information on the problem context and more information can result in better decisions, intuitive thinking system utilises the internal information that had already been experienced by the decision maker and coded as an intuitive shortcut in their mind. Therefore, this study proposes that the newly acquired information should not have an impact on the bullwhip mitigation performance of the SC manager.

> *H2b:* Being provided with additional information does not impact the demand tracking performance of managers with a cognitive propensity for their intuitive thinking system (System 1).

In addition to these two main hypotheses related to information presence, this study will also analyse how the impact of information changes in other environmental conditions. The main reason for this additional step is related to the difference of information presence from other conditions. Besides being a situational factor, information is also an operational solution to mitigate the BWE. Therefore, situation-dependent changes in the information utilisation capability of managers with a different cognitive propensity will also be an important step in mitigating the BWE. Moreover, since individuals make their decision in extensively dynamic situations, this perspective will also increase the external validity of the study by bringing the context closer to real life.

b. Time pressure in the Bullwhip Effect and Behavioural Studies

Section 2.4.2 (*p. 38*) introduced the concept of time pressure and other related terms, such as time constraint, its positive and negative impacts on decisions, and finally the SC areas that time pressure can influence. Following the work of Thomas et al. (2010) in which they investigated the unfavourable impacts of time pressure on supplier selection in SCs, this thesis assumes that similar effects can also be seen in the order management decisions of SC managers. This is because the key factor to consider here is not the scarcity of time alone, but its scarcity compared to the time required to evaluate the inputs to create an optimal output.

Availability or scarcity of time allocated for a decision task has a significant influence on the utilisation of dual thinking systems, as the time spent is one of the core characteristic differences between them. Since the rational thinking system is slower and requires more time to produce a solution, it cannot cope with the pressure of time, and thus the quality of output decisions diminishes (Evans et al., 2009; Carter et al., 2017). Therefore, managers who have a propensity to favour their rational thinking systems may face negative consequences when time is limited. The intuitive thinking system, on the other hand, is utilised more when there is limited time for a decision. As mentioned in Section 2.3, this system is subject to bias as a result. Therefore, when time is a limitation, both systems may fail to make a valid decision.

But it is worth remembering that there are some positive consequences of time pressure on decision making as discussed (*see* Section 2.4.2, *p. 38*). The study by Evans et al. (2009) also indicates that time pressure does not always catalyse the impact of biases, where the authors put forward that belief bias is observed disregarding the time pressure that participants experienced. Additionally, self-inductivity of the intuitive thinking system can also suppress the emergence of certain biases as the systems learns and replenishes from the mistakes, which are biases in this context (Furlan, Agnoli and Reyna, 2016; Carter et al., 2017). By this way, managers with cognitive propensity to intuitive thinking system may produce better decisions as they are experienced in utilising their intuition (Carter et al., 2017).

This study proposes that the repetitive and continuous nature of the order management decisions may activate the self-inductivity feature of the intuitive thinking system and managers can learn from their bias-led mistakes. Therefore:

> *H3a:* Managers with a cognitive propensity for their intuitive thinking system (System 1) may perform better than managers who have a propensity for using their rational thinking system (System 2) in demand tracking, when there is time pressure in the decision-making environment.
>
> *H3b:* Time pressure improves the demand tracking performance of managers with a cognitive propensity for their intuitive thinking system (System 1).
>
> *H3c:* Time pressure aggravates the demand tracking performance of managers with a cognitive propensity for their rational thinking system (System 2).

Time pressure also impacts on the utilisation of information in the decision-making process by increasing the selectivity of decision makers in terms of type and volume (Fraser-Mackenzie and Dror, 2011) and decreasing the information flow and processing (Thomas et al., 2010). In line with this, van Riel et al. (2003) propose that time pressure hinders the utilisation of both the rational thinking system and information utilisation. On the other hand, authors do not expect any influence on the information utilisation of the intuitive thinking system. As the impact of information has not already been expected in the first instance, in *Hypothesis H2b,* this study asserts further hypotheses in parallel with these ideas:

> *H3d:* The effect of additional information presence on System 1 managers' demand tracking performance (managers with a cognitive propensity for their intuitive thinking system) is not impacted by the time pressure in the decision-making environment.
>
> *H3e:* The effect of additional information presence on System 2 managers' demand tracking

performance (managers with a cognitive propensity for their rational thinking system) is aggravated by the time pressure in the decision-making environment.

c. Problem complexity in the Bullwhip Effect and Behavioural Studies

SC management is a completion of processes, which requires a balance between the demand and supply flows. While successfully matching these two elements is difficult on its own, it becomes harder with additional complexities, such as the delays and disruptions arising in both the supply and demand sides, and a company's dependency on the management quality of other SC partners (Steckel et al., 2004). In addition to these, changes in the external environment, such as the fast development of disruptive technologies and demand volatility related to evolving consumption trends, also increase the ambiguity of the decision-making environment and complicate the decision-making process (Kaufman et al., 2017).

van Riel et al. (2003) investigate the relationship between cognitive thinking systems and the problem complexity. Initially, they propose a positive bond with the complexity of the problem and the characteristics of the rational thinking system, such as the higher computational processing power, higher cognitive control and awareness. Although this bond is self-evident, the literature does not provide similar clarity for the relationship between the intuitive thinking system and problem complexity. As well as van Riel et al. (2003), Carter et al. (2017) emphasise the interconnectedness of both thinking systems, where the intuitive thinking system has the role of bringing the response into the right ballpark in order to ease the rational thinking system's problem-solving process (*see* Subsection 2.3.2 [b-III], *p. 33*). However, the same resources and some more (*see* Frederick, 2005) also mention the drawbacks of pure intuition utilisation for such problems that require the opposite features of the intuitive thinking system. Therefore, this study believes that when managers have preference for their intuitive thinking system, they may surpass the collaboration with the rational thinking system and finalise the decisions with their gut feelings. As discussed, this may result in suboptimal decisions, considering the problem is more complex and have many aspects to consider carefully. Therefore, we initially propose the following hypotheses for the bullwhip mitigation performance of SC managers, when they face more complex problems:

H4a: Managers with a cognitive propensity for their rational thinking system (System 2) perform better than managers who have a propensity for their intuitive thinking system (System 1) in demand tracking, when the nature of the problem is more complex.

H4b: Problem complexity aggravates the demand tracking performance of managers with a cognitive propensity for their intuitive thinking system (System 1).

H4c: Problem complexity improves the demand tracking performance of managers with a cognitive propensity for their rational thinking system (System 2).

The last point to take into consideration is the relationship between information presence and problem complexity. While complexity increases the amount of uncertainty in the environment, information acts in the opposite way. Therefore, the increased clarity of the environment may be diminished by the additional environmental or task-related complexity. While this may harm information utilisation of a manager with a propensity for their rational thinking system, it will not have any additional impact on the performance of managers with an intuitive propensity because of their inherent preference towards lower utilisation of environmental information. Consequently, this study proposes these final hypotheses:

H4d: The effect of additional information presence on System 1 managers' demand tracking performance (managers with a cognitive propensity for their intuitive thinking system) is not impacted by the increase in problem complexity.

H4e: The effect of additional information presence on System 2 managers' demand tracking performance (managers with a cognitive propensity for their rational thinking system) is decreased by the increase in problem complexity.

3.3 Conceptual Model and Chapter Summary

This chapter has synthesised the three main bodies of literature and identified the relationship aspects. As the produced hypotheses are the extract of these relationships, they are combined and presented to provide the summary of this chapter. Additionally, all of the explained relationships are illustrated via the following conceptual model (Figure 3.1).

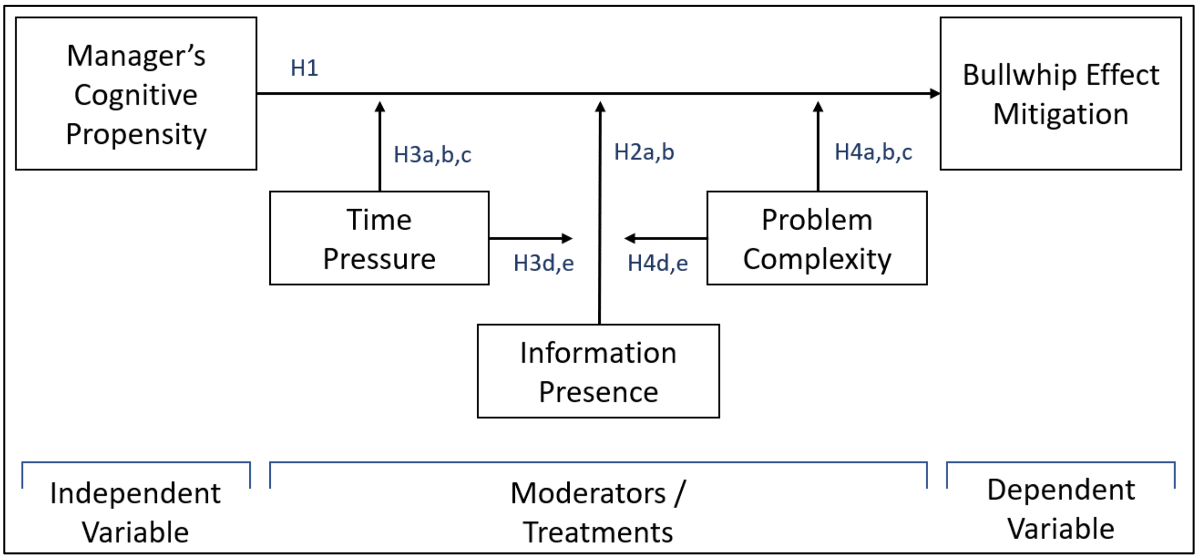


Figure 3.1 Conceptual Model

H1: Managers with a cognitive propensity towards their rational thinking system (System 2) perform better than managers who have a propensity to use their intuitive thinking system (System 1) in tracking the customer demand to mitigate the bullwhip effect.

H2a: Managers with a cognitive propensity for their rational thinking system (System 2) perform better in tracking the customer demand to mitigate the bullwhip effect when they are provided with additional information.

H2b: Being provided with additional information does not impact the demand tracking performance of managers with a cognitive propensity for their intuitive thinking system (System 1).

H3a: Managers with a cognitive propensity for their intuitive thinking system (System 1) may perform better than managers who have a propensity for using their rational thinking system (System 2) in demand tracking, when there is time pressure in the decision-making environment.

H3b: Time pressure improves the demand tracking performance of managers with a cognitive propensity for their intuitive thinking system (System 1).

H3c: Time pressure aggravates the demand tracking performance of managers with a cognitive propensity for their rational thinking system (System 2).

H3d: The effect of additional information presence on System 1 managers' demand tracking performance (managers with a cognitive propensity for their intuitive thinking system) is not impacted by the time pressure in the decision-making environment.

H3e: The effect of additional information presence on System 2 managers' demand tracking performance (managers with a cognitive propensity for their rational thinking system) is aggravated by the time pressure in the decision-making environment.

H4a: Managers with a cognitive propensity for their rational thinking system (System 2) perform better than managers who have a propensity for their intuitive thinking system (System 1) in demand tracking, when the nature of the problem is more complex.

H4b: Problem complexity aggravates the demand tracking performance of managers with a cognitive propensity for their intuitive thinking system (System 1).

H4c: Problem complexity improves the demand tracking performance of managers with a cognitive propensity for their rational thinking system (System 2).

H4d: The effect of additional information presence on System 1 managers' demand tracking performance (managers with a cognitive propensity for their intuitive thinking system) is not impacted by the increase in problem complexity.

H4e: The effect of additional information presence on System 2 managers' demand tracking performance (managers with a cognitive propensity for their rational thinking system) is decreased by the increase in problem complexity.

4. Research Design

"Education is not the answer to the question. Education is the means to the answer to all questions."

BILL ALLIN

4.1 Chapter Introduction

Scientific research consists of multiple steps. Systematically following these steps, the researcher enables the scientific validity of the work. Saunders, Lewis and Thornhill (2009) provide a systematic framework to conduct a scientific research, which is known as the research onion (*see* Figure 4.1).

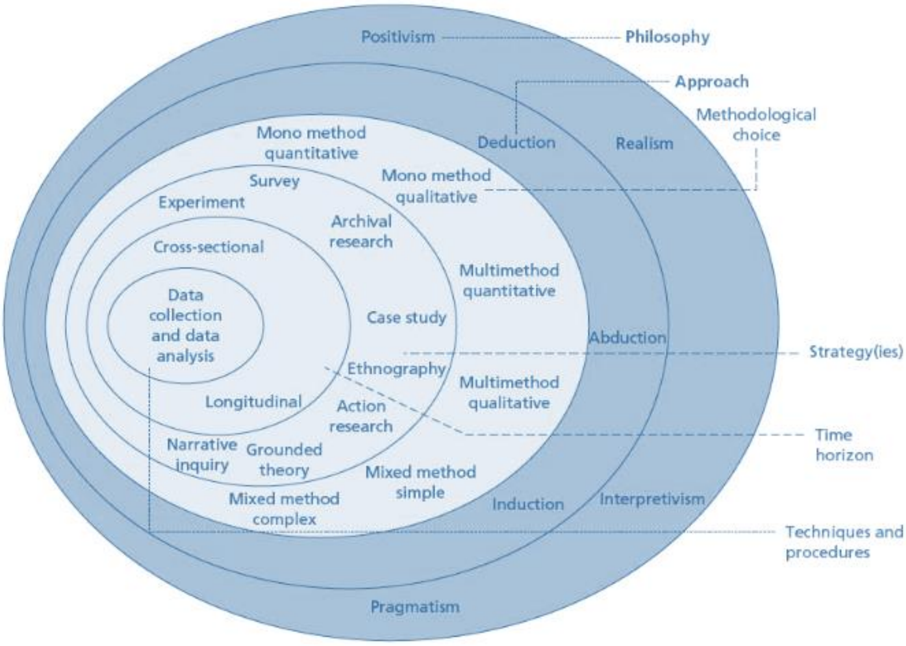


Figure 4.1 Research Onion (Saunders et al., 2009)

This research follows the layers within the research onion. This chapter begins by presenting the fundamentals of this research (*see* Section 4.2) by explaining the research philosophy of the researcher, the research approach and the selected research method.

In Section 4.3, the research method of this study – experiment – will be expressed, including the definitions of variables, their manipulation in the study and the experiment process.

Sections 4.4 and 4.5 will demonstrate the techniques that were used for the data collection and analysis respectively.

Section 4.6 will combine the insights about the quality of this work in six main dimensions: external validity via sampling and generalisability, construct validity via measurement and nuisance control, internal validity via experiment design and lastly the conclusion validity.

This will be followed by the ethics of this study in Section 4.7 and the chapter will be finalised by the summary (Section 4.8).

4.2 Research Fundamentals

4.2.1 Research Philosophy

The philosophical orientation of the researcher is the inception point of a research process; this is because the philosophy of a research is relatively constant. It does not change in accordance with the varying dynamics of the project, and it is constant in the investigator's mind. Depending on that consistent stance, the researcher defines how s/he perceives and observes a problem, and which methods can be adopted for the solution. Therefore, it directly impacts the approach to a problem, and the selection of appropriate methodology, data collection and analysis methods (Saunders et al., 2009).

The spectrum of philosophies for researchers does not have a consensus in the literature. For example, Bell and Bryman (2007) express five main philosophical perspectives: positivism, interpretivism, realism, objectivism and constructionism. Saunders et al. (2009), on the other hand, prioritise postmodernism and pragmatism to replace the last two aspects, while Eriksson and Kovalainen (2015) combine all of them and extend this philosophical classification by adding other perspectives other such as hermeneutics and poststructuralism. In contrast, the literature harbours a consensus on two main principles. Firstly, all scientific philosophies are appropriately located in the spectrum between objectivity and subjectivity (Collis and Hussey, 2013). Secondly, there is no single philosophical recipe for problems. Instead, they provide distinct solution aspects, which contingently suit different problems (Saunders et al., 2009).

To discover the philosophical positioning, a researcher should discuss the spectrum in three levels: ontological, epistemological, and axiological evaluation (Bell & Bryman, 2007; Saunders et al., 2009). Ontology refers to the search for existence and aims to find out the nature of reality. Epistemology concerns the perception and acquisition of knowledge (Saunders et al., 2009; Eriksson and Kovalainen, 2015). Lastly, axiology explains the researcher's role and impact in the project (Saunders et al., 2009).

In the light of these aspects, Saunders et al. (2009) compare the philosophical positions amongst each other. Following their work, Table 4.1 presents the comparison of two ends of the philosophical spectrum: positivism and interpretivism. This will help illuminate the characteristics of the spectrum where other philosophies are also placed.

The researcher of this study has also followed through this procedure to identify his research philosophy. In the ontological perspective, the researcher's ideas are positioned at neither of these opposite poles. The researcher believes that there must objective perception of the reality; however, it then becomes hard to identify everything in formulae. The other pole, interpretivism, also stays over-subjective for the researcher's point of view, by leaving the truth only to the interpretation of the researcher. The researcher, instead, adopts a philosophical position that mixes both poles. He believes that reality is based on objective structures and causal relationships, which are also dynamic and open to improvements by novel findings. This positioning is represented by 'critical realism' in the spectrum of philosophies.

Table 4.1 Positivism-Interpretivism comparison in three aspects (*adapted from* Saunders et al., 2009)

	Positivism	Interpretivism
Ontology	• Real, external, independent • One true reality (universalism) • Granular (things) • Ordered	• Complex, rich • Socially constructed through culture and language • Multiple meaning, interpretations, realities • Flux of processes, experiences, practices
Epistemology	• Scientific method • Observable and measurable facts • Law-like generalisations • Numbers • Casual explanation and prediction as contribution	• Theories and concepts too simplistic • Focus on narratives, stories, perceptions and interpretations • New understandings and worldviews as contribution
Axiology	• Value-free research • Researcher is detached, neutral and independent of what is researched • Researcher maintains objective stance	• Value-bound research • Researchers are part of what is researched, subjective • Researcher interpretations key to contribution • Researcher reflexive

Critical realism's ontological assertions also best suit the researcher's perspective. According to this aspect, knowledge of a phenomenon is a social construct. In other words, it connects the observable and measurable law-like objective knowledge with the subjective knowledge of narratives and interpretations. This perspective has two-fold benefits. First, the extent of the objectivity empowers the subjective knowledge forms by systematically structuring them in a social web of cumulative knowledge. Secondly, gathered subjective knowledge fills the unexplained or unobserved gaps of perceived objective knowledge.

Lastly, in the axiological perspective, the researcher believes that a researcher should take an active role to some extent. Although the investigator may be involved in a research project by preparing the research settings for example, s/he should preserve her/his objectivity to minimise researcher-led bias in the project. This also fits the specifications of critical realism.

To conclude, the researcher defines his philosophy as critical realism.

4.2.2 Research Approach

Following the structure of Saunders et al.'s (2009) research onion, the second decision to make is the selection of a research approach. The research approach can be explained as the procession of the research, considering two elements: theory and data (Bell & Bryman, 2007).

Depending on the position of theory and data, the literature provides two main research approaches: deduction and induction. The first accepts theory as the starting point of research (Eriksson and Kovalainen, 2015) and searches for its generalisability in the collected data by proposing and testing hypotheses. The ultimate goal of this approach is to either verify the theory in the tested data context or refute it. The inductive approach, however, represents the opposite direction of the process. It starts with observation of the reality and collection of the data and then aims to generalise the findings of the data into a new theory. The literature also provides an additional third approach, namely abductive reasoning, which is essentially a composition of the main two approaches. By simultaneously utilising the theory and observation, this approach provides conclusions about the phenomenon in focus (Saunders et al., 2009).

To identify the best research approach out of these three, it would be helpful to remember the objectives of this study. This research targets checking the application of an existing theory in a new context. Therefore, the researcher believes that a deductive approach is the best path to follow to conduct this study and accordingly adopts it.

4.2.3 Research Method: Questionnaire-based Online Experiment

Business and management fields provide a wide range of research methods for researchers to conduct their studies (Remenyi et al., 1998). The selection of the right research method is a critical step in research, because each method provides different advantages that come with different disadvantages. It is important to note that there is no single way to reach the goal. Easterby-Smith, Thorpe and Jackson (2012) provide a comparative framework of selected research methods that are scattered over the spectrum of epistemological philosophies. As Table 4.2 illustrates, there are no strict borders among the research methods and different methods may be effectively utilised for the very same research as long as they are appropriately justified.

Table 4.2 Epistemological comparison of selected research methodologies in management studies (Easterby-Smith et al., 2012)

	Positivist	Relativist	Constructionist
Action research	*		**
Case method	*	*	*
Experimental methods	**	*	
Grounded theory			**
Quasi-experimental research	**	*	
Survey research	*	**	*

*** Highly related, * Related*

Among these research methods that are frequently utilised in business and management studies, the case study is preferred for exploratory studies where there is a need for an elaborate study to deeply understand a phenomenon. It offers high levels of detail for the case but requires external confirmation via further studies in additional cases or larger scale studies (*see* Yin, 2009). Derived from the case study, action research and grounded studies aim to construct well-built theories respectively by either increasing the role of the researcher in the case or enlarging the data sources and hence increasing the generalisability of the case. The survey method, on the other hand, is preferred in order to discover the generalisability of previously explored findings in a case. It can explain the relationship among variables. However, it is only capable of finding out the correlational relationships among variables and cannot bring any explanation for the causal relationships (Knemeyer and Walker Naylor, 2011). Experimental methods, however, allow researchers to manipulate the variables, so that they can identify the causality among the variables in focus. In this way, they can be utilised to find out the fine details of extant hypotheses which were previously asserted by specific case studies or broad surveys (Katok, 2011a,b; Siemsen, 2011).

After comparatively analysing and understanding the advantages and disadvantages of potential methods, researchers need next to define the highest match of these trade-offs with the dynamics and targets of their study. As stated previously, this study looks for the relationship between managers' cognitive propensity and their bullwhip mitigation performance in different environmental settings. In other words, it searches for the net impacts of certain independent variables on the performance. However, the context of SCM harbours a myriad of uncertainties, which are required to be minimised. To do so and consequently to reach the net effects, i.e., causality, among the variables, it requires the isolation of all potential factors that may be correlative with the dependent variable. Therefore, the researcher has decided that experimental methods are the best fit for this research by complying with its requirements for internal purity.

Having been extensively used in fields such as economics, medicine, psychology and sociology, experimental methods have also gained interest in the operations and SCM fields, together with the increasing number of behavioural researches in these disciplines (Knemeyer and Walker Naylor, 2011). Gino and Pisano (2008) provide a wide selection of studies that utilise behavioural experiments. In those examples, the researchers investigate previously tested OM theories with a behavioural lens, so that they can detect the sub-optimality which traditional theories deviate. However, observations in those experiments may also bring unexpected results which can result in further theory development (Siemsen, 2011). Katok (2011a,b) categorises these three different utilisations of experimental methods into three main aspects. First, experiments are used to test previously produced hypotheses and refine theories. Second, they may bring additional insights to the existing theories. Lastly, they may be preferred to test the newly deployed institutional designs.

Experimental methods stand out as having two main advantages: control and responsiveness (Katok, 2011b). Due to the high level of control, researchers can manipulate the variables temporally to observe the causation they targeted. These temporal changes in the variables are called *treatments*. By deploying a specific treatment on and cancelling out others, researchers can remove all the extraneous impacts and observe the causality among the tested variables. In this aspect, researchers can also tap into the benefits of using control groups that would work as benchmark points. Besides, randomisation can also be utilised to minimise the unwanted external interventions. The second attribute of the experimental research method is its responsiveness. Conducting experiments are relatively low-cost compared to other methods, such as the case study, where the researcher may need to reach expensive databases, or the survey where it may be necessary to recruit a high number of participants. Likewise, experiments are also comparatively quicker to conduct. Therefore, if a mistake is made during any part of the experiment or if there is a need to revise any settings, experiments can be efficiently redone due to their flexibility in terms of time and cost. This operational efficiency can also be utilised alongside the development of the experiment tool by conducting quick and low-cost pilot studies (Knemeyer and Walker Naylor, 2011; Siemsen, 2011; Castañeda, 2019).

On the other hand, this method has its own disadvantages, like all other methodologies. Having high control over the settings of the research enables experiments to have a reliable internal validity. However, this high level of control, at the same time, isolates the research from any real-life context and endangers the external validity of the study (Katok, 2011a; Knemeyer and Walker Naylor, 2011; Siemsen, 2011). The literature proposes some other ways to deal with this lack of generalisability. Among them are simulation of the real-life context using scenarios or vignette-based experiments, recruiting more suitable participants that are related to the problem context (Siemsen, 2011), or complementing the research with other research methods that can provide higher external validity. However, this last option would, expectedly, result in the loss of some advantages that come with the selection of experimental methods. Therefore, in order to preserve the internal validity of the research while enhancing its external validity, some derivatives of experimental methods have been produced in the literature: laboratory, field and natural experiments (Diamond, 1983). Table 4.3 shows the main differences amongst these types.

Table 4.3 Comparison of experimental research methods

	Laboratory Experiment	Field Experiment	Natural Experiment
Randomisation	***	**	*
Investigator control	***	**	*
Manipulation of variables	***	**	*
External validity	*	**	***

**** High, ** Semi, * Low applicability*

As the table above depicts, the experiments that demonstrate all the essential characteristics are called laboratory experiments, which comes from the highly controlled environment of laboratories. The other end of the spectrum is the natural experiment, where the researcher does not have any control except the selection of the experiment site. To test the hypotheses, the researcher needs to observe the phenomena, understand, and identify the impact of causal variables in the environment. Between these two poles, there is the field experiment. It is a preferred method to increase the external validity of a study by defining the experiment area as being in the real field. However, it comes with the expense of a reduction in internal validity, because at least one of the main characteristics of this experiment cannot be exerted by the researcher. For this reason, field experiments are also named quasi-experiments. The field may possess many uncertainties and external situations that cannot be manipulated or randomised by the researcher (Katok, 2011b; Knemeyer and Walker Naylor, 2011).

Although any of these experimental methods can be used as long as it is appropriately justified and applied, laboratory experiments have become a major method in BOM (Katok, 2011a). As well as being related to the critical necessity to minimise the external impacts on human decision making, the complexity of the SC environment has also been another additional factor to abstract (Katok, 2011b). In line with this abundance of usage, laboratory experiments in the behavioural aspects are specifically named as *behavioural (laboratory) experiments* as well (Siemsen, 2011; Castañeda, 2019).

In line with the requirements of behavioural isolation in an SC setting, this study has conclusively adopted employing the laboratory experiment method among the depicted experimental method types. However, the literature has also provided some specific terms that indicate the data collection method and platform, such as questionnaire-based experiment or online experiment. Therefore, to be more specific, this study can best be described as a behavioural online experiment. The justification for the selected data collection platform will be explained in Section 4.4.2 (*p. 74*).

The last aspect to analyse within the scope of the research method is the time span and analysis approach of the study. In these aspects, and conforming to the dynamics of the research, this study adopted a cross-sectional quantitative approach. The reason for the selection of a one-off study over a longitudinal approach is related to the manipulation technique of the independent variable, which is the manager's cognitive propensity. Although it will be explained in more detail later in the chapter (*see* Subsection 4.3.5 [a-l], *p. 69*), the measurement technique of cognitive propensity requires participants to answer the questions subconsciously. When they are exposed to the same questions once more, the measurement quality of the questions can be impacted, despite there being no consensus on the topic yet. Therefore, the researcher has decided to adopt a cross-sectional study to protect its validity. Likewise, the dependent variable also has an impact on this decision. The task that measures bullwhip mitigation performance can be learned by the participants (*see* Subsection 4.3.3 [b], *p. 68*) and they can perform better in the second run. In turn, this will harm the internal validity of the research.

The reason for the quantitative analysis technique is, on the other hand, related to the objectives and constructs of the study. The study aims to identify performance differences in various situations. The best way to demonstrate these differences is to produce objectively comparable, quantitative outputs. In line with this, the constructs of the study – especially the dependent variable – require quantitative measurement.

4.3 Experiment

Bachrach and Bendoly (2011) determine five major steps for conducting an experiment:

1. Conceptualising the research question,
2. Operationalisation and design,
3. Methodology and collecting data,
4. Validity testing and interpretability, and
5. Effect and relationship testing.

In the previous chapter, the research questions were conceptualised via the help of reviewed literature. This section will demonstrate the remaining three steps of the experiment. First, the operationalisation of the variables will be explained in detail. This will be followed by the description of the application of the experiment, and finally its validity tests via pilot studies.

4.3.1 Defining the Variables

The conceptual model constructed in Section 3.3 (*see* Figure 4.2) illustrates the five main variables to be examined to answer the proposed hypotheses of this study. To sustain this clarity throughout the study, there are two essential questions to answer beforehand. First, what does each variable mean within the framework of the study? Second, how does the study manipulate or measure them?

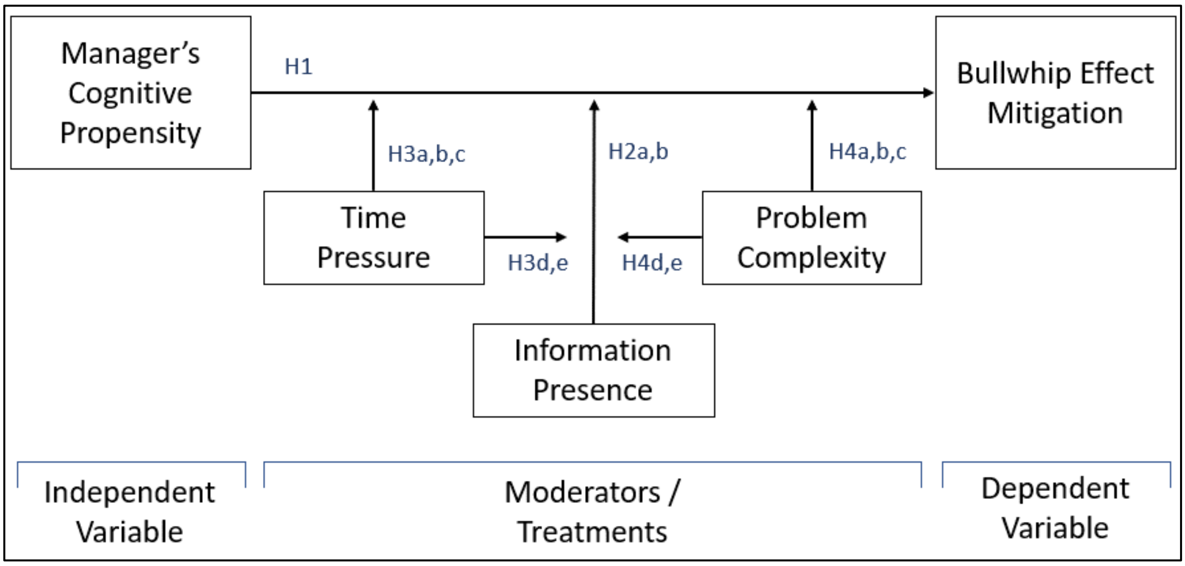


Figure 4.2 Conceptual Model (Also Figure 3.1)

Starting with the first question, the independent variable of this study is related to the *manager's cognitive propensity.* As previously explained in Section 2.3.2 (*p. 30*), individuals utilise two thinking systems when they are making their decisions: intuitive and rational. These systems work in collaboration; however, individuals may be inclined to use either system relatively more than the other. In the SC context, a manager's cognitive propensity for using either system will be employed as the independent variable in this research.

Secondly, the dependent variable is the demand tracking performance. As explained in Section 2.2.1 (*p. 17*), the BWE is a very common SC disruption and management of demand and orders have a substantial impact on its mitigation.

Finally, this research has three moderating variables. Three environmental situations were determined as the focus of this research: *information presence, time pressure* and *problem complexity*. Respectively, information presence is the extent of demand information along the SC. Time pressure is the situation where participants are asked to make their decisions in a relatively shorter time than usually required. Likewise, problem complexity is the increased complexity in the problem on which the participants are required to produce a decision.

4.3.2 Utilising Variables

To grasp the variables better, the next step is to understand how they are utilised (either manipulated or measured).

a. Independent Variable Manipulation: Measuring Cognitive Propensity by Cognitive Reflection Test

Manipulation of independent variables is an essential part of conducting an experiment. To observe the targeted impact, at least one independent variable must be manipulated, otherwise the study remains as a solely descriptive study (Katok, 2011b). In this study, the independent variable, namely cognitive propensity of an SC manager, will be manipulated *a posteriori* via a method called the *Cognitive Reflection Test* (CRT).

I. WHAT IS THE CRT?

The CRT was introduced by the seminal work of Shane Frederick (2005), and since then it has been one of the widely preferred methods to measure the cognitive differences of individuals (Stieger and Reips, 2016; Thomson and Oppenheimer, 2016; Szaszi et al., 2017; Stupple et al., 2017). It looks a short questionnaire with three simple questions (Nisula, 2016); however, its unique characteristics make it preferable despite its simplicity. The questions that compose a CRT prompt the participants to give easy, instantaneous answers at first glance. The literature (e.g., Frederick, 2005; Cueva et al., 2016) names them impulsive answers; however, these impulsive answers are generally wrong and require the individual to focus and think more analytically about the question to notice the trick (Pennycook et al., 2016; Stieger and Reips, 2016; Bialek and Pennycook, 2018). To avoid any misunderstanding, it is helpful to quote the original ideas of Frederick (2005, p.27):

> *"…the three items on the CRT are easy in the sense that their solution is easily understood when explained, yet reaching the correct answer often requires the suppression of an erroneous answer that springs "impulsively" to mind."*

The original three CRT questions that were utilised in Frederick (2005) and their correct and intentionally incorrect answers are:

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? (Incorrect: *10 cents*, Correct: *5 cents*)
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (Incorrect: *100 minutes*, Correct: *5 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? (Incorrect: *24 days*, Correct: *47 days*)

Frederick (2005) initially used the test to examine the impact of individual differences on two decision-making features: time and risk preference of the individual. Later on, the popularity of the test increased as already mentioned. The seminal work of Frederick has been cited 4,424 times to date. Besides, interest towards the test has also grown over time. In their meta-study, Brañas-Garza, Kujal and Lenkei (2015) review the 118 academic works that directly utilise the CRT and descriptively present the increase in time until 2015 (*see* Figure 4.3).

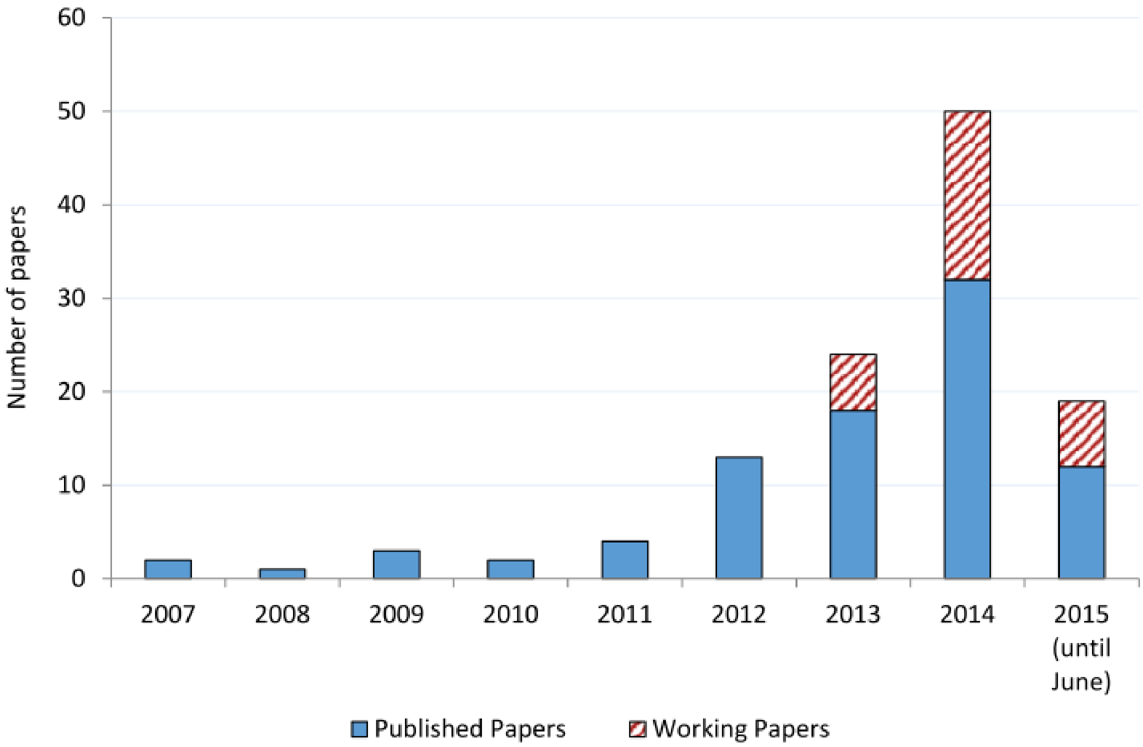


Figure 4.3 Distribution of academic works that utilised the CRT by 2015 (Brañas-Garza et al., 2015)

As in Frederick's original work, subsequent studies also used the test to investigate the relationships between individual differences and various aspects depending on the field. Among those aspects, some examples are heuristics-and-biases tasks (Del Missier, Mäntylä, and Bruine de Bruin, 2012), creativity (Barr et al., 2015), multiple decision-making (Oechssler, Roider, and Schmitz, 2009; Campitelli and Labollita, 2010; Koehler and James, 2010; Hoppe and Kusterer, 2011; Toplak et al., 2011, 2014), religious and paranormal disbelief (Gervais and Norenzayan, 2012; Pennycook et al., 2012; Shenhav, Rand, and Greene, 2012; Cheyne and Pennycook, 2013), withdrawal decisions (Kiss, Rodriguez-Lara and Rosa-Garcia., 2016), social preferences (Corgnet, Espin and Hernán-González, 2015; Peysakhovic and Rand, 2016; Ponti and Rodriguez-Lara, 2015), SAT scores (Obrecht, Chapman and Gelman, 2009) (Stieger and Reips, 2016; Thomson and Oppenheimer, 2016; Szaszi et al., 2017; Blacksmith et al., 2019; Pan et al., 2020).

The CRT was also utilised in the OM field, especially with the complementary works of Brent Moritz (e.g., Moritz et al., 2013; 2014; 2020; Narayanan and Moritz, 2015; Ovchinnikov et al., 2015). In these studies, the authors essentially investigated the role of individuals' cognitive differences in inventory decision-making performance in various settings.

II. WORKING MECHANISM OF THE CRT

To understand how the CRT indicates individual differences, two crucial questions should be answered clearly. First, how are the CRT questions processed in the mind, when an individual encounters these questions? Second, what does the CRT measure (Szaszi et al., 2017)?

Starting with the first question, the literature mainly separates the CRT answering mechanism into two major phases. In the first step of answering CRT questions, individuals are expected to notice this incorrectness and suppress it. As said in the earlier paragraphs, resulting from the unique nature of the CRT, the questions are assumed to lead the individual to an impulsive but incorrect answer (*see* Blacksmith et al., 2019). Szaszi et al. (2017) actually concretise this assumption by collating some related literature. Authors initially quote Mata, Schubert and Ferreira (2013) who affirm participants' awareness of that impulsive answer. Likewise, Travers, Rolison and Feeney (2016) also conduct a mouse-tracking experiment where they discovered that participants begin their consideration at that impulsive or intuitive incorrect answer. Therefore, individuals are initially required to suppress these answers when they are asked to solve CRT questions (Stanovich, 2016). Following this step, individuals are then expected to analytically compute the right answer (Erceg and Bubić, 2017; Patel, 2017; Szaszi et al., 2017).

III. WHAT DOES THE CRT MEASURE?

Summarising these two steps, the second question can also be instantly answered: What does the CRT measure? Mechanisms indicate that to successfully answer a CRT question, a participant needs to have a certain level of reflection capability to notice the potential impulsive answer as well as a sufficiently numerical and analytical capability to solve the problem. In other words, it can be seen as the measure of both cognitive reflection and numerical capability (Brañas-Garza et al., 2015).

However, it is critically important to mention that this measurement has limits, mainly due to the ambiguity in its usage (Blacksmith et al., 2019). Despite its growing popularity in various fields, it is relatively nascent field. So, it is still in evolution and there are multiple challenging opinions regarding its properties and utilisation (Erceg and Bubić, 2017; Szaszi et al., 2017). The main confusion has been related to the interpretation of CRT scores. While Frederick (2005) built the test to measure the capability of overriding an intuitive but incorrect answer, subsequent authors' interpretations have varied. For example, Gino and Ariely (2017) accepted it as a measure of intelligence, Cokely and Kelley (2009) and Liberali et al. (2012) as cognitive impulsivity, Weller et al. (2013) as numeracy, and lastly Pennycook et al. (2012) as cognitive style (Blacksmith et al., 2019).

Therefore, for an appropriate and valid manipulation of the CRT, it is important to understand what it will mean for a study. In line with this necessity, the author has identified two main interpretation clusters in the literature for the interpretation of CRT results: dual-process perspectives and cognitive ability.

Initially when Frederick (2005) released the test, his definition of reflection was based on the individual's "ability or disposition to resist reporting the response that first comes to mind" (p. 35) (Campitelli and Labollita, 2010). This was also defined by others as "the tendency to override the impulsive answer" (e.g.*,* Thomson and Oppenheimer, 2016; Moritz et al., 2020; Pan et al., 2020). According to this perspective, the fewer number of impulsive answers given makes the individual more reflective. However, there have been authors that went beyond the term reflection. For instance, Toplak et al. (2011, 2014) considered the CRT as a predictor of rational thinking, whereas Campitelli and Labollita (2010) saw it as a disposition towards actively open-minded thinking (Pennycook et al., 2016). Agreeing on these, Stupple et al. (2017) asserted that CRT indexes analytical thinking which is a more generic concept than rationality. In line with this, Bialek and Pennycook (2018, p.1953) identified the CRT as a tool to measure the 'propensity to engage in analytic or deliberative reasoning in lieu of gut feelings or intuitions'.

Among the studies that aim to determine what the CRT measures, there has been another spin-off from the reflection in an opposite direction. While the reflection aspect asserts that the number of non-intuitive answers indicate the power of the individual to resist the impulsive answer, this school of thought is based on the idea that the CRT shows the cognitive miserliness of an individual via the number of intuitive responses (e.g*.,* Toplak et al., 2014; Thomson and Oppenheimer, 2016; Szaszi et al., 2017). In Subsection 2.3.2 (b) (*p. 30*), the laziness of the human mind was already explained by highlighting that our mind is more prone to spend less effort when making decisions. In line with this, cognitive miserliness is defined as the 'unwillingness to go beyond the default' (Stupple et al., 2017, p.3). Building an analogy to the relationships between reflection and the propensity for analytical thinking, some ideas have also emerged from the cognitive miserliness idea and regard the CRT as a measure of the propensity for intuition (Alós-Ferrer et al., 2016; Pennycook et al., 2016). Combining all these, a comprehensive perspective has also emerged which considers the CRT as the predictor of an individual's propensity to use either thinking system in decision-making (e.g*.,* Stieger and Reips, 2016; Moritz et al., 2020; Šrol, 2018; Damnjanović et al., 2019).

Secondly, some researchers stressed the CRT's capability to measure individuals' cognitive abilities. This idea, at its core, believes that even though participants show the required reflectivity to notice the impulsive answer, they must also show some level of cognitive ability to solve the questions (Szaszi et al., 2017). However, this certainly does not mean that the CRT is a pure measure of cognitive ability but is 'aided by reading comprehension and mathematical skills' as Frederick (2005, p.35) stated (Blacksmith et al., 2019, p.601). In line with this, many studies have presented a correlational relationship between participants' CRT scores and their numeracy skills (e.g., Alós-Ferrer et al., 2016; Thomson and Oppenheimer, 2016; Stupple et al., 2017; Szaszi et al., 2017; Bialek and Pennycook, 2018).

In conclusion, there is not a consensus on the measurement scope of CRT yet (Stupple et al., 2017) and it is still dynamic and evolving by the contributions of accumulating studies. Considering that, this study has adopted the major and the most comprehensive idea that considers CRT as the measurement tool of an individual's propensity to preferably using either thinking system when they make a decision.

IV. COMPARISON OF CRT TO OTHER ALTERNATIVE METHODS

As there are various methods to solve a research problem, there are also multiple means to identify the thinking disposition of an individual besides the CRT. They can be primarily categorised into three: self-reported such as Need for Cognition (NFC), Faith in Intuition (FI) and Rational Experiential Inventory (REI) (e.g*.,* Petty et al., 2009; Alós-Ferrer et al., 2016; Pennycook et al., 2016; Nisula, 2016; Stupple et al., 2017; Szaszi et al., 2017); questionnaire-based method such as heuristics and biases tasks (Šrol, 2018); and verbal methods (Gavirneni and Isen, 2010).

It is important to keep in mind that there is no single right solution; instead, every method has up and downsides. CRT, for example, was criticised in some respects mainly based on its ambiguity. Some authors (e.g*.,* Erceg and Bubić, 2017; Blacksmith et al., 2019; Erceg, Galić and Ružojčić, 2020) recently questioned the validity of the CRT in two respects: its prediction capability and insufficient conceptualisation of the topic. The former criticism can be connected to the aforementioned ambiguity. As an evolving tool, studies with challenging assertions will bring clarity to the topic as they accumulate. Secondly, CRT is criticised for insufficient internal validity. It is a short and unique test, but it has not been built on the systematic construct validity approach. Instead, authors have measured its power by its correlation with other, similar tools. Similar concerns have been stated for the other methods too. For instance, self-report methods may lack true reflectivity as they are consciously expressed by the decision-maker (Šrol, 2018). Likewise, interpretation of verbal methods is limited to the conscious expressions of the participant and unconscious thinking processes can easily be missed (Szaszi et al., 2017).

There are advantages to, the CRT embodies as well. Starting from the seminal work of Frederick (2005), many studies have proved the CRT's prediction power by presenting their correlation with the various potential tools mentioned above. In Frederick's (2005) study, it came up as either the best or second-best test among four decision-making aspects and the only one that can measure them all. Being only a three-item test to be completed in a few minutes and producing better results than tests with up to 215 questions that take up to around 3.5 hours, CRT has become attractive and popular among researchers.

On balance, the CRT has been selected for this study in lieu of other methods as the advantages outweigh the disadvantages in this context. Some attributes of this study, which will be explained in the following sections, have also contributed to this decision. For example, the order management task that was utilised to measure the dependent variable required a certain level of cognitive effort (*see* Subsections 4.3.2 [b-III, V], *p. 62*). Therefore, utilising a heavier and longer method could have impacted on participants' NP performance. Likewise, the experiment was conducted online to reach the participant number that is required to measure all the treatments of this experiment (*see* Section 4.4.2, *p. 74*). As a side effect, however, this led to diminution of experimenter control (*see* Subsection 4.3.5 [b], *p. 72*). Consequently, a method which needs the least amount of experimenter supervision or guidance was needed. Enabling both requirements, while providing the equivalent prediction power at the same time, CRT was conclusively opted for in order to identify the individual thinking style propensity.

V. CREATION OF THE CUSTOMISED CRT

Alongside the concerns regarding the construct validity of the CRT, questions themselves were also questioned for various reasons. The main consideration that emerged was the increasing popularity of the test. Though it is a positive indicator, the simplicity of the test has caused a vast usage ratio among studies from different domains and hence a continuously increasing number of people have come across the three original CRT questions (Welsh and Begg, 2017; Blacksmith et al., 2019). Moreover, in the environments of stationary participants, such as online or laboratory subject pools, participants' familiarity with the questions has become a more serious issue (Thomson and Oppenheimer, 2016).

Originating from this concern, researchers have started to evolve the test (Stieger and Reips, 2016). Two main approaches were followed for this dynamic evolution: tweaking the original questions and expanding the test by developing new questions (Toplak et al., 2014; Damnjanović et al., 2019). Independently from the evolution approach, the literature presents three main criteria that researchers have considered when customising a CRT: the difficulty level of questions, measurement of numeracy and number of questions. Firstly, overcomplexity or longevity would make it more difficult to notice the trick in the question (*i.e., flooring effect*), while at the other end of the spectrum, oversimplicity would result in a higher ratio of correct answers than the original questions (*i.e., ceiling performance*) (Thomson and Oppenheimer, 2016; Patel, 2017; Welsh and Begg, 2017). Secondly, measurement of numerical abilities has been another matter for consideration when revising a CRT. As well as researchers deliberately including questions with a strong capability to measure numeracy, some also preferred having non-numerical questions to focus only on reflectivity (Szaszi et al., 2017). Lastly, the actual number of questions has also been a point of interest. Many researches considered three questions to be insufficient for the purposes of a CRT, and added additional questions (Stieger and Reips, 2016; Erceg and Bubić, 2017). However, since too many questions would be both cognitively tiring and distracting, an upper limit has also been adopted and many researchers generally prefer to have five to seven questions.

In line with these considerations, various customisations have been made to the original CRT. After a comprehensive but non-exhaustive review of those, 44 questions have been identified in the literature. Utilising the aforementioned three categories with pilot studies, the 44 questions were evaluated, and the following five questions were selected for this study:

1. Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class? (from Toplak et al. (2014); correct answer is 29, intuitive incorrect answer is 30)

2. 25 soldiers are standing in a row 3m from each other. How long is the row? (from Oldrati et al. (2016); correct answer 72m; intuitive incorrect answer 75m)

3. You go to bed at eight. You set your old analogue alarm clock to wake you up at nine. How many hours of sleep will you get? (from Oldrati et al. (2016); correct answer 1h; intuitive incorrect answer 13h)

4. A frog fell into a hole 30m deep. Every day it climbs up 3m, but during the night, it slides 2m back down. How many days will it take the frog to climb out of the hole? (from Ackerman (2014); correct answer 28 days; intuitive incorrect answer 30 days)

5. A man buys a pig for £60, sells it for £70, buys it back for £80, and sells it finally for £90. How much has he made? (from Toplak (2014); correct answer is £20, intuitive incorrect answer is £10)

For all these questions and their evaluation, please see Appendix A (*p. 127*); for a detailed explanation of the pilot processes that lead to these five questions, please see Appendix B (*p. 130*).

VI. MANIPULATION

CRT has been elaborately explained in its various aspects so far. To conclude the section, the manipulation rule applied in this study will be explained.

CRT questions are designed to be answered in three ways: *correct*, *intuitive incorrect* and *non-intuitive incorrect* (Cueva et al., 2016; Pennycook et al., 2016; Thomson and Oppenheimer, 2016; Damnjanović et al., 2019). In line with this, participants who had provided three or more *intuitive incorrect* answers out of the five CRT questions were *a posteriori* categorised as participants with a propensity for intuition. Conversely, participants with fewer than three intuitive incorrect answers were labelled as participants with a propensity for rationality.

It should also be noted that this research has not studied the difference between the correct and non-intuitive incorrect groups. Because the main difference between these two groups is their numeracy capabilities. They have a similar level of cognitive reflection, they noticed the impulsive answers and started to ponder more on the question. Therefore, they were categorised in the same cognitive group while discarding their numerical differences in this instance as it was not the focal point of the study.

b. Dependent Variable Measurement: Order Placement Performance via a Supply Chain-wide Newsvendor Game

I. BULLWHIP EFFECT AND ORDER PLACEMENT PERFORMANCE

This study has been motivated by the research gaps related to the BWE and its relationship to the cognitive characteristics of SC managers. Occurrence of bullwhip and its causal factors have been provided in Section 2.2.1 (*p. 17)*. Following that, Subsection 2.2.2 (*p. 23*) has explained another crucial construct related to bullwhip: order management. Its importance in the bullwhip domain comes from its dual role. Firstly, order quantity is impacted by the bullwhip throughout the SC alongside the inventory quantities and total inventory costs. Secondly, placed orders are the main conveyor of fluctuating information among the SC partners. In this way, they result in the occurrence of bullwhip at the same time. However, this also means that effective order management strategies can also be utilised to mitigate the BWE in the SC. In line with this idea, many researchers have measured managerial order and inventory management performance as the proxy to measure the occurrence and evolution of the BWE in the SC. To measure the order placement performance, however, individual performance was generally benchmarked to an optimal performance (*see* Croson et al., 2014; Moritz et al., 2020). Alternatively, the performance of two groups were comparatively analysed to identify the least cost performance.

II. TRIAL OF BEER DISTRIBUTION GAME

To conduct these analyses, experimental methods have employed various game-based approaches (e.g*.,* Gina and Pisano, 2008; Katok, 2011a; Chen, Su and Zhao, 2012). Among those, the *Beer Distribution Game (BDG)*, in short *The Beer Game*, was heavily used to understand the occurrence of and reasons for the BWE (Chen and Disney, 2007; Narayanan and Moritz, 2015). The game was originally devised as a board game by the Sloan School of Management in the early 1960s for system dynamics research, but then gained popularity in the SC and BWE fields after the work of Sterman (1989a,b) (Strozzi et al., 2007; Edali and Yasarcan, 2014). The game is played by four individuals who represent the sequent firms of a serial SC, namely factory, distributor, wholesaler, and retailer (Moritz et al., 2020). Participants observe their inventory level and place orders to the upstream SC partners. Throughout the game, they aim to meet the total customer demand with a minimum inventory holding cost (Coppini et al., 2010; Moritz et al., 2020). To fulfil these aims, participants consider certain aspects when they place orders (*see* Section 2.2.2, *p. 23*). These aspects cover the openly known quantities of inventory in-stock, supply line inventory and incoming demand, as well as the unknown parameters such as managers' cognitive perceptions of the desired inventory level, perception of supply line inventory and demand expectations. Managerial differences regarding these cognitive approaches determine their order management policies, and therefore their bullwhip mitigation performances.

This study was initially designed to operate a customised BDG to observe the order placement performance of participants. However, as the study developed, some concerns were raised for the utilisation of BDG in this study. Firstly, as already stated, the game was designed to be a board game in the first instance. Since it is a complex game with detailed procedures, it requires a proper experimenter control at the beginning of the game as well as the total game duration. As a result, participants have to fully understand the game and act as intended. However, as explained in Subsection 4.3.5 [a-II] (*p. 70*), this study utilises 5 variables (1 dependent, 1 independent and 3 moderators) and 12 experimental treatments. These consequently increased the required participant number and led to challenges with recruitment of the target population (*see* Section 4.4.2, *p. 74*) in a laboratory setting. Since utilising the board game version became harder, a computerised online version of the BDG was developed by the researcher as a solution, following the previous applications in the literature (*see* Oliva and Gonçalves, 2005; Moritz et al., 2020). Although computerised version of BDG has brought a solution for participant recruitment, it had some drawbacks as well, of which the biggest caused by the complexity of this study's treatments. This complexity required further control and supervision of the experimenter to ensure the participants' comprehension of the game, edgeways its validity. The valid applicability of that version was tested via pilot studies (*see* Appendix B-1, *p. 131*), which ultimately demonstrated an imbalance in the comprehension of participants among the treatments. Even though that condition was met, there was another challenge with the BDG selection. As explained in the previous paragraph, participants in the BDG are required to consider multiple aspects in their ordering decision, such as the in-stock inventory, supply line inventory and the demand expectation for the next period. However, BDG allows participants to make only one decision, which is the decision of order quantity for the upcoming period. In other words, standard BDG applications are based on inferring the aforementioned parameters using the only available data, which is the order quantity of the participants. To do so, the ordering formula (*see* Section 2.2.2, *p. 23*) and analysis methods such as Partial Least Squares* and Genetic Algorithm* are utilised to predict the most meaningful set of parameters. However, this method of analysis requires a much higher number of participants to put forward a valid analysis, which was difficult to achieve especially when there are multiple unknown parameters and few known values.

To provide a valid measurement in the online platform, it was decided to simplify the game and to do so, another common inventory management game, namely the *Newsvendor Problem* (NP), was utilised.

III. NEWSVENDOR PROBLEM

The Newsvendor Problem differs from the BDG in four main aspects. First, the former requires one agent in comparison to four roles in the latter. Secondly, it is a single period game while the standard BDG lasts approximately 30 periods. Third, it is based on the zero-stock inventory; in other words, the excess inventory is not stocked for the next period but is salvaged in return of a deposit. Lastly, participants are provided with only two sets of data in the NP: demand and finance information (Gavirneni and Isen, 2010). Therefore, participants are expected to consider the stochastic demand profile, material cost, selling and salvage prices, and place an order amount that will result in the highest profit for that period (Castañeda, 2019). To sum, the NP is a simplified version of the BDG in multiple aspects (Katok, 2011b).

In line with this comparison, the simplifications of the NP can bring solutions to the two main drawbacks of the BDG. Firstly, participants are required to analyse only the demand information in comparison to the multiple aspects of the BDG. This, in return, facilitates the recruitment of a sufficient number of participants for a valid analysis. Secondly, simplicity directly impacts on the easier comprehension of the game, even in the online platform where an effective experimenter control is lacking (*see* Appendix B, *p. 130*).

However, the NP also has some deficiencies in the research of the BWE per se. The original game is conducted as a single period game via a single agent and this does not suitably correspond to the SC research which is based on the connections of multiple stakeholders. Additionally, this setting does not allow the application of the moderator treatments – higher information access, time pressure and

* Partial Least Squares method, in essence, assign some predicted figures for each parameter in a formula and among all the alternative predictions, it aims to find the optimal set of parameters that will equal to the closest possible figure to the known value of the formula. This proximity of that predicted figure is calculated by the squared deviations from the known value and the parameter set with the least-squared-error is chosen as the solution. The problem with the Partial Least Squares is that it requires the experimenter to set an initial value for each parameter and for each initial parameter configuration, different least-squared results can be found. Therefore, it is essential to mention that this method is more suitable to predict the fine details where the approximate values of the parameters are predicted. To fill this gap of the Partial Least Squares and Genetic Algorithm methods can be utilised. As contained in its name, this method is based on changing the gene of the formula for each run. In other words, it sets different initial values for each parameter, compares their proximity using the squared deviations and determines all the potential closest configurations for the decision-maker to choose.

problem complexity. Therefore, combining and customising the BDG and NP, another game was developed by the researcher to measure the dependent variable of this study with validity.

IV. CUSTOMISED ORDER MANAGEMENT GAME

While the customised game is built on the basics of the NP in terms of the zero-inventory policy, the business range of the game has been selected as the 'supply chain', being adopted from the BDG. Therefore, it can also be called the 'supply chain-wide Newsvendor Problem'. Table 4.4 presents the key characteristics of the customised game together with the respective source game and the reasons.

For further details regarding the evolution of the tool, please see Appendix B (*p. 130*) where the pilot studies are explained.

Table 4.4 Customisation summary of the OMG

Attribute	Adopted application	Adoption from	Explanation
Business scope	Supply chain	BDG	• To provide the feeling of the SC • To operate the moderator treatments
Demand profile	Seasonal	BDG	• To measure the demand tracking capability of the participants
Period amount	Multiple (15 periods)	BDG	• To provide enough data spectrum • To bring the feeling of order and inventory flow
Period-end inventory	Zero-inventory	NP	• To simplify the game for the online platform for easier comprehension and higher validity • To decrease the construct to analyse, hence, to keep the participant requirement at an applicable level
Lead time and supply line policy	Next day, full delivery	NP	• To simplify the game for the online platform for easier comprehension and higher validity • To decrease the construct to analyse, hence, to keep the participant requirement at an applicable level
Item cost and sale price	High profit margin setting	NP	• In the high profit margin settings where material cost is much lower than the selling price, participants' biases are observed better than the low profit margin settings
Period-end feedback	Immediate	NP	• To increase the comprehension of the game by helping participants to analyse their previous decisions and demand pattern • To bring the feeling of order and inventory flow • Casteñada (2019) states that the feedback interventions do not invalidate the test
Agent number	One echelon (Distributor)	NP	• To eliminate the impact of other participants' performance by providing all participants with the same conditions • That one echelon was decided to be the 'distributor', because the feeling of bullwhip is felt towards the upstream SC. Besides, selecting an upstream echelon is more suitable to utilise the information presence treatment (Coppini et al., 2010; Moritz et al., 2020).

Considering the BWE is a SC phenomenon, this game's one-echelon formation may require further clarification regarding its applicability. The BWE is occurred by unexpected changes in the customer demand, even with a tiny change, and finally, its consequences are observed both spatially and temporally. Spatially, the fluctuation of orders and inventory increase towards the upstream SC. Downstream echelons like retailer or wholesaler are impacted less than upstream echelons like distributor and factory. In the temporal perspective, each echelon experiences the BWE in time unless they have a systematic operational strategy to solve it. In sum, the occurrence of the BWE can be observed for the supply chain as well as for each echelon of the SC, especially the upstream echelons as they are the major stakeholders experiencing the phenomenon. In line with this, this study utilises the OMG to observe the reaction against an emerging BWE in the distributor echelon in this 15-period time. In other words, this game is not designed to observe the generation or visualisation of the bullwhip, but instead to demonstrate the demand tracking capability of the potential supply chain managers, which is essential to manage and mitigate the bullwhip effect.

Combining all the aspects presented in Table 4.4, the customised Order Management Game can be summarised as follows. The game is played by one participant for 15 periods, who has the role of distributor. Duties of remaining SC partners are simulated in the experiment to eliminate the undesired interaction between the performances of participants playing in the same SC. Participants are provided with two types of information, i.e., constant and variable. First, the constant information set was informing participants about the inventory policy (zero-inventory), financial policy (cost, sale and salvage prices) and lead time (next day delivery). These were shown to the participants as visualised throughout the game, while they were also explained in the instructions before the start of the game. Since these are stationary, they were primarily aimed to help participants better understand the basics and flow of the game. Secondly, the variable information provided to the participants was related to the customer demand. It is also presented in the instructions where participants were told how they can use that information. As depicted there, participants were asked to analyse customer demand history, understand its trends and then accordingly place their order to the factory to receive the ordered demand the next day and fulfil the forecasted incoming customer demand. After each decision, they are provided with feedback about their forecasting performance and expected to perform better in the proceeding periods by utilising this feedback.

V. MEASUREMENT OF THE PERFORMANCE

Utilising these customised settings, participants were presented with periodic demand information for 15 periods and were asked to place orders for the upcoming periods (*see* Figure 4.4). They were expected to analyse the trends in the demand history in the downstream SC, forecast the upcoming demand and adjust their orders accordingly.

The game essentially measures the demand tracking capability and forecast accuracy of the SC manager of the distributer echelon in a SC. This measurement is accepted as one of the proxies of BWE occurrence and mitigation performance. Because demand management is a crucial component of BWE analysis together with the management of inventory and incoming supplies. While earlier studies analysed all proxies at once by primarily utilising BDG, OMG focuses on a single proxy to isolate the effects of the others. Subsequent studies can utilise the remaining proxies and they can be combined in a final study that will enable identifying their single and combined impacts on the BWE occurrence and mitigation. Utilisation of a single proxy is especially crucial for this study with regards to its methodology, namely experiment. Experimental studies are based on the maximum isolation of the constructs, so that the causal impacts can be observed without the minimum infiltration of external and undesired factors.

Consequently, the variance between participants' demand forecasts (i.e., placed orders) and actual demand received determined their performance. Therefore, higher variance has meant a poorer demand tracking capability, while the lower variance has indicated better performance.

The application of the tool in the experiment will be explained in Subsection 4.3.5 [c] (*p. 73*).

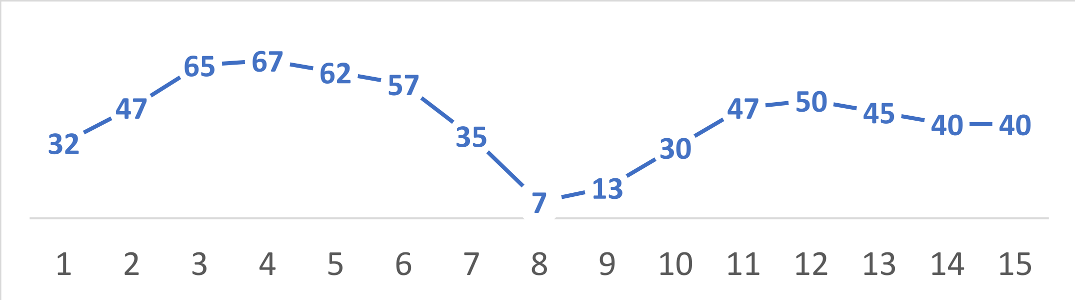


Figure 4.4 Seasonal demand flow for the 15 periods

c. Moderator Manipulation: Treatments

Although the backbone of this study is based on the impact of the independent variable on the dependent one, it is also supported by the moderating factors. Moderators were configured in two treatments. Constructing at least a set of treatments is essential for the experimental methods (Katok, 2011a), because an observation must be compared with a control or comparison group to identify the direction and power of the impact. In line with this, the construction of treatments within the order management task of the experiment will be explained for each moderator in this section.

It is also critical to remind that the aforementioned game flow and measurement technique were designed to be applied the same across all conditional treatments. Even though participants may be assigned to the treatments with different amounts of information, time limitations or problem complexity, they all had the same task to analyse the downstream SC demand with all provided information and place their orders for the upcoming period.

I. MANIPULATION OF INFORMATION PRESENCE

The first moderator is about the amount of information provided to the participants. Previous experiment applications in the OM field have manipulated various information attributes according to their requirements. Katok (2011b) identifies three ways of manipulating the amount of information in a NP: (1) manipulating the amount of provided order options, (2) manipulating the provided feedback information regarding the previous periods, and (3) manipulating constraints regarding participants' ordering styles. Likewise, BDG (*see* Steckel et al., 2004) can easily manipulate all the information provided such as visibility of supply-line inventory in the upstream SC, change of lead times or visibility of the customer demand in the downstream SC partners. As explained in Subsection 4.3.2 [b-IV] (*p. 63*), this OMG provides constant information for the inventory and lead times. Besides, it also provides variable information of customer demand and participants are asked to utilise it to make their order placement decisions. In line with this, manipulation of this first moderating variable -information presence- is also designed to be applied by changing the accessed tiers in the SC. One group of participants was provided with the demand information of only the direct customer while the other group was additionally given the demand of the retailer and end customer. Demands of those SC partners were designed to follow the same demand trend with one period lag for each partner and to give earlier signals regarding the potential changes in the direct customer demand. For instance, by looking at the demand of end customer, participants could have had a rough idea about the incoming order that will come after two periods. This additional information was aimed to bring further help for the participants to understand the demand trend in advance.

II. MANIPULATION OF TIME PRESSURE

As a straightforward construct, time pressure is generally manipulated by changing the amount of time given to the participant for completion of a task (Hwang, 1994). Following Evans et al. (2009), this study has manipulated the time into two groups. First is the baseline group where participants do not feel any additional pressure of time, but they are expected to complete the task in the time they are given. In the second group however, they are forced to complete the task in a much shorter time than required. The time counts down on the game screen and participants are required to catch up with time so as not to lose any performance points.

The amount of time given has been determined as a result of the second pilot study (*see* Appendix B-2, *p. 142*). It has been found that participants complete a period of the task in 20 seconds. That amount of time was determined as a suitable time to allow participants to complete the period, but at the same time, would give them a feeling of time pressure. However, pilot studies have also shown that the time spent on a period decreases throughout the game. That meant a learning curve due to the simplicity of the game. Therefore, to keep the valid impact of time pressure in all periods, the time allocated to the participants was decreased to 15 seconds per period.

Lastly, time pressure treatment has also been manipulated in the presence of higher information. To allow participants to process that extra information, in the treatment where participants are under time pressure with higher information, they were provided with five additional seconds per period.

III. MANIPULATION OF PROBLEM COMPLEXITY

SCs are complex systems in themselves, with various stakeholders, actions and constraints. Order management is also a complex combination of different decision processes (Moritz et al., 2020). On the other hand, all these realities were simplified in this experiment to limit the frame of the study to the target variables and to ensure the participants' successful comprehension and completion of the tasks. Therefore, it needed a careful preparation phase for the manipulation of problem complexity variables. Two critical points were considered:

- To ensure an effective manipulation, participants of one treatment must be exposed to a task that is sufficiently more complex than the baseline treatment.
- At the same time, that complexity should not prevent participants' comprehension. In other words, the task must preserve its minimum simplicity.

The literature on SCM and the BDG provides some applications. Van Riel et al. (2003), for example, relate the problem complexity to the number of cues for a task. Likewise, Croson et al. (2014) assert that the difficulty of a task may vary according for various reasons, some of which are feedback processes and time delays. Although these may be perfectly appropriate for other studies, utilising these means in this study may be unintentionally mixed with other moderating variables. Since that would harm the valid measurement of the study, this potential harm has been regarded as a constraint when manipulating the problem complexity variables. Instead, another aspect related to the structure of the SC has been chosen for this manipulation. Participants of the baseline treatment were provided with a single channel SC, while the other group went through a double channel SC. When preparing the complex treatment, the amount of the customer demands was divided into two with ratios of 1/3 and 2/3. Therefore, while both treatments presented the same amount of demand for each period, in the complex treatment, it required participants to analyse the demand in two different sales channels. It is also critical to mention that since they were not told that both marketing channels have similar demand trends, they should have considered their forecasts separately and then place the total as their order to the upstream SC partner, i.e., factory.

Although the level of complexity seems low, it has been regarded as suitable due to two main reasons. First, since the OMG was intentionally designed to be simple enough to manipulate the performance of participants in the online settings without any additional requirement of experimenter guidance, the level of complexity moderation was arranged in line with the simplicity of the overall game. In other words, the aim was to increase the complexity of the game in comparison to the other treatments. Secondly, the literature provides examples of behavioural experiments employing various treatments to measure the interaction between problem complexity and cognitive load. For example, Wang et al. (2014) provide a wide range of selections measuring the cognitive load in complex problems. Authors' citation from Campbell (1988) also shows task complexity's suitability to be measured by increasing the information load, diversity or rate of information change. Therefore, even though the change may seem simple, it may be complex enough to observe the performance variations concerning cognitive differences.

4.3.3 Construct Validity I - Control of Nuisance Variables

Target variables and their utilisation in the boundaries of this study have been explained so far, however, there is another set of variables that is of high importance, especially for the experimental methods: nuisance (or control) variables. As the experiments, including this experiment, are specifically used for identifying the causal relationships, unwanted effects of potential but unavoidable variables must be controlled to ensure that they do not violate the validity of the findings. In this study, nuisance variables related to both tasks of the study (CRT and OMG) were detected and controlled.

a. CRT-related Nuisance Variables

Regarding the CRT phase, two aspects were controlled: gender related differences and experience with the test, i.e., previous exposure. Starting from Frederick (2005), previous literature (e.g., Cueva et al., 2016; Szaszi et al., 2017; Brañas-Garza et al., 2015) demonstrated that male participants scored higher than females in CRTs. Therefore, it has been decided to control for this factor so as not to skew or distort the results.

Secondly, previous exposure has been another point of interest in the CRT research. Many authors (e.g., Haigh, 2016; Patel, 2017) asserted that being experienced with the CRT questions violates the validity of a CRT, basing their assertion on three reasons:

- The three original CRT questions are intentionally devised to be short, seemingly easy at first sight but aiming to deceive the participant. In other words, once a person notices this hidden trick, then the question would lose its operability for that person (Frederick, 2005).
- The test has been increasingly popular since its first usage among the psychology and behavioural studies, which increased the chances of a person seeing the test (Toplak et al., 2014). Besides, its popularity was also nurtured by popular science books such as 'Thinking, Fast and Slow' by Kahneman (2011) (Welsh and Begg, 2017).
- The popularity of the test has also been accelerated by studies using online recruitment tools. Since these platforms are composed of a stationary subject pool, the ratio of potential participants who had already been exposed to the CRT has dramatically increased (Thomson and Oppenheimer, 2016).

However, there are also researches that assert the opposing argument (e.g., Brañas-Garza et al., 2015; Bialek and Pennycook, 2018; Šrol, 2018). According to them, overexposure to CRT questions does not harm the test's validity because people who have noticed the trick have already answered the questions correctly and they are expected to give correct answers again. People with lower cognitive reflection, however, will not have already noticed the trick in the first instance. Therefore, the questions are still novel to them and preserve their validity.

This study, however, is based on the first school of thought, especially considering participants from the subject pools that are recruited multiple times. Greater exposure may increase the possibility of noticing the tricks in the questions. Based on this concern, this study has replaced the original CRT questions with relatively novel ones, as suggested by Stieger and Reips (2016). It has also embodied a set of control questions for any previous exposure to the utilised CRT questions (Haigh, 2016; Welsh and Begg, 2017).

b. OMG-related Nuisance Variables

In the second task of the study, there has been only one concern regarding the interference of a nuisance variable: the learning curve. The literature shows some example studies that have shared the same consideration (e.g., Narayanan and Moritz, 2015; Pan et al., 2020). The potential impact of the learning effect in this study has been considered, based on the following three reasons:

1. Simplification for online execution,
2. Repetitive structure of the game for 15 periods,
3. Differences between the working mechanisms of intuitive and rational thinking systems (*see* Subsection 2.3.2 [b], *p. 30*).

The third reason is particularly important. As intuitive thinking system is based on utilising experience or emotion-based internal knowledge, individuals with a cognitive propensity for intuition may learn the game faster. Since this may also change the performance of that group, it was decided to control for any potential occurrence of learning curve interference.

4.3.4 Construct Validity II – Control of Operability

To ensure the validity of the study, another aspect to control is the operability validity of the constructs. In other words, utilised methods, techniques, or tools must measure or manipulate the associated constructs as targeted. This section will present the control of operability for each construct respectively.

a. Manipulation Check of CRT

This study has adopted five CRT questions which are different from the original three. To ensure that they hold the same validity as the original questions, they were evaluated throughout the pilot studies from two main aspects.

Firstly, CRT questions are expected to have a certain level of difficulty as well as impulsivity. Previous researchers (e.g*.,* Narayanan and Moritz, 2015; Stupple et al., 2017; Bialek and Pennycook, 2018) measured this by descriptively analysing the ratio of correct and incorrect results. In Frederick (2005), approximately 30% of the participants answered all the questions correctly. Szaszi et al., (2017), later, studied the success ratio of the original three questions where the ratio of correct answers was between 21-41% and that of the intuitive incorrect answers was between 47-71%. To provide the optimum difficulty, this study has aimed to have nearly half the participants (50%) provided intuitive incorrect answers. More information about how the questions have been analysed can be found in Subsection 4.3.2 (a-V) (*p. 59*) and Appendix B (*p. 130*).

Secondly, the response time is utilised to control the manipulation quality of the CRT questions used. Many studies (e.g*.,* Alós-Ferrer et al., 2016; Szaszi et al., 2017; Blacksmith et al., 2019) have demonstrated the difference between the intuitive incorrect and other answers (correct and incorrect). As the intuitive answer is the one that comes to mind first, they are answered more quickly than the other answers which require further analytical consideration. Patel (2017) mentions a critical aspect regarding the issue which happens when a participant becomes confused and spends a longer time than the average. Nevertheless, this study believes that response time can be an indicator of the operability of the questions in the majority of cases.

b. Measurability Check of OMG

The most important aspects to control in the OMG are the comprehension and attention of the participants. To ensure the appropriate application of these two factors, this study has applied the following precautions.

Firstly, ease of comprehension had been and essential factor throughout the development of the game. Besides simplifying the game for the online platform, additional care and effort were spent on the clarity of the tool in its all features, including the clarity of the language, comprehension of the visuals, flow of the game and the experiment overall. Secondly, participants were provided with clear and detailed instructions of the game (*see* Appendix C-2, *p. 144*). Finally, to ensure participants' comprehension, they were strongly encouraged to read the instructions by setting a compulsory reading time before starting the game and asked for affirmation of comprehension. Participants who did not affirm that they understood the game were excluded from the study.

Regarding participants' attention, the length of the game was kept as short as possible with 15 periods. Since the task of the game is repetitive, it could have been boring and caused participants lose their attention and thus give random responses. Moreover, participants were offered a strong monetary incentive (£100.00). Utilisation of an incentive is a very common method in economics studies, especially offering real money (Katok, 2011a, b). This study applied a staged incentivisation, where a raffle was conducted among the participants who had met a certain level of minimum performance. Additionally, throughout the game, participants were informed about their progress towards the raffle eligibility limit. Doing this was targeted at consolidating participant attention and motivation. Lastly, a minimum level of performance was determined for the study, which was very low and only participants who gave random answers would perform lower than that level. Therefore, participants were clearly told that they must meet that level to be accepted in the game.

During the evolution of this study into its final form, some other manipulation control techniques were also applied, such as an extra instruction video, pre-training for the game, additional attention and comprehension checking questions, and a different method of incentivisation. However, as the game was quite simplified compared to the previous versions, it became very straightforward with a single-stage decision that repeats for 15 periods. Therefore, these precautions were gradually dropped. Details of this evolution are provided in Appendix B (*p. 130*).

4.3.5 Experiment Process

Having learnt about the variables, the next step is to combine them in this study as a complete experiment. To do so, this section will first present how this study met the internal validity requirements for an experiment, then the precautions that were taken to mitigate the low generalisability risks, and finally the application of the experiment.

a. Internal Validity - Experiment Design

Since the experiments are methods with high internal validity, there are some criteria that must be ensured and some that must be considered in the development of the tool. Their application in the design of this experiment was analysed in the following six subtitles.

I. ESSENTIAL 1: EXPOSURE OF THE TREATMENTS: WITHIN OR BETWEEN-SUBJECTS?

Utilisation of participants in an experiment is mainly done in two ways, within-subjects or between-subjects designs. In the within-subjects design, a participant receives all the treatments successively. It is used to measure the differences or temporal changes within the person. The between-subjects design, however, is utilised to measure the differences among subjects. In this design, a participant receives only one treatment (Katok, 2011a, b). The comparative analysis of both designs is shown in Table 4.5:

Table 4.5 Comparison of experiment designs

Attribute	Within-subjects	Between-subjects
Output analysis	Differences within the subject	Differences within the cases
Treatment exposure	Participant is exposed to all treatments	Participant is exposed to only one treatment
Potential nuisance variable	Order Effect	Individual differences
Participant requirement	Lower – as a participant takes all the treatments	Higher – as all treatments separately require the minimum number of participants

Reflecting on the first two attributes, this study has adopted the between-subjects design. First, this study aims to discover the performance differences among the individuals that are in different conditions. Second, as the treatments in this experiment are all designed in the same style with minor differences, when a participant receives one treatment, s/he would be familiar with the game and cannot receive another treatment, otherwise, it would harm the validity of the study. Since both these conditions can only be reached with the between-subjects design, it was applied in this study.

Regarding the other two attributes, individual differences has not been a concern for this study as a nuisance variable, since individual differences is one of the independent variables of this study. Lastly, the required number of participants were recruited for the between-subjects design, despite it being much higher than the within-design option.

II. ESSENTIAL 2: OPERATING THE TREATMENTS: FULL FACTORIAL OR FRACTIONAL DESIGN?

In the most basic experiment design, it is required to apply all treatment configurations to have a holistic observation of all possible conditions. This is called *full factorial* design. For example, if an experiment has two independent variables of a and b, 2x2=4 combinations (a1b1, a1b2, a2b1, a2b2) are analysed in the full factorial design. However, when an experiment has more independent variables, the number of combinations for the full factorial design increases dramatically. For example, a 4-variable study requires 16, and a 5-variable requires 32 treatment combinations. It may be hard to realise that, because every combination requires a minimum number of cases, in other words participants. The number of potential participants in the targeted population or in the subject pool may be inadequate. Likewise, in studies requiring a certain level of budget to prepare the conditions that treatments should have, a high number of treatment combinations would cause a large amount of cost (Katok, 2011b).

As a solution for these potential problems, experiments can also be designed in *fractional design*. To decrease the number of combinations, some may be dropped according to their impact on the study (Katok, 2011a).

This study has four independent variables, all of which have dual treatments. Therefore, a full factorial design would have required 2x2x2x2=16 treatment combinations. However, the objectives of this study do not include the interaction of all variables. The impacts of all moderating variables on the independent variable are the focus of interest, together with the impacts of time pressure and problem complexity moderators on the information presence moderator. However, the interaction between time pressure and problem complexity is not targeted. Hence, treatment combinations related to that interaction were dropped and this study adopted a fractional design in the form of three separate full factorial (2x2) designs, which resulted in 12 treatment combinations as shown in Table 4.6.

Table 4.6 Treatment combinations

Treatments	Cognitive Propensity	Information Presence	Time Pressure	Problem Complexity
1	Intuitive	Low	-	-
2	Rational	Low	-	-
3	Intuitive	High	-	-
4	Rational	High	-	-
5	Intuitive	Low	+	-
6	Rational	Low	+	-
7	Intuitive	High	+	-
8	Rational	High	+	-
9	Intuitive	Low	-	+
10	Rational	Low	-	+
11	Intuitive	High	-	+
12	Rational	High	-	+

III. ESSENTIAL 3: RANDOMISATION

The third essential factor for the design of the experiment is the random assignment of the participants to the treatments (Katok, 2011a). In this study, participants were randomly distributed into the six conditions (L--,H--,L+-,H+-,L-+,H-+) that are associated with the moderator manipulation (*see* Table 4.6). Manipulation of the independent variable (cognitive propensity), however, was done as an *ex post* analysis since it was a personal characteristic and it was not possible to manage its distribution during the experiment. Therefore, it was already randomised.

IV. POTENTIAL RISKS

Conducting an experiment is a risky situation to ensure its successful application. In experimental methods, the experimenter holds a critical place with high level of control, while the online platform greatly limits the activities of the experimenter. Therefore, experiments using online platforms are designed to be highly clear and understandable to minimise the shortcomings of experimenter inefficiency (Brañas-Garza et al., 2015). This study has paid particular attention to ensure participants' comprehension, as explained in detail in Subsection 4.3.4 (b) (*p. 69*).

However, it is important to mention a side risk that comes with the efforts related to comprehension. Providing clear instructions overall for the experiment and specifically for the game is one of the major means of supporting the tool's understandability. However, if the instruction is written in a language that directs the participant in a certain direction in the experiment, then the framing (investigator) effect can be observed. This has also been kept in mind, alongside the preparation of the instruction texts, and it was carefully avoided in order not to affect the participants' decisions in any aspect.

Another risk that is commonly met in experimental methods is the ordering effect. However, this aspect has not been evaluated as a risk for this study for three reasons. First, because of the between-subjects design, participants in this experiment do not receive more than one treatment, so, it is not applicable to this research. The ordering effect for the questions utilised in the tasks of the experiment is also not a matter of concern. CRT questions are independent questions that are not connected to each other and they measure the same constructs while they provide the same quality specifications. Questions of order management tasks, on the other hand, are individual pieces of a whole 15-period demand cycle. So, participants must follow this cycle and the order of the questions cannot be changed for any purpose.

Lastly, experiments are required to be designed to avoid framing effects if not intended as a construct of the study. One potential risk of framing could have been, in this study, an undesired impact of moderating treatments on participants' utilisation of cognitive thinking styles. Actually, the interaction of treatments and utilisation of cognitive thinking styles constructs the core of this research. The main motivation for this study has come from the deficiency of a configurational approach to the relationship between cognitive propensity and order management performance. This study has brought a solution to this deficiency by observing the performance of participants with a propensity to different cognitive thinking systems in various situations. These situations were specifically selected concerning the strengths of the thinking systems, i.e., time pressure for the intuitive thinking system and problem complexity for the rational thinking system, while the information presence condition was related to both systems and resembling their reaction towards an operational solution. It is critical to remind that despite individuals may have a propensity to use either cognitive thinking system, they are still capable to use the other thinking system up to a level. Therefore, while the main aim of this configurational approach was to observe participants' performance under a catalyser condition, their performance in the adverse condition was also another interest of this study.

However, beyond these interactions, the experiment was designed to minimise the framing effect and be neutral towards participants from either cognitive thinking group in either condition. To ensure this, the CRT section was presented to participants as the brainstorming questions to avoid participants' awareness. Because the processing of CRT questions in the mind starts subconsciously. In other words, participants are needed to be unaware of the tricks of the questions. By covering the CRT test as a brainstorming phase, this requirement was met. Likewise, in the OMG section, when participants were guided about the game in the instructions before starting the game, they were not directly mentioned that they are receiving the more complex condition or time-restricted condition. Instead, instructions were designed to provide information only about the randomly assigned condition and avoid participants compare their situation with the others. Since they were not aware of the differences among the conditions, any additional attention to intentionally utilise either thinking system was not considered as a risk for this experiment.

b. External Validity I - Generalisability

In contrast to the high internal validity, experimental methods lack external validity. In other words, because of the controlled settings and the abstractness from real-life impacts, it becomes harder to generalise the findings (Stevens, 2011). To bring solution to this, the literature provides some solutions.

Firstly, one school of thought (e.g., Boyer and Swink 2008; Grant and Wall 2009) considers that supporting the findings of laboratory experiments with another methodology, such as a survey or field experiment, can create triangulation, which boosts the external validity of a study. Another group of researchers (e.g., Cantor et al., 2014; Helfat and Peteraf 2015) believes that utilising experts in the field before and during the experiment can resolve the problems related to external validity, by verifying the suitability of the scenarios and treatments in the experiment. It is believed that expertise can compensate for the lack of generalisability, especially in comparison to participants with less knowledge or training on the topic. Third, the experimental settings can be supported by a real-life story (Creane, 2008; Cantor et al., 2014). It can be a basic cover story or a detailed scenario-based (vignette-based) experiment. Lastly, external validity of an experiment can be increased by using deception. This is commonly used in experiments in the psychology, economics, and business fields;

however, the experimenter has to be careful with the level of deception, as it may exceed the limits of ethics (Katok, 2011a,b).

Among these options, this study firstly utilised the support of expert professionals, before and during the experiment. During the preparation of the experimental tool, the experimenter's *Supply Chain Research Group* provided a continuous feedback flow as a panel. Later, SC professionals (either working in the area, having worked in or had higher education) were selected as the target population of this study (*see* Section 4.4.2, *p. 74*). Secondly, a cover story was utilised in the instructions and settings of the experiment. The SC of the story was characterised as a newspaper business SC (*see* Appendices C-1, 2, *p. 143*). To increase the reality of the context, visuals of the tools were designed to reflect that story (*see* Appendices C-1, 3, *p. 143*). Lastly, deception was also used in the CRT part of the experiment. Mentioning the questions belonging to a CRT would violate the internal validity of the test, because it is a tool to measure the subconscious cognitive processes. If the participants had been told it was a CRT, they could have behaved more cautiously and that could have altered their performance. Therefore, they were told that the test was a brainstorming test that is being applied only to prepare the participant for the OMG (*see* Appendix C-2, *p. 144*).

c. Application

The experiment was prepared in the online survey platform, named *Qualtrics*. However, for the recruitment of the participants, another online tool was utilised, *Prolific* (*see* Section 4.4.2, *p. 74*). In Prolific, the study was posted to the access of only the target subject pool by attaching a short description of its basics (*see* Appendix C-2, *p. 144*). Participants who were willing to join the study were then directed to the Qualtrics platform.

At the start of the experiment, they are given a detailed instruction regarding the overall experiment. They are provided with the Participant Information Leaflet (PIL) (*see* Appendix D-1, *p. 152*). Following their consent, they are asked for their gender and then they start with the first stage of the experiment, the CRT. They are told that it is a set of brainstorming questions that they need to complete in 5-7 minutes* (*see* Appendix C-2, *p. 144*). All the participants answer the five CRT questions, and following that, they are asked if they have seen any of the questions before (for measurement purposes).

Next, participants pass to the second phase, the *Order Management Game (OMG)*. This part was aimed to be completed in 10 minutes*. In this phase, participants are randomly distributed to the six treatment settings: low information, high information, no time pressure, high time pressure, low problem complexity, high problem complexity. For all participants, a detailed instruction for that part is also given to them and then they start the game. In the first period of the game, participants are presented with a visual of the newspaper SC with the cost, salvage and sale prices of the newspapers, and the customer demand of the last two days. After considering the demand direction and financial information, they are asked to place an order for the next day. After placing the order, they are presented with a feedback page, which shows the following information: order placed by the participant, actual demand received for that period, period-end financial analysis and cumulative financial analysis, and cumulative demand statistics (average demand and standard deviation).

The game is designed to be self-interactive. In other words, information presented in the subsequent questions is shaped according to the participants' answers. To realise that, an embedded coding tool (JavaScript) has been utilised (the codes can be found in Appendix C-4, *p. 147*). As they progress in the game throughout the 15 periods, they are presented with information that is customised according to their previous orders. Finally, at the end of the 15th period, participants are provided with their result and informed if they are eligible or not for the raffle.

* Durations were determined according to the results of the pilot studies.

After completing the whole data collection, the raffle was drawn among the 498 eligible participants, using the randomisation formula in Microsoft Excel. The promised £100.00 was sent to the participant drawn over the Prolific platform as a bonus payment. The raffle process was recorded via screen record and the other participants were also informed about the winner, together with a link to the video. Video can be reached via the link and QR code provided at Appendix C-5 (*p. 151*).

The complete tool for all settings and codes used can be found in Appendices C-1 (*p. 143*) and C-4 (*p. 147*).

4.4 Data Collection

The data collection process of this study is described in two subsequent steps. This section will first express this study's 'unit of analysis' and this will be followed with the 'sampling and participant recruitment' processes.

4.4.1 Unit of Analysis

As stated at various points, this study aims to identify the performance variations of SC managers with different cognitive propensities under three moderating conditions. It is also mentioned in Subsection 4.3.2 (b-IV) (*p. 63*)that participants will be playing the role of an SC manager in the distributor echelon of a newspaper SC. Therefore, it can be asserted that this study's unit of analysis is the SC manager.

4.4.2 Sampling and Participant Recruitment

There are four key considerations to ensure the validity of a sample: target population, sampling method, sampling platform and sample size. Each will now be discussed in turn. The first of these essential factors is the application of the study in the appropriate context, of which a part is the target population. As also mentioned in the previous section, this study measures the varying performance of SC managers. However, it is necessary to indicate what defines a 'supply chain manager' for this study. Previous research using similar inventory management games (e.g., Oliva and Gonçalves, 2005; Knemeyer and Walker Naylor, 2011) recruited students or professionals. Many studies (e.g., Katok, 2011b; Bolton, Ockenfels and Thonemann, 2012) have demonstrated that there is not much difference between the student or professional subjects; however, some (e.g., Stevens, 2011) showed strong opposition to the reliance solely on student subjects, especially undergraduates, stating that it would harm the external validity of the experiment. Therefore, considering both opinions, this study has targeted existing or potential SC managers, who are either working in a related business position or have graduated from a related course and thus have the potential to work in a related job.

The second consideration is sampling method. Sampling is the selection process of subjects from the population; it is composed of two main groups: probability and nonprobability sampling methods. To maintain the representativeness of the study, the former was selected over the latter. Probability sampling methods are further divided into six groups depending on the purpose of the study. Among those, this study has applied '*cluster sampling*'. Cluster sampling is conducted by the employment of a predetermined group or cluster that presumably has the same heterogeneous features as the rest of the population. Together with the simple random sampling and systematic sampling, this method aims to preserve the generalisability of the research. Although the other two methods provide higher generalisability, when it is not possible to reach the whole population and provide equal participation chance, cluster sampling may result in higher validity (Sekaran, 2016).

The cluster of this study was found via online participant recruitment platforms. In the behavioural decision-making field, as well as in behavioural operations, the recruitment of participants from online platforms is a common practice. *Amazon Mechanical Turk (MTurk)* and *Prolific* are the most frequently used platforms. This study has used *Prolific,* as its pre-screening options allowed a more suitable subject recruitment for the target population of this study. Furthermore, some studies have found that MTurk utilisation can have distorting impacts on the CRT studies, caused by multiple exposure (*see* Stieger and Reips, 2016; Thomas and Clifford, 2017). Prolific was considered as a safer option for the validity of the study.

In Prolific, the number of potential participants – namely the size of the cluster of the target population – was 2,143. Out of that number, it was targeted to recruit 600 participants. In experimental studies, a minimum of 30 observations are targeted per treatment. This comes from the mainstream analysis techniques that are used in experimental methods such as ANOVA or t-test that require fitness to the Central Limit Theorem. Initially it was targeted to have 12x30=360 valid observations. However, since the manipulation of the cognitive propensity is not randomly conducted by the researcher but manipulated via *ex post* analysis, it is not possible to have perfectly equal distribution even though the CRT questions were designed to produce this (*see* Subsection 4.3.2 [a-VI], *p. 60*). In addition, invalid observations must be considered, which will be excluded from the final number. Therefore, to ensure the minimum number of 30 cases per treatment, the target was 50 participants per treatment, which results in a total of 600.

4.4.3 External Validity II – Screening

To ensure the validity of measurement, another important point is to control the potential interference of any statistical noise. To do this, observations are screened in two approaches: *ex ante* and *ex post* (Thomas and Clifford, 2017).

For the *ex ante* screening, the whole Prolific subject pool was pre-screened to provide a sample cluster of the target population. The screening was done according to their background and only potential participants with background education in the '*Business, Management and Marketing*' fields were identified as eligible to voluntarily join the study.

There could have been more *ex ante* screening, such as job specification; however, it would possibly eliminate numerous suitable candidates for the study. Therefore, the remaining participant eligibility control was done via *ex post* screening. For the elimination of these unsuitable cases, participants' answers in the CRT and OMG were analysed and the suspicious or outlying answers were excluded in order to reduce the statistical noise and preserve the validity of the findings. Some examples of the suspicious answers were 1) the same figure given to all questions and 2) overly low or high figures in comparison to the expected response range. Outlying answers, on the other hand, were identified by the participants' OMG performance. While they play the OMG, they were supposed to make a profit by their decisions and ideally reach the threshold of SIM£1,500 if they understand the game and concentrate on it. Participants who have negative profit results were accepted as outliers, considering either they did not completely understand the flow of the game or they were not sufficiently attentive throughout the game.

4.5 Data Analysis

Data analysis was structured around three steps: data cleansing, preparation of the test samples for the hypotheses, and application of the main tests.

First, before any analysis, evident outliers that show no sign of attention, but appear to give random and inattentive answers, were eliminated. Some examples to these invalid participations are as providing the same answer for all the questions or too high numbers that cannot be a product of attentive calculation.

Secondly, the raw dataset was transformed into the manipulated performances that are ready to compare. It started with the calculation of participant orders' deviations from the actual demands for 15 periods. Next, these 15 periods were divided into four phases as seen in Figure 4.5.

To recap, the independent variable of this study was the order placement performance of an SC manager, which is to be measured by the capability to follow the changes in customer demand. High follow-up capability required a timely and accurate reaction to the change in the direction of the demand. Thinking backwards, a late or slow reaction to change meant the underreaction of a participant. Underreaction was utilised to measure the lack of demand follow-up capability of a participant and it was observed by greater deviation from the incoming order.

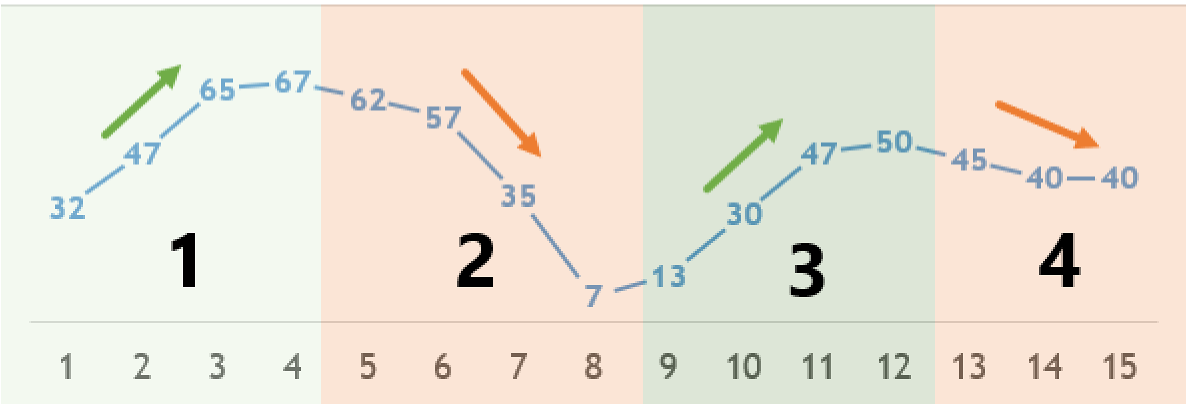


Figure 4.5 Separation of the periods

Throughout the 15 periods of this simulation, the direction of the customer demand was changed three times during the phases, as shown by the arrows (Figure 4.5). In line with these changes, two categories were created: change to decrease (combination of phases 2 and 4) and change to increase. When the demand started to decrease, participants' underreaction was detected by the continuing high ordering, called overordering. Likewise, in underreaction to the change to increase, the underreaction's indicator has been underordering. Considering these measurement criteria, initially two proxies were created: underordering in the change to increase (coded as CINCUN, representing **C**hange-to-**INC**rease-**UN**derorder) and overordering in the change to decrease (coded as CDECOV, representing **C**hange-to-**DEC**rease-**OV**erorder). The addition of these proxies led to the primary construct that was used to measure the participant performance and to test the hypotheses: underreaction to change in either direction (CUN).

This study has utilised another proxy as well, namely cdecov, as the secondary indicator of the performance. This addition of proxy was based on two reasons. The first reason was related to the potential impact of profit and loss balance of the designed OMG. In the instructions (see Appendix C-2, *p. 143*), participants were informed that the cost of material was SIM£2, the sale price is SIM£5 and the salvage cost is SIM£1 if there is any unsold stock left at the end of the period. Therefore, when participants overorder, they lose SIM£1 per unsold item (material cost-salvage cost), whereas they lose SIM£3 per unmet customer demand (sale price – material cost). Considering this cost asymmetry, overordering may be preferable compared to underordering and it may impact the participants' behaviour by leading them to be more flexible for overordering and less for underordering.

Additionally, the seasonal trend of the demand has also had an impact on the management of ordering behaviour. OMG starts with an increasing demand where participants are expected to increase their placed orders in line with the increasing incoming demand. However, since they are also trying to understand how the game proceeds and demand increases, it was also taken into account that they will be wary in comparison to the upcoming periods of the game. Therefore, they were not expected to easily place dramatically high orders. In the second demand increase stage, however, they were believed to be more reactive than the first stage as they were going to have a clear understating of the flow of the game. Likewise, in overordering, participants are again expected to be wary while they follow the decreasing demand. However, they were also considered to be more flexible about overordering and its additional cost of as the loss of overordering is a third of underordering.

To sum, it was considered that participants may show higher variance when they may potentially overorder and lower variance in periods that may lead to underordering. This behaviour-triggering difference between over and underordering was taken into consideration in the analysis stage. While the main measurement construct was designed as the total variance between the placed order and received demand (Cun) (for Stages 2,3,4), variance in the overordering stages (Stage 2 and 4) was also created as the secondary measurement construct. In this way, the variance of the overordering stages was isolated potential diminishing impact of the underordering stage (Stage 3).

Another reason for this dual measurement method was related to the spectrum difference of directions. In other words, while there is no limit to underreact (overorder) when the demand started to decrease, the underreaction (underorder) to the demand that starts to increase has a minimum limit of 0. Therefore, although CUN is the primary indicator of performance, it may be influenced by CINCUN's one-sided limit of underreaction. Because of that, CDECOV was selected as the secondary indicator to ensure the results of hypotheses tests.

Both indicators were prepared for each hypothesis sample, and then the outliers were eliminated from the samples to avoid any statistical interference issues. With this step, the preparation of the dataset was completed for the test of hypotheses.

As mentioned in Section 4.2.3 (*p. 51*), experimental methods are mainly utilised to test hypotheses in relation to the causal relationships. To do so, independent variables are manipulated into two treatments, where generally one represents the control group and the other is the test group. In line with this, analysis of experiments mainly focuses on the comparison of these two groups. This study has also applied a similar analysis approach by comparing the differences among the treatments in accordance with the associated hypotheses.

To comparatively analyse two groups, there are frequently utilised analysis methods such as ANOVA and t-tests. The selection of the analysis method depends on various criteria. As well as the method being appropriate to enable reaching the targets of the study, the dataset must also be suitable for the selected method. To apply a type of t-test, one of the requirements of the dataset is its fit to the Central Limit Theorem. In other words, this method can measure the differences among the samples that are normally distributed. When this condition is not met, another set of tests can be applied, which are called 'non-parameter' or 'distribution-free' tests.

In line with this requirement, this study first checked the normality of the samples for each hypothesis test using the *Shapiro-Wilk Test*. While some tests were found to fit the normality requirement, there were some that violated it. Therefore, a non-parameter counterpart of the two-sample t-test, namely the *One-way Mann-Whitney U-Test*, was chosen for this study. This method was also used in the preceding studies, such as Narayanan and Moritz (2015). Since the indicators, i.e. deviation, have only one direction, the test was conducted as *one-sided*.

Additionally, to control the impacts of nuisance variables, some further tests were also utilised. To measure the impact of previous exposure on the CRT performance, *linear regression test* was used. To detect any potential learning curve, however, participants' performances at the beginning and end of the game was compared by using *independent t-test*.

Analysis was conducted via *Microsoft Office Excel* and *SPSS* programs. While the former was used to clean the dataset and prepare the right format, the latter was utilised to apply the required statistical analysis to test the study's hypotheses and reach the results.

4.5.1 Conclusion Validity – Analysis

Conclusion validity relates to the reliability of the analysis and the resulting outcomes. It is predominantly based on two aspects: reliability and statistical power. The former was enabled by the outlier elimination. All the sample sets that are independently used in hypothesis tests were screened in advance of the analysis, so that analysis has not been skewed by extreme observations. Secondly, the statistical power of $p<0.05$ was targeted but, in order not to miss any important interference (Type II error), cases with approximate p values were also discussed to provide insight for the future research directions (Trochim and Donnelly, 2001).

Besides, the family-wise error rates (FWER) of the hypothesis tests were also calculated with the following formulae: $FWER = 1 - (1-\alpha)^n$. Even though the FWERs for the hypothesis groups are higher than the threshold level of 0.05, the analysis and the inferences in this study focused on the significance of individual hypotheses.

4.6 Quality of the Research

All types of studies are expected to show rigour, no matter what method or research technique they use. Stentoft and Rajkumar (2018) categorise the essential quality requirements of a study in two phases and three groups as illustrated in Figure 4.6:

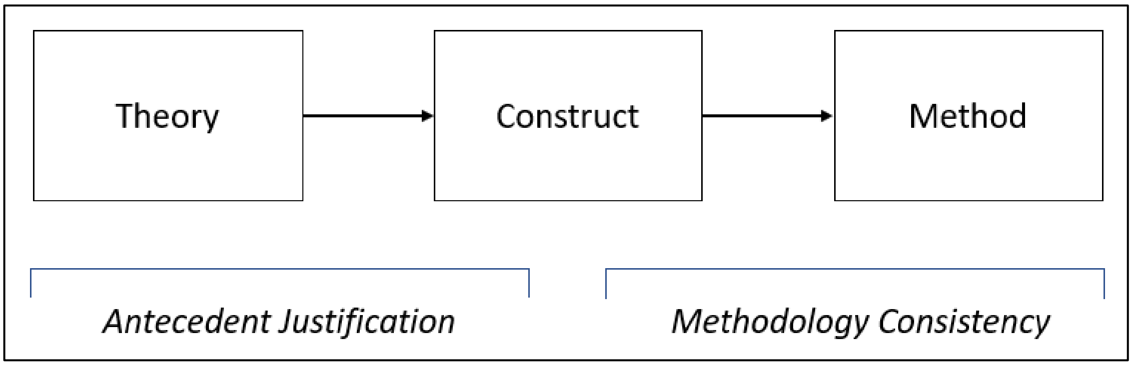


Figure 4.6 Rigour consistency of a research

The guiding works of Trochim and Donnelly (2001), Bachrach and Bendoly (2011), Katok (2011a,b), Thomas et al. (2017) and Castañeda (2019) elaborate on these specific elements. However, instead of putting them all into a list, it is better to synthesise and reinterpret them for each study. In line with this, the quality of this research was investigated under the following four main titles which have been discussed in the most relevant position throughout the chapter:

- **Construct validity** was analysed in two parts: control of nuisance variables (Section 4.3.3, *p. 67*) and control of operability (Section 4.3.4, *p. 68*).
- **Internal validity** investigated the design features of the experiment (Section 4.3.5 [a], *p. 69*)

- **External validity** was also handled in two parts. The first part explained the action to ensure the generalisability of the study (Section 4.3.5 [b], *p. 72*), while the second part focused on the screening actions undertaken throughout the participant recruitment (Section 4.4.3, *p. 75*).
- The last quality control section (Section 4.5.1, *p. 78*) was on **conclusion validity** which looked into the reliability of the analysis.

4.7 Ethics

Throughout the study, ethics have been at the centre of this study's considerations. Ethical approval was granted from the University of Warwick's BSREC Research Ethics Sub-Committee with the reference number of BSREC 25/19-20 (Amendment Number: 02) (*see* Appendix D-2, *p. 152*). Before starting any data collection, the first approval was given on 13 December 2019 and, as the study evolved during the pilot studies, the ethical approval was updated two more times until its final form's approval on 09 June 2020.

Although the ethical approval form can be reached via the reference number on request, it is practical to express some highlights of the ethical considerations of this study. The first critical point is the provision of confidentiality and data anonymisation. These have been ensured from the very beginning, when recruiting participants via Prolific. Prolific keeps the participant information private and the cases were only identified by participants' Prolific ID numbers. The Prolific ID number was only used for two purposes: confirmation of valid participation after completion and the raffle among the eligible participants. In relation to confidentiality, this study has not collected any personal information from the participants, expect their gender.

Secondly, participants were paid in return for their time and effort. Each participant was paid £2.10. This amount was calculated according to the recommended price range of Prolific. While the average recommended payment was £7.80 per hour, this study decided to pay higher than the average to incentivise participation and paid £8.40 per hour, which made the payment £2.10 for the length of this study – 15 minutes.

Lastly, another critical aspect of ethics is the participant consent. A Participant Information Leaflet (PIL; *see* Appendix D-1, *p. 152*) was provided to the participants in a downloadable form at the beginning of the study. It included all the details related to the study from the ethical issues, such as confidentiality, anonymisation, withdrawal and compensation, to the details of the study. Participants were encouraged to read the PIL, and their start in the experiment has been taken as their consent.

4.8 Chapter Summary

This chapter explained the methodological backbone of this research in detail. It started with the fundamentals of the research in relation to the philosophical approach of the researcher and the selected methodology. Next, the application of the experiment method was explained, including the definition and utilisation of variables, validity of the employed constructs, design phase of the experiment and its application in the field. The chapter then presented the key points of data collection and analysis. The quality of the research has been at the core of the study and the validity and rigour of each step was justified where appropriate. Another key consideration has been the alignment of decisions regarding the research philosophy, research approach and method, data collection and analysis methods throughout the research. The preparation of the experiment tool has taken a long time and effort. It has gradually evolved in its latest form over eight months. The three pilot studies are worth looking at to better understand the study and can be seen in the appendices, including all the justification for the amendments.

5. Results

"In nature we never see anything isolated, but everything in connection with something else which is before it, beside it, under it and over it."

JOHANN WOLFGANG VON GOETHE

5.1 Chapter Introduction

The results chapter of this study is composed of four major sections.

To begin with, Section 5.2 presents the descriptive analysis of the findings, including the distribution of the participants in the pilot and main studies, gender distribution and the completion time.

Next, Section 5.3 analyses the required quality control checks. Subsections subsequently check the impacts of nuisance variables and operability, which are followed by the design control of randomisation.

Section 5.4 then explains the preparation of the required datasets for the main analysis. This includes the coding of the measurement constructs, exclusion of outliers and finally the normality control of measurement samples.

Lastly, Section 5.5 presents the main analysis visually and statistically in three hypothesis categories, followed by the chapter summary at Section 5.6.

5.2 Descriptive Results

This study recruited a total of 755 participants. Initially, the first pilot study was distributed within the personal circle of the researcher via social media connections and mails. In total, 239 people were asked to participate and 85 of them joined the study (35.56%). Out of the 85 participants, however, only 31 have completed the pilot study in full (36.47%). The low rate of participation and completion as well as the user feedback on the complexity of the online environment have led the experimenter to consider the applicability of the designed game. The further analysis of the first pilot study is provided in Appendix B-1(III) (*p. 137*).

The tool was revised with the guidance and recommendations provided by the researcher's supervisors and research group colleagues. The revised tool, which is the current version explained in Chapter 4, was tested with 26 participants that were recruited via Prolific. The manipulation and measurement capabilities of the new tool were controlled (*see* Appendix B-2, *p. 142*). After seeing its validity via CRT manipulation power, previous exposure of CRT questions, successful completion rate for the OMG, it was accepted as a valid part of the main data collection which was continued via Prolific.

For the main study, 670 participants were recruited in total. 47 cases were initially excluded; 21 because of inattentive and random answering and 26 for quitting or answering too quickly. The distribution of participants is illustrated in Figure 5.1.

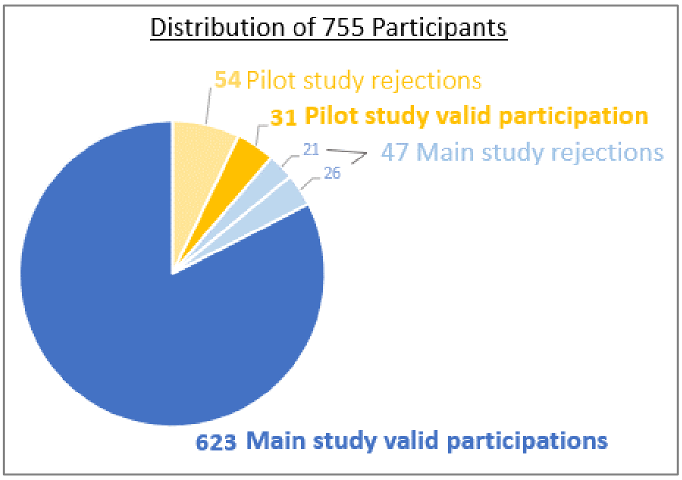
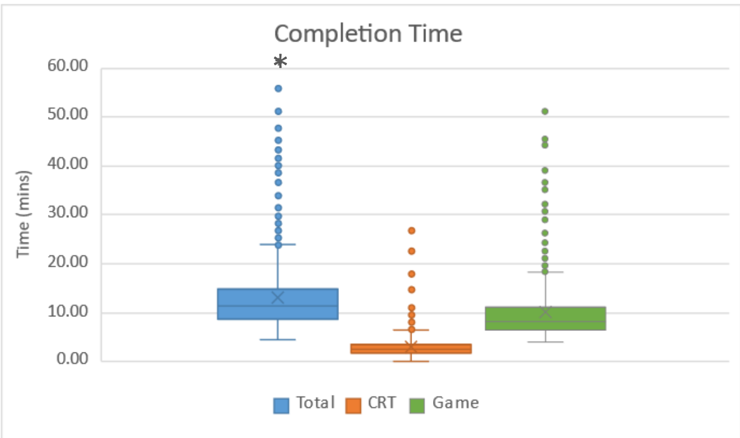


Figure 5.1 Distribution of total participants

As explained in Chapter 4, the experiment tool consisted of two main sections: the *CRT* and *OMG.* Before starting the CRT section, participants were also asked about their gender. It was intended to have an equal distribution of gender to avoid any gender-related interference. The analysis has shown that 58% of the participants were male and 42% were female. Although the distribution is close to the target, any potential impact of gender was controlled for, as explained in Subsection 4.3.3 (a) (*p. 67*) (*see* also Subsection 5.3.1[a], *p. 82*).

The experiment tool was expected to be completed in approximately 15 minutes. The results show that the average completion time was close to the target with 13.15 minutes (min: 4.43, max: 113.83). The outlier analysis was also done to eliminate any possible skew of mean value by excluding the values over 2 interquartile range from the inner quartiles. 39 cases (6%) were excluded from the upper extreme and the revised average of completion time was calculated as 11.64 minutes (*see* Figure 5.2).



	CRT	Game	Total
Mean	2.81	9.05	12.11
SD	4.68	1.50	3.56

Figure 5.2 Boxplot of the completion time of the experiment

5.3 Quality Control Analysis

The previous chapter explained that this study has ensured the construct validity in two stages. First, through the control of nuisance variables' impact on the main variables of this study and then through the check of operability. Each will now be discussed in turn.

5.3.1 Control of Nuisance variables

a. CRT-related Nuisance Variables

Subsection 4.3.3 [a] (*p. 67*) has stated two critical nuisance variables: gender control and previous exposure.

With the exception of the second pilot study (26 observations), all participants were asked for their gender. Of the 597 participants remaining in the main study, 348 were male (58%) and 248 were female (42%)*. First, the distribution of thinking systems was compared between genders (Figure 5.3). The distributions of thinking dispositions for the total sample and for genders were well balanced across both genders.

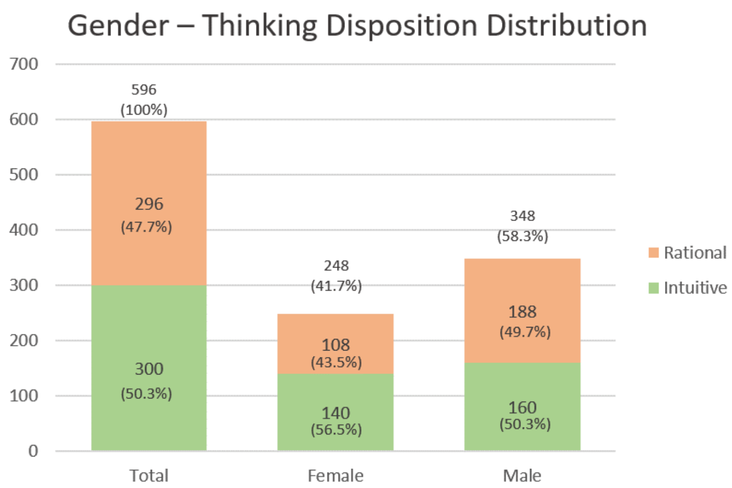


Figure 5.3 Gender – Thinking disposition distribution

Secondly, the distribution of the genders within the thinking dispositions was also analysed, as it was the main reason for controlling the gender ratios. Previous literature has shown a difference among the CRT performance of male and female participants, where generally male participants perform better in the test. Therefore, it was expected to obtain a similar result and it turned out to be so, as demonstrated in Figure 5.4.

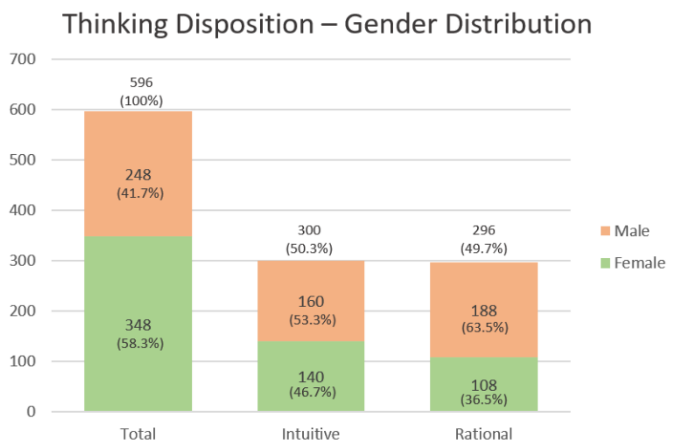


Figure 5.4 Gender – Thinking disposition distribution

* One participant refused to state gender information.

As the figure shows, the distribution of participants who are inclined towards utilising their rational thinking systems more in their decisions is skewed towards the male category with 63.5% as opposed to 36.5% in the female category. However, it is also seen that this skew has not harmed the distribution of thinking systems overall in the experiment, which ended up with nearly 50% distribution of participants for each thinking system. Considering the target was to have either category with at least a 40% ratio, this target was reached with a highly equal distribution, without being affected by the potential skew threat of gender distribution.

The second nuisance variable related to the CRT's manipulation validity is participants' previous exposure to the CRT questions. After completing the CRT questions, participants were asked to assert if they had seen the questions before joining this study. In this study, 623 participants were asked five CRT questions. Hence, from a total of 3,115 CRT exposures, only 30.9% (964) of the questions were seen by the participants prior to this study. Figure 5.5 presents the distribution of a) participants according to their number of previously seen questions and b) the comparison of questions.

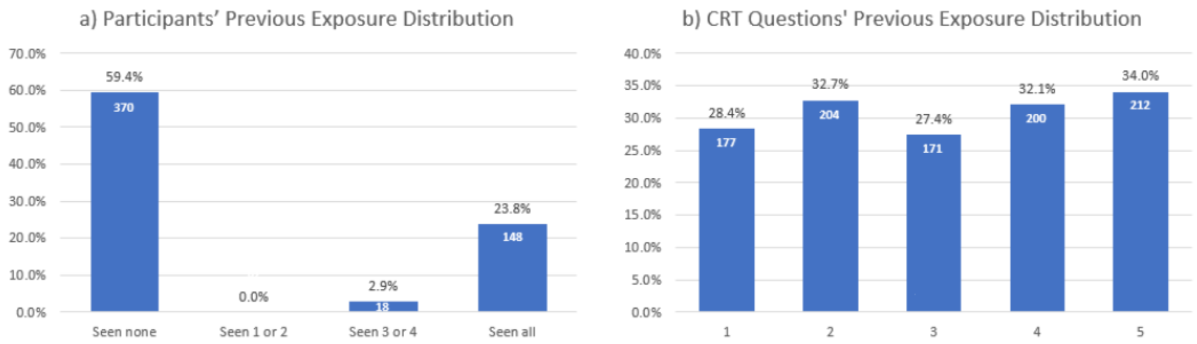


Figure 5.5 Previous exposure distribution for a) participants and b) CRT questions

As the figure depicts, more than 50% of the participants have not seen any of the selected five CRT questions before, while only nearly 20% of the utilised sample have seen all the questions. When analysing questions individually, on the other hand, all questions have similar previous exposure, ranging between 27.4% and 34.0%. Considering previous exposure concern was one of the aspects to select the questions for the customised CRT of this study (*see* Subsection 4.3.2 [a-V], *p. 59*), these findings show that the process of selecting the CRT questions was successfully applied in this aspect.

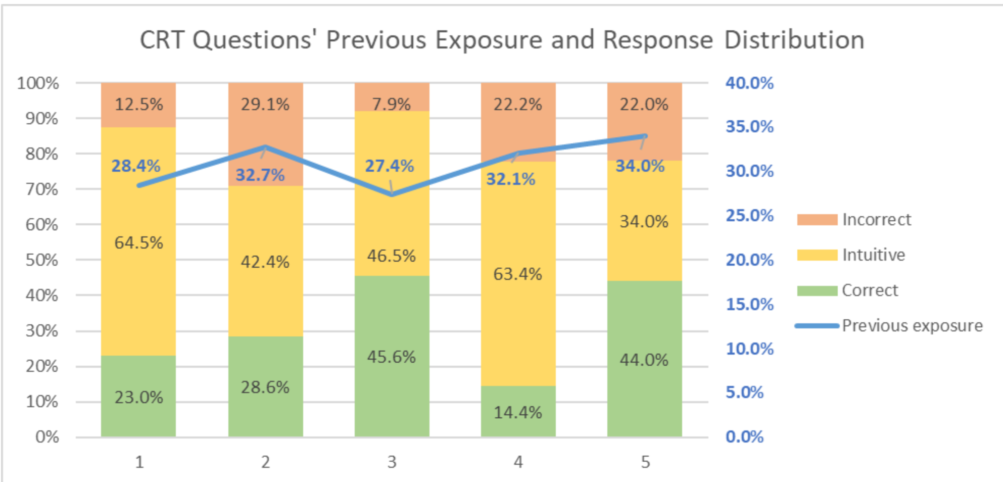


Figure 5.6 Previous exposure distribution alongside CRT responses

When it comes to the relationship between previous exposure and CRT performance, Figure 5.6 provides the initial descriptive analysis. As seen, the trend of previous exposure (blue line) does not look to be associated with the ratio trend of correct answers per questions (green bars). Therefore, it is presumed that the reason for the variation in the correct answer amounts of questions is not related to the previous exposure. In another aspect, the overall performance difference between participants who had seen the majority of the questions before (>2) and those who had not (≤2) was also visualised. As seen in Figure 5.7, a similar pattern is observed for both conditions, meaning the absence of a difference coming from the exposure.

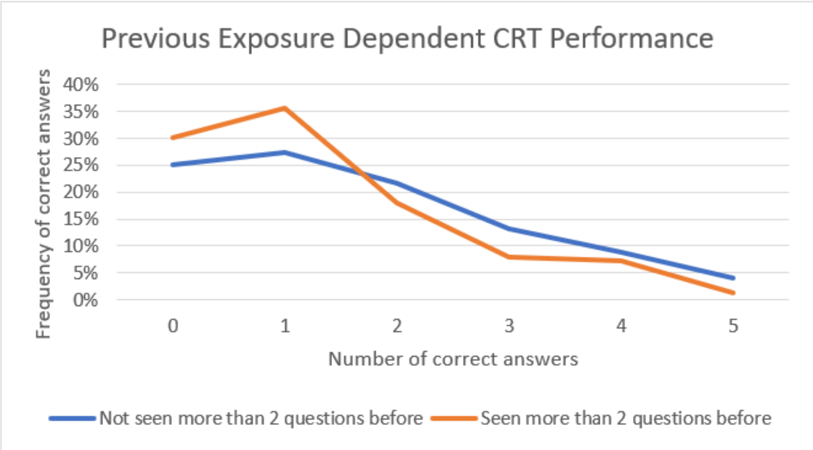


Figure 5.7 Previous exposure-dependent CRT performance

However, to reach a statistically dependable conclusion, the impact of previous exposure on the CRT performance was tested by using linear regression. The results, conversely, showed a statistically significant impact of previous exposure on the CRT performance (p=0.036) with a negative way interaction (β=-0.55). However, the explanation power of these results was extremely low (R^2=0.007), meaning it has a very minor impact on the explanation of CRT performance differences (*see* Appendix E-1 for full analysis table, *p. 152*).

Any CRT performance difference related to the previous exposure to the CRT questions was not detected. In other words, the CRT manipulation of this study is not impacted by the potential nuisance variable of previous exposure.

b. OMG-related Nuisance Variables

There has been only one nuisance variable determined regarding the measurement of independent variables in this study: learning curve. To measure the impact of this nuisance variable, the order management performance of participants was comparatively measured from two opposite points of the game. These points were determined as the 5^{th} and 13^{th} periods as being the first days of decreasing demand trend.

The comparison test was conducted for each of 12 treatment conditions by using independent samples' t-tests. The following test results with the visualised graphs summarise the results of the tests (*see* Appendix E-2 for full analysis table, *p. 153*). According to the findings, a learning curve – meaning a positive way performance difference between the beginning and end performances – is not seen in any treatment. Instead, generally a negative performance is seen in Period 13, compared to a similar situation in Period 5. Although this decrease in performance is not statistically significant for intuitive participants in most of the treatments, the rational group shows a poorer performance in Period 13 in all treatments except one.

Figure 5.8 Visualised comparison of beginning and end performances

This may be related to the high effort asserted by the rational thinking system for a task. As the high effort consumption does not last for long, the effective performance of participants might have been attenuated as the game progressed. However, the length of the task can be another research aspect for future studies.

5.3.2 Control of Operability

The second aspect of the construct validity was the appropriate operability of the constructs as targeted.

a. Manipulation Check of CRT

The manipulation capabilities of the utilised CRT questions were measured from two aspects: distribution of the answers and the response time.

A CRT question is expected to possess some level of impulsive trigger as well as a similar level of difficulty. In other words, it should be tricky enough to deceive intuitive participants but not all. Likewise, it should be difficult enough to be solved correctly only by rational participants. Previous studies have identified a balanced ratio in the distribution of the answers (*see* Subsection 4.3.4 [a], *p. 68*). Therefore, a similar balance was also searched for in the results of this study.

At the beginning of the chapter, it was shown that the *ex post* manipulation of the CRT provided an almost level distribution between the two thinking dispositions: 50.3% intuitive and 49.7% rational. This level separation was searched for each of the five CRT questions (Figure 5.9). Although some questions are more impulsive (1 and 4) and some are easier to solve (3 and 5), out of 3,115 questions (5x623), nearly half of them were replied to with an intuitive approach, while for the other half, participants utilised the rational approach (correct + non-intuitive incorrect answers).

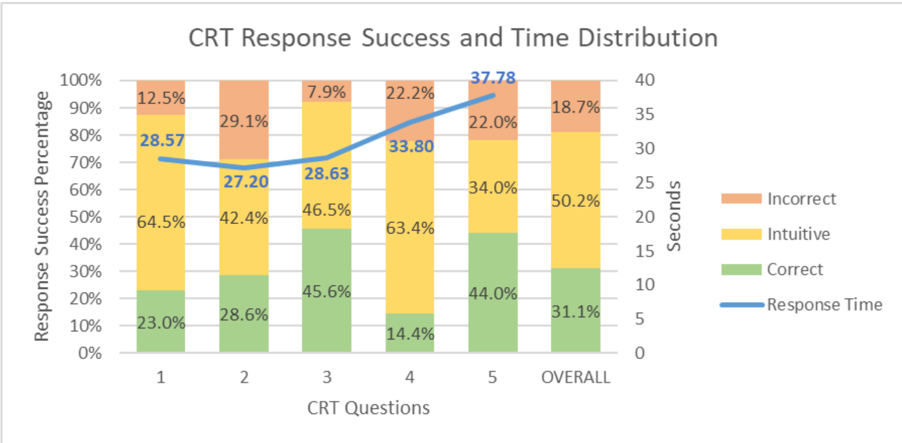


Figure 5.9 CRT response success and time distribution

The manipulation power of the questions was also analysed by participants' response times (Figure 5.9). In the instructions, participants were recommended to spend 5-7 minutes for the CRT section; however, the section was completed in about three minutes. The average answer time varied between 27.2-37.78 seconds. At first sight, it was considered that the test was easier than targeted. However, no trend is seen between the response time and success. An increase in the response time is observed, which can be related to the decreasing cognitive performance. Nevertheless, it was observed that there is no unusual observation related to the response time and success ratios, and in the overall analysis, the targeted manipulation performance was reached.

b. Measurability Check of OMG

As explained in Subsection 4.3.4 [b] (*p. 69*), the comprehension and attention of participants were maintained by careful considerations throughout the design of the OMG. Therefore, the need for additional *ex post* comprehension and attention checks were removed from the study, while in the pilot stages multiple control questions were effectively utilised (*see* Appendix B-1, *p. 131*).

The success of participants in the game can be an indicator of the overall clarity and measurability of the game. Figure 5.10 presents the comparative ratios of participants with different success levels. As nearly 75% of whole sample managed to pass the raffle eligibility level, it can be inferred that the game has possessed a sufficient level of comprehensibility for the majority of participants.

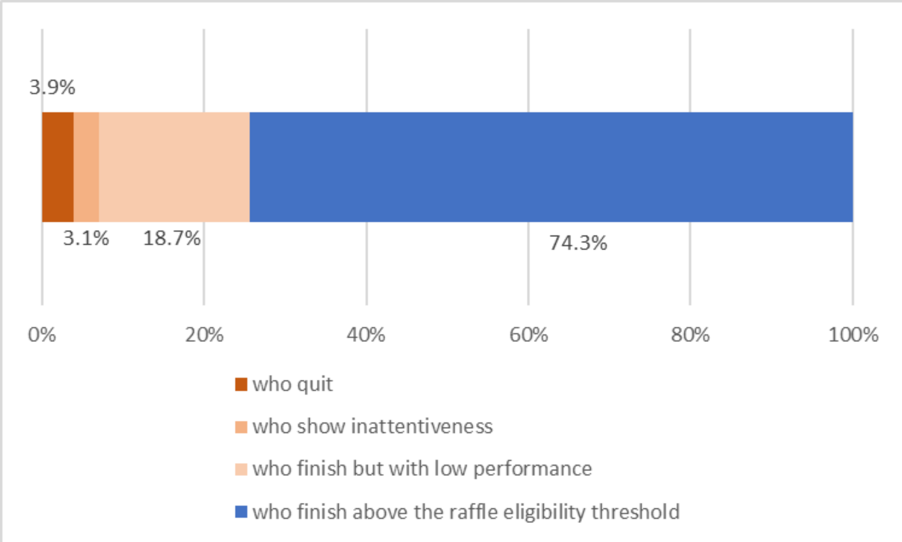


Figure 5.10 Participant performance distribution

5.3.3 Control of Internal Validity: Randomisation

The internal validity of this study has been determined by following three key experiment design rules (*see* Subsection 4.3.5 [a], *p. 69*). While the first two rules – decisions between within-design or between design and full-factorial or fractional design – were determined in the design process of this experiment, randomisation required a control because of the *ex post* distribution of the independent variable (cognitive propensity). Each treatment category was targeted to provide a minimum of 30 valid cases. The target was reached, as seen in Table 5.1.

As shown, the distribution of six scenarios varied between 95 and 119, with an average of 104 cases. When they are also separated into the cognitive propensity, we reach the number of treatment combinations. Table 5.1 demonstrates that all the treatments have had higher than the threshold number of 30, with the minimum value of 36.

Table 5.1 Distribution of participants throughout the treatment combinations

Information Presence	Cognitive Propensity	Baseline	Time Pressure	Problem Complexity
Low	Intuitive	36	62	49
	Rational	60	57	50
Total		96	119	99
High	Intuitive	47	65	52
	Rational	48	52	45
Total		95	117	97

5.4 Pre-analysis Dataset Preparation

Before testing the analysis, initially datasets were prepared in three stages: preparation of constructs from the raw database, defining and removing the outliers for each test sample, and finally checking the normality.

The raw dataset that was exported from the Qualtrics platform had provided all types of required information, from the answers to time spent per question. However, it was necessary to prepare the specific sample datasets for each hypothesis test. This was completed in two stages. First, by analysing participants' CRT answers, they were manipulated into the predefined groups of intuitive and rational dispositions, as explained in Subsection 4.3.2 (a-VI) (*p. 60*).

Then, for the measurement of participants' performance in the OMG, the predetermined measurement constructs were created by using the answers given to 15 periods. Section 4.5 explained the two measurement constructs used in this study. In line with those instructions, measurement samples were prepared for each hypothesis and, to ease the analysis, these samples were coded in a systematic way. Overall, the hypotheses have compared either of the two groups: the difference between two cognitive propensity groups within a treatment, and a propensity group's performance difference in two treatments. The tests for the former condition were coded starting with treatments numbered from T1 to T6. Likewise, for the second condition, the tests were coded for the propensity groups with P1 or P2. Lastly, the code indicates the analysis groups. For example, when comparing the intuitive propensity group's treatments 2 and 4 performance difference, it was coded as P1.2v4. Lastly, the analysis constructs were added to the code. For a better comprehension of the coding, Figure 5.11 provides explanatory examples for both hypothesis types.

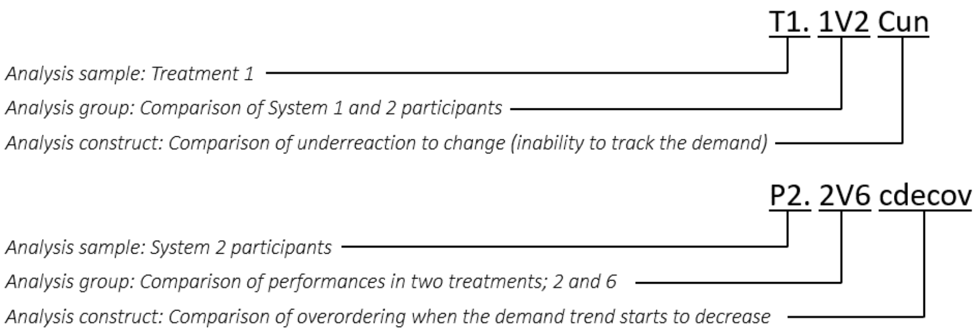


Figure 5.11 Coding examples for hypothesis measurement samples

After preparing the 26 measurement samples for 13 hypotheses, these datasets were transferred to SPSS software. Initially, outliers of all measurement samples were defined and excluded. The total exclusion ratio is 1%, with the number being 140 out of 13,052 cases.

At this point, the datasets have become ready for the targeted tests. As mentioned in Section 4.5, it was necessary to check the conformity of the samples in normal distribution to determine the analysis technique. For that purpose, the Shapiro-Wilk test was applied for all the measurement samples. As seen in Table 5.2, out of 26 measurement samples, 21 have rejected the null hypothesis which asserts that the measured sample is normally distributed (except the highlighted ones). Therefore, to apply the same technique for all and to reach standardised results, all hypotheses were tested via the non-parametric counterpart of the t-test, namely the Mann-Whitney U-Test.

Table 5.2 Normality tests for the measurement samples

		Kolmogorov-Smirnov			Shapiro-Wilk		
	Hypotheses	Statistic	df	Sig.	Statistic	df	Sig.
H1	T1.1V2Cun	0.078	93	.200	0.979	93	0.132
	T1.1V2cdecov	0.073	93	.200	0.977	93	0.104
H2a	P2.1V2Cun	0.085	104	0.064	0.97	104	0.018
	P2.1V2cdecov	0.121	104	0.001	0.949	104	0.001
H2b	P1.1V2Cun	0.104	80	0.032	0.974	80	0.105
	P1.1V2cdecov	0.076	80	.200	0.985	80	0.479
H3a	T3.1V2Cun	0.149	115	0	0.885	115	.000
	T3.1V2cdecov	0.118	115	0	0.897	115	0.00
H3b	P1.1V3Cun	0.12	96	0.002	0.941	96	.000
	P1.1V3cdecov	0.103	96	0.014	0.95	96	0.001
H3c	P2.1V3Cun	0.104	111	0.005	0.943	111	.000
	P2.1V3cdecov	0.108	111	0.003	0.919	111	.000
H3d	P1.2V4Cun	0.089	107	0.037	0.976	107	0.046
	P1.2V4cdecov	0.06	107	.200	0.984	107	0.237
H3e	P2.2V4Cun	0.099	97	0.021	0.952	97	0.001
	P2.2V4cdecov	0.127	97	0.001	0.934	97	.000
H4a	T5.1V2Cun	0.125	92	0.001	0.893	92	.000
	T5.1V2cdecov	0.141	92	0	0.894	92	.000
H4b	P1.1V5Cun	0.133	79	0.001	0.925	79	.000
	P1.1V5cdecov	0.13	79	0.002	0.9	79	.000
H4c	P2.1V5Cun	0.13	106	0	0.915	106	.000
	P2.1V5cdecov	0.116	106	0.001	0.936	106	.000
H4d	P1.2V6Cun	0.097	91	0.033	0.966	91	0.017
	P1.2V6cdecov	0.103	91	0.019	0.963	91	0.012
H4e	P2.2V6Cun	0.088	87	0.091	0.942	87	0.001
	P2.2V6cdecov	0.135	87	0	0.936	87	.000

5.5 Main Study Analysis

5.5.1 Baseline Hypotheses

Following the preparation of the datasets, hypotheses were tested using a One-sided Mann-Whitney U-Test, starting with the baseline hypotheses.

Initially analysing the relationships visually, it was seen that all comparison groups show a similar trend of deviation from the actual order received (Figure 5.12). However, in the statistical analysis (Table 5.3), all null hypotheses based on the similarity of comparison groups were rejected. This means that the compared measurement samples do not belong to the same population. In other words, there are performance differences between the compared groups.

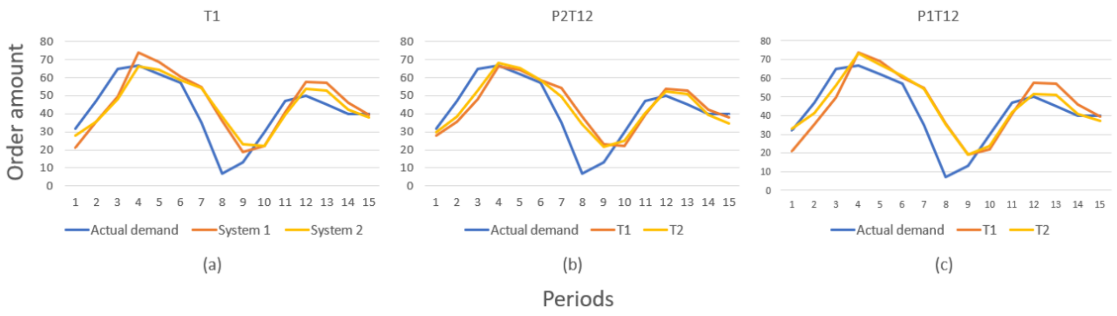


Figure 5.12 Average order comparison for baseline hypotheses (a) H1, (b) H2a, (c) H2b

In Table 5.3, we can see that the *cdecov* constructs have provided a higher statistical significance with a confidence level of 95%, while *Cun* constructs have provided 90% confidence. The reason for this difference is assumed to be related to the limit of possible overreaction when the order started to increase. Explaining with an example, when the actual order falls below 10 units at Period 8, if participants cannot forecast the upcoming increase in demand, they would continue to place a lower order; in other words, underreact to the change. However, since their minimum order is 0, their underreaction is limited. On the other hand, if they underreact to a demand change towards decrease, they can continue to place higher orders as much as they want without any limitation. For this reason, *cdecov* – overreaction given to the demand starting to decrease – may show higher statistical significance.

Table 5.3 Mann-Whitney U-Test results for baseline hypotheses

Hypotheses	Measurement sample and comparison group description	Total N	1st N	2nd N	1st Group Mean	2nd Group Mean	1st Group Mean Rank	2nd Group Mean Rank	Asymptotic Sig. (1-sided test)
H1*	T1.1V2Cun across CT1.1V2	93	34	59	160.7	150.9	52.1	44.06	0.083
	T1.1V2cdecov across CT1.1V2	93	34	59	81.59	72.95	53.68	43.15	0.035
H2a*	P2.1V2Cun across CP2.1V2	103	59	44	150.9	129.3	55.95	46.7	0.06
	P2.1V2cdecov across CP2.1V2	103	59	44	72.95	62.45	56.33	48.73	0.044
H2b	P1.1V2Cun across CP1.1V2	80	34	46	160.7	142.8	44.88	37.26	0.074
	P1.1V2cdecov across CP1.1V2	80	34	46	81.59	69.2	46.51	36.05	0.023

*Affirmed with 95% confidence, Family-wise error rate: 0.14

These statistical inferences showing the performance differences have affirmed the H1 and H2a hypotheses of this study:

H1: Managers with a cognitive propensity towards their rational thinking system (System 2) perform better than managers who have a propensity to use their intuitive thinking system (System 1) in tracking the customer demand to mitigate the bullwhip effect. (*AFFIRMED, p=0.035*)

H2a: Managers with a cognitive propensity for their rational thinking system (System 2) perform better in tracking the customer demand to mitigate the bullwhip effect when they are provided with additional information. (*AFFIRMED, p=0.044*)

Hypothesis H2b proposes a difference in the performance of System 1 participants when they are provided with additional SC-wide demand information and when they are not. However, a difference has been detected against the expectation and the hypothesis was rejected.

H2b: Being provided with additional information does not impact the demand tracking performance of managers with a cognitive propensity for their intuitive thinking system (System 1). (*REJECTED, p=0.023*)

5.5.2 Time Pressure Hypotheses

Similar steps were also followed for the second set of hypotheses. Visually, the most significanct difference was seen in the intuitive participants' performances under time pressure (Figure 5.13).

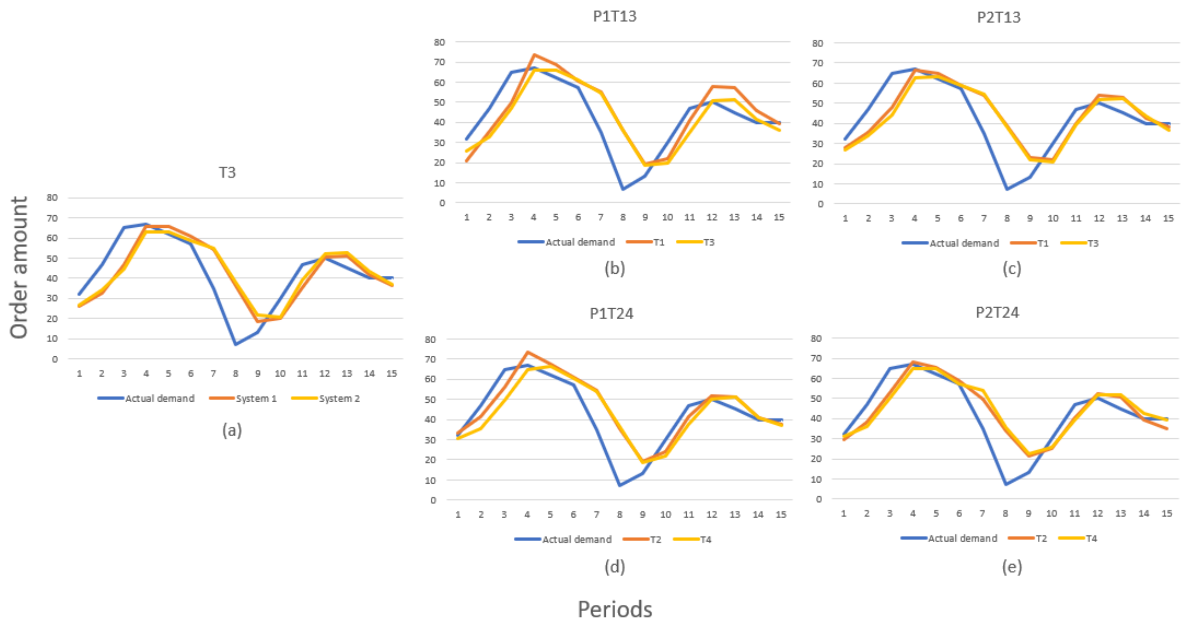


Figure 5.13 Average order comparison for time pressure treatment hypotheses
(a) H3a, (b) H3b, (c) H3c, (d) H3d, (e) H3e

In line with the graphs, statistical findings could not show any statistically significant differences between the compared groups among the hypotheses searching for the inter-group differences, namely H3a, H3b, H3c and H3d (Table 5.4). To begin with, H3a provides the lowest statistical confidence value *(p=0.492, 0.430*) and it has been rejected as the null hypothesis of indifference could not be refuted.

H3a: Managers with a cognitive propensity for their intuitive thinking system (System 1) may perform better than managers who have a propensity for using their rational thinking system (System 2) in demand tracking, when there is time pressure in the decision-making environment. (*REJECTED, p=0.492, 0.430*)

Although the other three hypotheses (H3b, H3c and H3e) were also refuted, their statistical confidence values were much higher (respectively *p=0.115, 0.097 and 0.095)* and closer to the thresholds (p=0.05). Therefore, it would be helpful to examine what they could have meant, especially considering their potential impacts on the further development of the topic. Out of these three hypotheses, H3b was looking for a performance improvement of participants with a cognitive propensity to System 1, whereas the other two -H3c and H3e- aimed to find a performance reduction in participants with a cognitive propensity to System 2 when they are faced with time pressure. Considering the statistical confidence values' proximity to the thresholds, it could have been inferred that participants with a propensity to the opposite thinking systems were impacted reversely by the time pressure. While the intuitive participant group (System 1) has the potential to improve their performance, the rational participant group (System 2) may potentially be affected negatively under time limitations.

H3b: Time pressure improves the demand tracking performance of managers with a cognitive propensity for their intuitive thinking system (System 1). (*REJECTED, p=0.115*)

H3c: Time pressure aggravates the demand tracking performance of managers with a cognitive propensity for their rational thinking system (System 2). (*REJECTED, p=0.097*)

H3e: The effect of additional information presence on System 2 managers' demand tracking performance (managers with a cognitive propensity for their rational thinking system) is aggravated by the time pressure in the decision-making environment. (*REJECTED, p=0.095*)

Table 5.4 Mann-Whitney U-Test results for time pressure treatment hypotheses

Hypotheses	Measurement sample and comparison group description	Total N	1st N	2nd N	1st Group Mean	2nd Group Mean	1st Group Mean Rank	2nd Group Mean Rank	Asymptotic Sig. (1-sided test)
H3a	T3.1V2Cun across CT3.1V2	113	61	52	165.1	163.7	57.58	56.09	0.392
	T3.1V2cdecov across CT3.1V2	112	61	51	75.31	72.63	57	55.9	0.430
H3b	P1.1V3Cun across CP1.1V3	95	34	61	160.7	165.1	48.4	47.78	0.459
	P1.1V3cdecov across CP1.1V3	96	34	62	81.59	77.42	53.1	**45.98**	0.115
H3c	P2.1V3Cun across CP2.1V3	110	59	51	150.9	160.3	**51.83**	59.75	0.097
	P2.1V3cdecov across CP2.1V3	110	59	51	72.95	72.63	55.8	55.16	0.458
H3d	P1.2V4Cun across CP1.2V4	107	46	61	142.8	149.1	51.78	55.67	0.261
	P1.2V4cdecov across CP1.2V4	107	46	61	69.2	68.33	54.68	53.48	0.422
H3e	P2.2V4Cun across CP2.2V4	96	45	51	133.9	150.2	**44.52**	52.01	0.095
	P2.2V4cdecov across CP2.2V4	98	47	51	71.32	74.18	47.15	51.67	0.216

* Affirmed with 95% confidence, Family-wise error rate: 0.23

In contrast to the previous hypotheses, H3d proposes an indifference in System 1 managers' high-information-provided performance when they are tested under time pressure. The findings associated with this hypothesis have not provided any statistically significant difference between the compared groups. While the lack of proof of a difference does not prove the existence of similarity, it has been considered an interesting finding not to show any performance change, although it is seen in the System 2 managers. Therefore, this hypothesis has also been included as affirmation in the boundaries of this study, although it is subject to further research.

H3d: The effect of additional information presence on System 1 managers' demand tracking performance (managers with a cognitive propensity for their intuitive thinking system) is not impacted by the time pressure in the decision-making environment. (*AFFIRMED, p=0.261, 0.422*)

5.5.3 Problem Complexity Hypothesises

In the visual analysis of the last set of hypotheses, comparison of the average orders per period does not present much difference except for some substantial gaps seen in the performance comparison of System 1 managers in low and high problem complexity treatments (Figure 5.14).

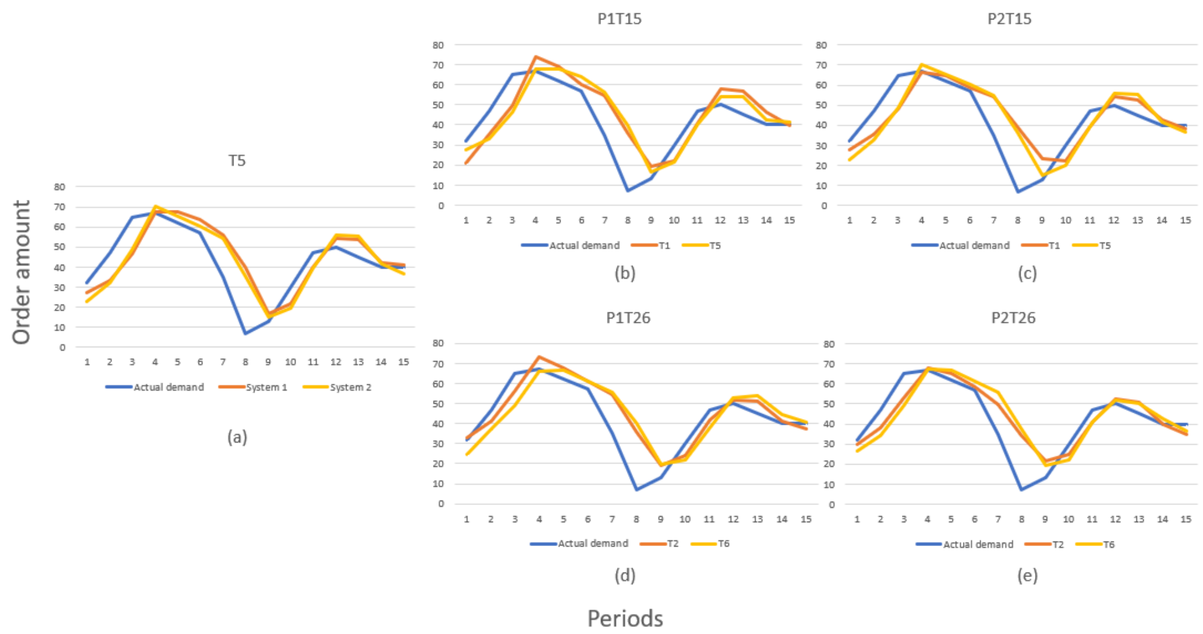


Figure 5.14 Average order comparison for problem complexity treatment hypotheses
(a) H4a, (b) H4b, (c) H4c, (d) H4d, (e) H4e

However, when the intergroup differences were checked statistically, a number of differences were found with high statistical significance (H4a and H4e). Starting with the first hypothesis (H4a), it was observed that System 2 managers continued to outperform System 1 managers in the higher complexity. Moreover, the performance gap between two managerial groups was increased.

> *H4a:* Managers with a cognitive propensity for their rational thinking system (System 2) perform better than managers who have a propensity for their intuitive thinking system (System 1) in demand tracking, when the nature of the problem is more complex. (*AFFIRMED, p=0.012*)

A statistically significant difference was also found for the performance difference of managers with low and high problem complexity, when they are also provided with higher amount of information (*P2.2V6)*. In the first hypotheses sets, it was observed that the performance of System 2 managers was increased by providing them with additional information. However, the positive impact of that information reduces when the complexity of the problem increases.

> *H4e:* The effect of additional information presence on System 2 managers' demand tracking performance (managers with a cognitive propensity for their rational thinking system) is decreased by the increase in problem complexity. (*AFFIRMED, p=0.045*)

The remaining hypotheses in this environmental condition were statistically rejected, however, it could be helpful to highlight the H4b which asserts that System 1 managers perform worse when they are faced with higher complexity. Even though its statistical confidence value (p=0.062) is slightly lower than the threshold (p=0.05), it may be an indicator for further studies on the topic.

> *H4b*: Problem complexity aggravates the demand tracking performance of managers with a cognitive propensity for their intuitive thinking system (System 1). (*REJECTED, p=0.062*)

The other rejected hypothesis is the assertion related to System 2 participants' performance increase with higher problem complexity. It was expected that System 2 managers would have performed better when they faced a more complex problem. However, this positive way difference has not been observed between the comparison groups.

H4c: Problem complexity improves the demand tracking performance of managers with a cognitive propensity for their rational thinking system (System 2). (*REJECTED, p=0.281, 0.366*)

In contrast to the previous hypotheses, H4d proposes an indifference in System 1 managers' high-information-provided performance when they face higher problem complexity. However, it was found that problem complexity aggravates participants' information-enabled performance as well in both measurement constructs. Therefore, this hypothesis has been rejected.

H4d: The effect of additional information presence on System 1 managers' demand tracking performance (managers with a cognitive propensity for their intuitive thinking system) is not impacted by the increase in problem complexity. (*REJECTED, p=0.006, 0.01*)

Table 5.5 Mann-Whitney U-Test results for problem complexity treatment hypotheses

Hypotheses	Measurement sample and comparison group description	Total N	1st N	2nd N	1st Group Mean	2nd Group Mean	1st Group Mean Rank	2nd Group Mean Rank	Asymptotic Sig. (1-sided test)
H4a*	T5.1V2Cun across CT5.1V2	92	45	47	187.9	162.5	52.96	**40.32**	0.012
	T5.1V2cdecov across CT5.1V2	90	43	47	87.86	77.72	50	**41.38**	0.059
H4b	P1.1V5Cun across CP1.1V5	79	34	45	160.7	187.9	**35.43**	43.46	0.062
	P1.1V5cdecov across CP1.1V5	77	34	43	81.59	87.86	38.16	39.66	0.385
H4c	P2.1V5Cun across CP2.1V5	105	59	46	150.9	158.3	51.47	54.96	0.281
	P2.1V5cdecov across CP2.1V5	106	59	47	72.95	77.72	52.58	54.65	0.366
H4d	P1.2V6Cun across CP1.2V6	91	46	45	142.8	183.1	**39.1**	53.06	0.006
	P1.2V6cdecov across CP1.2V6	90	46	44	69.2	87.09	**39.25**	52.03	0.01
H4e*	P2.2V6Cun across CP2.2V6	86	46	40	146.8	159.8	**39.25**	48.39	0.045
	P2.2V6cdecov across CP2.2V6	87	47	40	71.32	75.1	47.91	46.45	0.202

* Affirmed with 95% confidence, Family-wise error rate: 0.23

5.6 Chapter Summary

This study recruited 670 participants in total, which produced 623 valid cases with 47 rejections. The gender distribution was 42% male – 58% female. Participants completed the experiment in an average of 11.64 minutes.

Against the concerns related to the CRT, any violation related to gender distribution and previous exposure to the questions was not experienced. Despite the higher number of male participants, this did not influence the equal distribution of cognitive propensity and nearly equal numbers for each propensity group was reached. On the other hand, a negative impact from previous exposure was observed; however, since it had very low explanation power, it was not regarded as creating any interference for the study.

Another concern was related to any potential learning behaviour that would occur in either participant group, and consequently would positively skew its performance. Analysis, however, has shown a completely opposite skew. A reduction in System 2 participants' performance throughout the game was observed in four out of six treatments. Since it was not a performance enhancing learning, the reason for this temporal difference was considered to be related to the unique characteristics of the rational thinking system as it requires a higher cognitive effort for the solution of a problem, and the focused cognitive power may be consumed earlier than the task ends. Since intuitive thinking system shows a lower cognitive effort, the concentration might have lasted longer. However, it has been noted as a promising research area for future studies.

After completing the required controls, the hypotheses were tested with a One-sided Mann-Whitney U-Test. Out of the 13 hypotheses, five (H1, H2a, H4a, H4e) were affirmed with a 95% confidence level; one could not identify the difference (H3d) and the remaining eight were rejected. Figure 5.15 depicts the proposed conceptual model with the hypothesis results.

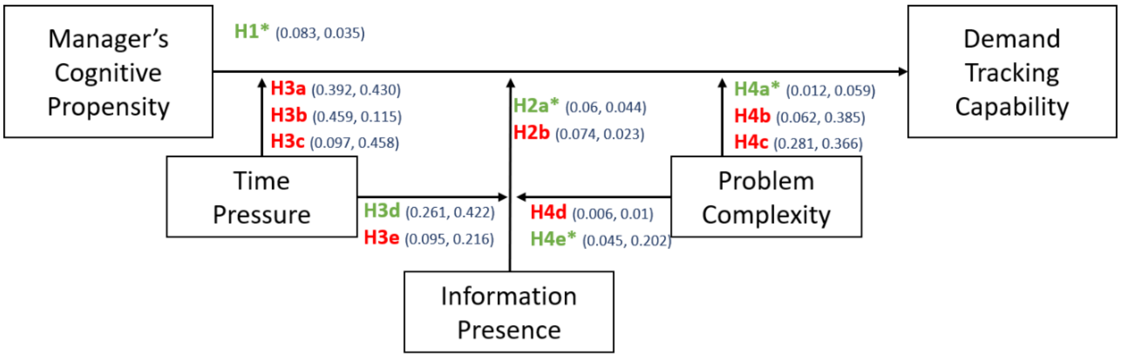


Figure 5.15 Hypothesis results for proposed conceptual model

6. Discussions

"Truth is found neither in the thesis nor the antithesis, but in an emergent synthesis which reconciles the two."

GEORG WILHELM FRIEDRICH HEGEL

6.1 Chapter Introduction

The chapters so far have reviewed the literature to advance the knowledge about the core concepts of this study; determined interrelationships and proposed related hypotheses; and finally tested the hypotheses utilising the designed experiment. This discussions chapter will present the inferences from the analytical findings in conjunction with their relationship to the extant literature.

The structure of the chapter follows the categorisation of the hypotheses. Hence, Section 6.2 will explain the findings related to the affirmation of the baseline hypotheses. The findings for this section are of high importance as they are the main connection point of this study to the extant literature.

Section 6.3 will then present the findings of the situational hypotheses in three categories. The findings of these hypotheses will be synthesised with the related literature.

Finally, the chapter will be summarised in Section 6.4.

6.2 Bullwhip Effect Mitigation through the Lens of Dual-Process Theory

The main motivation of this study was to identify the role of managers' cognitive propensities in their demand tracking capability as one of the core constructs of BWE creation as well as its mitigation. The literature provided a number of studies in the area, which were mainly composed of the serial works of Dr Brent Moritz. Starting with Moritz et al. (2013), they first measured the individual differences in NP performance. Then in Narayanan and Moritz (2015) they have enhanced the study by utilising a more complex inventory management simulation, the BDG. In Moritz et al. (2020), they investigated the various scenarios related to the distribution of managers with different cognitive propensity groups across the SC, and measured the varying impacts on the BWE occurrence. In all these invaluable studies, it was observed that participants who have a propensity for preferably using their rational thinking systems, System 2 or rational managers in the short definition of this study, perform better than System 1 or intuitive managers in mitigating the BWE throughout the SC. Contributing to the development of the BOM field, this study has focused on another aspect of the topic by investigating the potential performance variations in association with the impacts of decision-making environments. However, to be able to observe these expected variations, a benchmark was needed. Consequently, that main finding of previous studies was included in this study as hypothesis H1.

H1 was expected to affirm the findings of those studies. In other words, System 2 managers were considered to handle the BWE by showing a higher capability to follow the demand. As expected, the results of this experiment have confirmed these results. System 2 managers followed the changes in the demand trend better than System 1 managers. Consequently, the role of managerial cognitive differences in bullwhip mitigation strategies has been consolidated with this study. On the other hand, this affirmation has also ensured the validity of the customised CRT and designed OMG as they provided coherence with previously reached results multiple times.

H1 (Accepted): Managers with a cognitive propensity towards their rational thinking system (System 2) perform better than managers who have a propensity to use their intuitive thinking system (System 1) in tracking the customer demand to mitigate the bullwhip effect.

6.3 Environmental Conditions for Bullwhip and Cognition

The concepts of cognition and cognitive thinking systems were elaborately discussed in the literature review chapter. A point that was specifically highlighted was the common bias against the capabilities of the intuitive thinking system in comparison to those of the rational thinking system (*see* Subsection 2.3.2 [b-III], *p. 33*). However, together with the further research in the field, the relative strengths and weaknesses of both thinking systems were illuminated.

Besides these advancements, the concept has also progressed in scope. In other words, differences between the cognitive system utilisation have been researched in many contexts, one of which is the various SC decisions. As previously depicted, researchers have investigated various situations in which intuition can produce a better performance. Some examples are supplier selection and relationships (Katok, 2011b) and risk management (Ancarani et al., 2013).

Together with the studies preceding the first hypothesis of this study, order and inventory management has also become one of the SC aspects in which cognitive thinking styles bring a difference. However, more research with different perspectives is required to enlighten the dynamics of cognitive thinking systems in varied situations. Moving from this gap, observing the potential performance variations associated with three environmental conditions was targeted. The studies of Campitelli and Labollita (2010) and Ayal et al. (2015) had addressed the essentiality of external factors for making a decision. In the BSCM domain, they were also seen as critical for decision making (Schorsch et al., 2017). In line with these studies, the impacts of three environmental conditions on the order management performances of SC managers with different cognitive propensities were analysed. The following sections will provide a detailed discussion on each related hypothesis.

6.3.1 Information Presence in Bullwhip Effect and Behavioural Studies

The first set of situational hypotheses was built on the utilisation of additional demand information of further upstream SC partners. Section 2.4.1 (*p. 37*) reviewed the essentials of information utilisation for behavioural decision making as well as SC decisions. Following that, Subsection 3.2.2 [a] (*p. 42*) asserted the following hypotheses:

H2a (Accepted): Managers with a cognitive propensity for their rational thinking system (System 2) perform better in tracking the customer demand to mitigate the bullwhip effect when they are provided with additional information.

H2b (Rejected): H2b: Being provided with additional information does not impact the demand tracking performance of managers with a cognitive propensity for their intuitive thinking system (System 1).

In sum, the expectation was to receive opposite results from two groups of participants with different cognitive propensities, regarding the utilisation of additional information. To be able to test these assertions, participants in both cognitive propensity groups were given two information scenarios (*see* Subsection 4.3.2 [c-I], *p. 65*). In the first setting, participants were provided with the demand information of their direct customers (Low Information). In the second setting, on the other hand, they could have reached the demand information of further SC partners (High Information).

Considering the findings gathered, we have ended up with opposing results for the asserted hypotheses. Starting with H2a, it was observed that participants (representing the SC managers) with a cognitive propensity for preferring the rational thinking system in their decisions – System 2 managers – performed better in following the changes in demand trend when they were provided with further information from the downstream SC. Consequently, it can be inferred that further SC demand information can be utilised in demand tracking if the SC managers are of System 2 propensity. This finding is in line with the mainstream school of thought in the BWE literature (*see* Section 2.2). Many works have defined information access as a major bullwhip mitigation strategy by supporting the forecast accuracy as well as other order management attributes, such as decreasing safety stock and overall SC cost (Cui et al., 2015; Narayanan and Moritz, 2015). Lack of information, on the other hand, was accepted as a significant cause for the bullwhip (van Riel et al., 2003). Besides the SC aspect, information has been viewed as a significant element in the decision-making literature. Especially when considering its main characteristics, the rational thinking system utilises as much information as possible to critically analyse the problem and then produces the decision (van Riel et al., 2003). In aligning with this accumulation of both literature clusters, this study's findings related to H2a have shown that participants with a cognitive propensity for using their rational thinking system have utilised the additional information when it was provided and performed better in tracking demand and hence mitigating the bullwhip.

Besides these mainstream opinions about the role of information in bullwhip mitigation, the literature also harbours some research aspects that highlight the occasions where information cannot boost the inventory management activities. Among those situations, stationary demand situations (Steckel et al., 2004), deficient information attributes, such as incomplete or inaccurate information (Bendoly and Swink, 2007), or undependability of the information source (Croson et al., 2014), can be counted. The main inference from these examples has been that the presence of information is not a guarantee of a performance increase. In line with this, information utilisation of the intuitive thinking system has also been considered as one of the situations where the more information provision may not mean a higher performance. The reason for this assumption has been related to the inherent characteristics of the intuitive thinking system. While the rational thinking system elaborately utilises as much information as available and this leads to expect better decisions when accessing further information, it does not apply to the decisions given by the intuitive thinking system. Intuitive decisions are produced instantly, without spending much effort on reviewing newly acquired information. Instead, intuitive decisions may use inherently extant information which were gathered via previous experiences and then stored in easily accessible memory. Moving from this consideration, H2b was proposed which asserts an expectation that System 1 participants would not be different when they are provided with low or high information.

However, the associated findings indicated an opposite situation, and the hypothesis was rejected. Just as in the case of System 2 managers, System 1 managers' performance increased with the provision of additional information. Therefore, the initial outcome would be straightforward so that additional information can also be used for better decision making by the managers who have a cognitive propensity for using their intuitive thinking system more in their decisions. The reason for the unexpected information utilisation can be, first, related to the collaborative mechanism of cognitive thinking systems (*see* Subsection 2.3.2 [b-III], *p. 33*). It has been highlighted at many points that our decisions are the product of both thinking systems, although there may be a preference for using either. In line with this, more information might have triggered the rational thinking system even though managers are inherently more inclined to use the other. Second reason, on the other side, can be associated with the information attributes. This study was constructed as a single information type to preserve participant focus and comprehension throughout the experiment. Therefore, it might have been easy to process the additional information for intuitive managers. However, as already explained, the quality attributes of information, such as accuracy, timeliness, completeness and structure, can be factors affecting information usage. Gavirneni and Isen (2010), for example, have shown that the participants in their study did not seek additional information to enhance their decisions. In line with this research stream, variations in information attributes can be a factor to determine the information utilisation performance of either thinking system.

6.3.2 Time Pressure in Bullwhip Effect and Behavioural Studies

The second set of situational hypotheses is targeted at explaining the potential performance variations in demand tracking when managers make their decisions under time pressure. The main aspects related to time pressure and constraint have been explained in Section 2.4.2 (*p. 38*). As well as the role of time pressure in the SC decisions and behavioural decision-making, Subsection 3.2.2 [b] (*p. 43*) analysed the potential relationships among these elements and consequently produced these initial hypotheses to be tested:

> *H3a (Rejected):* Managers with a cognitive propensity for their intuitive thinking system (System 1) may perform better than managers who have a propensity for using their rational thinking system (System 2) in demand tracking, when there is time pressure in the decision-making environment.
>
> *H3b (Rejected):* Time pressure improves the demand tracking performance of managers with a cognitive propensity for their intuitive thinking system (System 1).
>
> *H3c (Rejected):* Time pressure aggravates the demand tracking performance of managers with a cognitive propensity for their rational thinking system (System 2).

The overall expectation was to observe a time pressure-related difference in the demand tracking performance of managers. The literature has produced studies that investigate the role of time pressure in various business aspects, such as consumer behaviour, negotiations, retailing, and accounting (Thomas et al., 2014), as well as the SC decisions, such as supplier evaluation and relationship management (Carter et al., 2017). Furthermore, that impact was also considered to be seen in the BOM (Thomas et al., 2014).

Time pressure was also regarded as critical from the cognitive decision-making perspective. That importance depended on the difference between the thinking systems' characteristics (*see* Subsection 2.3.2 [b], *p. 30*). As explained, the rational thinking system puts a high amount of focus and effort into the decision-making process. Its processing style is the serial processing done by solving the problems separately in order. Intuitive thinking system, on the other hand, has shown opposite features. It is miserly in effort and focus. Decisions are taken subconsciously without much concentration and effort. Adopting parallel processing can solve multiple problems simultaneously. These features, expectedly, result in a process time difference as well. While the intuitive thinking system produces instant decisions, the rational thinking system requires a considerable amount of time to finalise the decision. Moving on from all these time-related differences between the thinking systems, their demand tracking performance under a limited time was questioned in this study. Consequently, via H3a, it was researched whether managers with a cognitive propensity for utilising their intuitive thinking systems perform better then managers using the rational thinking system, when the decision is required to be made under time pressure. Against expectations, the findings have demonstrated that System 2 managers perform still better than System 1 managers in tracking demand. Nevertheless, a reduction in the performance gap between the two groups has still been observed, although it was not high enough to produce statistically significant inferences.

Subsequent hypotheses H3b and H3c also could not provide a statistically significant difference between the compared groups. H3b was proposing a higher performance for System 1 managers under time pressure, while H3c mentioned the opposite for System 2 managers, i.e. a lower performance when they are faced with time pressure. However, it was also observed that the statistical confidence values for these two hypotheses were close to the threshold value of 0.05. Considering that this study is 1) the first in this research direction, 2) utilises a single construct to analyse BWE-related performance 3) and employs a considerably simple experiment to measure the dependent and moderating variables, it would bring promising insights for future studies that may examine the dependent and moderating variables in further detail. Therefore, the direction of change in the performance of compared groups in these hypotheses may provide the following inferences even though they will require further research to be statistically affirmed.

To recap, a bias against the intuitive thinking system's decision-making performance, in comparison to the rational thinking system, has been observed in the literature. However, this study has shown that in a specific environment – time pressure – which is more suitable for the working mechanism of the intuitive thinking system, the demand tracking performance of System 1 managers may be increased, while that of System 2 managers may decrease. Even though System 2 managers may still track the demand better than System 1 managers, the potential in the opposite performance changes of thinking systems (H3b and H3c) have shown that it is a promising research area to investigate the performance differences concerning the environmental context.

In addition, manager groups' information utilisation was also investigated under time pressure conditions via these following hypotheses:

> *H3d (Accepted):* The effect of additional information presence on System 1 managers' demand tracking performance (managers with a cognitive propensity for their intuitive thinking system) is not impacted by the time pressure in the decision-making environment.
>
> *H3e (Rejected):* The effect of additional information presence on System 2 managers' demand tracking performance (managers with a cognitive propensity for their rational thinking system) is aggravated by the time pressure in the decision-making environment.

The reason for these additional hypotheses was to observe the effectivity of an operational condition under the other environmental conditions. Regarding the information utilisation of the rational thinking system, it was considered that System 2 managers will require more time to make their decision, when they are provided with additional information. However, when they cannot find the required time, it was considered that their information-empowered performance would reduce. These expectations were in parallel with the assertions of Thomas et al. (2010), where they connect time pressure with the reduction in information flow and processing. Likewise, van Riel et al. (2003) had also argued for the negative impact of time pressure on information utilisation. The related result of this study (H3e) unfortunately could not provide a statistically significant verification for these earlier research findings. However, the statistical confidence value (0.095) for H3e was also close to the threshold level (0.05) and it may be accepted as a promising outcome as well as it requires consolidation from the further studies that will be designed specifically for these associated treatments.

Regarding the information utilisation of System 1 managers, on the other hand, it was claimed that the time pressure would not impact at all, because, as explained in the previous section, additional information was not considered as a factor to make the performance of System 1 managers differ in the first instance. Therefore, the same presumption was made for this situation as well. Although information has unexpectedly shown supportive impact for the bullwhip mitigation performance of System 1 managers, this impact does not change when there is time pressure in the decision-making environment. This performance indifference can tell that System 1 managers continued to make their decisions in line with the main characteristics of the intuitive thinking system. Although, they tapped into the benefits of additional information, they did not show high amounts of time and effort in making the decisions. As a consequence, they could have sustained their information-empowered demand tracking performance under time pressure, while the supportive power of information was not observed for System 2 managers.

6.3.3 Problem Complexity in Bullwhip Effect and Behavioural Studies

The last hypothesis group has pursued describing potential performance variations in demand tracking when managers face with higher problem complexity. Section 2.4.3 (*p. 39*) provided a background to the problem complexity, especially highlighting its importance for the SC environment. Comprising multiple stakeholders, any SC decision can possess complexity. Inventory and demand management is also one of these decisions. While these tasks have various uncertainties *per se*, as the number of stakeholders and associated flows increase, their management becomes harder. The complexities in the SC were exemplified with other occasions, such as unexpected disruptions and delays throughout the SC (Steckel et al., 2004), and changes coming with the disruptive technologies and evolving consumption trends of consumers (Kaufman et al., 2017). In line with this, Subsection 3.2.2 [c] (*p. 45*) analysed the potential outcomes in relation to the cognitive thinking systems' coping capabilities with problem complexity. Consequently, the following hypotheses were proposed initially:

> *H4a (Affirmed):* Managers with a cognitive propensity for their rational thinking system (System 2) perform better than managers who have a propensity for their intuitive thinking system (System 1) in demand tracking, when the nature of the problem is more complex.
>
> *H4b (Rejected):* Problem complexity aggravates the demand tracking performance of managers with a cognitive propensity for their intuitive thinking system (System 1).
>
> *H4c (Rejected):* Problem complexity improves the demand tracking performance of managers with a cognitive propensity for their rational thinking system (System 2).

The complexity was constructed as an additional sales channel towards the downstream SC (*see* Subsection 4.3.2 [c-iii], *p. 66*), by proportionally dividing the total demand into these channels. Therefore, despite being provided with the same demand in total, it became harder to analyse the total demand and place the required order for participants who had taken the problem complexity treatments.

The main motivations behind these hypotheses were similar to those associated with the time pressure, despite being in the opposite direction. While time pressure was included in the study to uncover the potential capabilities of the intuitive thinking system, increased problem complexity has been considered as a suitable situation to realise the further capabilities of the rational thinking system. Therefore, initially it was considered that the higher performance of System 2 managers in comparison to that of System 1 managers would be sustained when the problem is more complex (H4a). This assumption was affirmed by the findings of this study.

Besides the initial hypothesis, the study has also investigated the impact of problem complexity on individual thinking systems individually (H4b and H4c). The suitability of the rational thinking system to solve more complex problems has already been highlighted so far, in conjunction with its features to assert higher focus and effort. Supporting that, van Riel et al. (2003) had also proposed a positive relationship between the utilisation of the rational thinking system and increased complexity. However, the findings of this study have shown that problem complexity does not contribute to the demand tracking performance of System 2 managers.

Regarding the performance of System 1 managers with more complex problems, van Riel et al. (2003) proposed a similar hypothesis. Their assumption was that the intuitive thinking system could have been successful in simplifying the problem especially by using managers' prior experiences. However, this study's related hypothesis (H4b) was dissimilar to these statements for two reasons. First, this study utilised the experiment as the research method. It isolated the measurement of variables from all other factors to observe the intervariable causal effects. Therefore, the specific interaction between the experience and problem complexity performance was not within the boundaries of this study. Second, although 'bringing the issue to the right ballpark' is among the features of the intuitive thinking system, the literature does not provide any opinion arguing that this feature works better with increased complexity. On the contrary, the higher complexity may, unsurprisingly, obstruct simplifying the problem. In return, this may result in a more suboptimal demand tracking performance. The findings related to hypothesis H4b have unexpectedly not provided a statistically significant difference, the statistical confidence value (0.0.62) was highly close to the confidence threshold of 0.05. In other words, System 1 managers' demand tracking performance expectedly decreased when the problem got more complex, but it requires further focused studies to consolidate these findings statistically.

Summing these three hypotheses together, it can be inferred that while it is apparent that System 2 managers outperform System 1 managers under higher problem complexity (H4a), their performance to deal with the complexity also shows a clear result. System 1 managers' demand tracking performance is inclined to decrease in a more complex SC. On the other hand, System 2 managers are not affected in the same way as System 1 managers and they can retain their demand tracking performance.

Following the same structure as for the time pressure treatments, lastly, the between information presence and problem complexity was researched via the following hypotheses:

> *H4d (Rejected):* The effect of additional information presence on System 1 managers' demand tracking performance (managers with a cognitive propensity for their intuitive thinking system) is not impacted by the increase in problem complexity.
>
> *H4e (Affirmed):* The effect of additional information presence on System 2 managers' demand tracking performance (managers with a cognitive propensity for their rational thinking system) is decreased by the increase in problem complexity.

Subsection 3.2.2 [c] (*p. 45*) highlighted the opposing characteristics of information and complexity through the lens of uncertainty. While information is a tool to reduce uncertainty for better decisions, complexity works the opposite way and increases the amount of uncertainty in the environment. Therefore, it was asserted that further problem complexity may harm the information-enabled clarity of the environment. Remembering that additional information was proposed as an enabler of rational thinking system, it was also proposed that the complexity could have reduced the impact of information in demand tracking performance for System 2 managers (H4e) and the findings have affirmed this proposition.

System 1 managers' performance, on the other hand, was considered unrelated to the information as the intuitive thinking system inherently utilises the information coming from experience. Therefore, it was proposed that any change in the demand tracking performance of System 1 managers would not be observed when they are faced with a more complex problem. However, as explained in Section 6.2, H2b has been refuted and it was observed that information contributes to the performance of System 1 managers. When the problem is more complex, however, information has not shown a similar contribution. Hence, System 1 managers' performance was decreased.

To sum, when the problem becomes more complex, System 2 managers continue to outperform System 1 managers in mitigating the bullwhip. Moreover, they can sustain their initial performance, while System 1 managers are influenced negatively by increased problem complexity. The positive impact of additional information, however, is blocked by the complexity of the problem for both manager groups.

6.4 Chapter Summary

This chapter has synthesised the findings of this experimental study with the extant knowledge of the related literature. The arguments have been demonstrated per the hypotheses in line with the structure of previous chapters.

Starting with the baseline hypothesis, results have shown that System 2 managers perform better in tracking the changes in demand trends and therefore are more capable of mitigating the bullwhip in SCs. This result had been already reached by other studies in the field. Affirmation of the extant knowledge in the baseline hypothesis has proved that this study has been built on the same validity characteristics as those studies. Moreover, as the word hypothesis implies, this finding was utilised as a benchmark to identify the differences with other tested conditions. Hence, the inferences made for the remaining hypotheses have also possessed the same level of validity.

Following the baseline hypothesis, the remaining situational hypotheses have researched the performance variations in relation to three environmental conditions: information presence, time pressure and problem complexity.

The literature has always regarded information as an operational solution for the mitigation of the BWE. It is accepted as the antidote of uncertainty that has always been among the biggest causes of the bullwhip. However, even access to information could not be the only and ultimate solution to the bullwhip, as there was the other side of the coin, namely behavioural reasons. As the closest, the baseline hypothesis has shown how cognitive differences impact on managers' bullwhip mitigation performance. As a further step, this study has also observed the relationship of managers' cognitive propensity with their information utilisation capability in tracking demand changes. It was observed that information boosts the performance of managers from both cognitive propensity groups. This was expected for the System 2 managers as they are more prone to analysing the situation in more detail and thoroughly. However, System 1 managers were expected not to efficiently utilise the provided additional information as the intuitive thinking system preferably uses the information coming from prior experience. This result can be explained with the collaborative utilisation of thinking systems. Even though individuals may be inclined to use either thinking system in decision making, it does not mean that they are purely intuitive or rational. Human beings possess both thinking systems and utilise both when required. Also in this experiment, System 1 managers have shown that they also tap into the benefits of additional information, just as System 2 managers do.

Secondly, the performance differences associated with the time pressure in the decision-making environment were investigated. The intuitive thinking system intrinsically makes instantaneous decisions, whereas the rational thinking system requires more effort and therefore time when making a decision. Moving from this difference, it was considered that the baseline inference associated with the outperformance of System 2 managers over their System 1 counterparts may not be valid when the time to finalise the decision is hardly sufficient and also limited. Therefore, propositions argued a positive way difference for the performance of System 1 managers in tracking demand, and vice versa for System 2 managers. Results, indeed, have observed an improvement in System 1 managers' demand tracking capability when they were under time pressure. However, this improvement was not sufficient to be proved with the required statistical significance level. This lack of statistical significance was also valid for the hypotheses associated with System 2 managers. Although a reduction was observed in their demand tracking performance under time pressure, System 2 managers still performed better than System 1 managers. In sum, despite the maintenance of System 2 managers' superiority, the performance of both manager groups converged. This has been a promising finding for future research to discover the outperformance situations related to the intuitive thinking system.

Thirdly, problem complexity was selected as the last environmental condition that may influence the bullwhip mitigation performance of managers with different cognitive propensities. Managing the complexity has been detected as one of the differences between the thinking systems. While the rational thinking system produces better decisions, by spending more focus, effort and time on the decision-making process, the intuitive thinking system makes quick and automated decisions by simplifying the situation. However, this simplification may cause missing critical points in the problem and result in higher deviation from the optimal decision. Therefore, it was proposed that System 2 managers continue to outperform System 1 managers, since both cognitive groups will be impacted by the complexity in opposite and diverging directions. As expected, it was observed in the results that System 2 managers continued to track the demand changes better than System 1 managers. Again expectedly, it was observed that System 1 managers were badly influenced by the high problem complexity and their performance dropped dramatically. However, this performance reduction was not enough to be proved with statistical significance and requires further research to be supported. System 2 managers, on the other hand, showed performance reduction in contrast to expectation; however, that reduction was not statistically significant. Therefore, it can be concluded that although problem complexity aggravates the overall demand tracking performance, System 2 managers do better in overcoming this difficulty in comparison to System 1 managers.

Lastly, the impact of information on demand tracking performance was also researched under the other environmental conditions for both manager groups. Overall, a substantial difference was detected between the impacts of environmental conditions on participants' utilisation of additional information. Firstly, it was observed that the performance of participants from either cognitive thinking group was not changed even if they are provided with additional demand information. On the other hand, problem complexity brings a substantial diminishing influence for both manager groups in facilitating the support of additional information. This finding consequently means that unless the right conditions are met, access to additional demand information from the downstream SC may not produce the targeted bullwhip mitigation results, even without consideration of the cognitive propensity of the managers.

7. Conclusion

"The press, the machine, the railway, the telegraph are premises whose thousand-year conclusion no one has yet dared to draw."

FRIEDRICH NIETZSCHE

7.1 Chapter Introduction

This chapter will conclude the thesis in four sections.

First, achievement of the research questions will be addressed in Section 7.2.

Sections 7.3 and 7.4 will present the theoretical contributions and practical implications of this thesis.

The limitations of this study will be mentioned in Section 7.5 together with the future research directions.

Finally, the chapter will be summarised in Section 7.6.

7.2 Revisiting the Research Questions

The best way to comprehend the successful completion of a project is done via the control of the questions asked at the beginning. Therefore, it is helpful to remember those research questions and to check if they have been answered:

RQ: How do environmental conditions impact on the bullwhip mitigation performance of managers with different cognitive propensities via their demand tracking capabilities?

- How does the cognitive propensity of supply chain managers impact on their demand tracking capability?
- How does additional information presence differ from the performance of supply chain managers with different cognitive propensities?
- How does time pressure in the decision-making environment differ from the performance of supply chain managers with different cognitive propensities?
- How does the increased complexity of the problem differ from the performance of supply chain managers with different cognitive propensities?

To provide a holistic answer to the overarching research question, sub questions were subsequently answered. Initially, related literatures were thoroughly reviewed. Factors leading to the BWE were identified together with the proposed solutions to date. Next, the working mechanism of human cognition was explored via the dual-process theory. Investigating the similarities and differences between the cognitive thinking systems and potential decision-making environments in the SC, their impacts were hypothesised in various configurations. However, before all configurations, it was checked how managers' bullwhip mitigation performance varies under neutral conditions in relation to their cognitive propensities for either of the two thinking systems. Results showing a difference between managerial groups where System 2 managers outperformed System 1 managers has answered the first sub question.

While this finding affirmed the extant knowledge in the field, the remaining questions sought for the context-dependent performance variations and doing so enhanced the field. Among the contextual situations, the first set of hypotheses was prepared to answer the second sub questions by measuring how additional information presence impacts on bullwhip mitigation performance of both managerial groups. It was observed that information improved the performance of both managerial groups, although it was better utilised by System 2 managers. Moreover, further hypotheses measured the impact of information in conjunction with other environmental conditions. It was seen that the positive impact of additional information presence dropped in other situations, except System 1 managers' performance under time pressure.

The third sub question investigated the performance variations that were caused by time pressure in the environment. Depending on the suitability of the intuitive thinking system's characteristics in deciding in a short time, it was considered that time pressure may increase the performance of System 1 managers, while aggravates that of System 2. However, the related findings could not provide significantly significant proofs for these suggestions. It was observed that System 2 managers continued to outperform System 1 managers for this specific problem, while System 1 managers' performance increased with time pressure, and System 2 managers' performance dropped. However, these performance difference should be tested in further studies that will be specifically designed towards the targeted dependent or moderating variable. In addition, managers from both cognitive thinking groups preserved their information-boosted performance even under time pressure.

The last sub question sought for problem complexity's impact on both managerial groups' bullwhip mitigation performance. The expectation was to observe performance improvement in System 2 managers' performance, but reduction in that of System 1 managers. Even though the findings showed performance reduction for both managerial groups, the amount of reduction was much lower in System 2 managers in comparison to System 1 managers. While these performance reductions could not be proven with statistical significance, further studies may provide additional insight especially for the reduction of System 1 managers' performance when faced with more complex problems.

In conclusion, this study aimed to bring further observation on the cognitive propensity-dependent performance of supply chain managers regarding their variations in different environmental contexts. The sub questions examined three environmental conditions and revealed that supply chain managers' bullwhip mitigation performance depends on these environmental conditions as well as their cognitive propensities.

The answers to these questions have filled the aforementioned gaps by contributing to the literature. However, there are also practical implications of the findings of this study. Following the research balance framework of Stentoft and Rajkumar (2018) (Figure 7.1), the subsequent sections will provide the theoretical contributions and practical implications of this study.

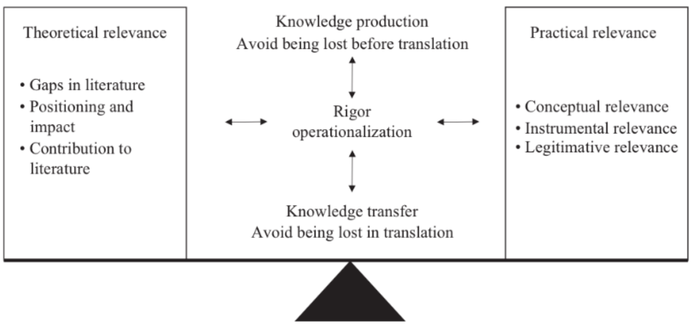


Figure 7.1 Research balance framework (Stentoft and Rajkumar, 2018)

7.3 Theoretical Contributions

This study has embodied a three-fold contribution. The first contribution has been made to the literatures of utilised concepts: BSCM in the wider perspective and specifically, BWE occurrence and mitigation. Secondly, the adopted theory has been extended with the perspective of this study. Lastly, a prepared experiment tool has brought novel aspects for the measurement of this study's variables.

Starting with the first perspective, it was mentioned in detail that behavioural aspects have comprised an important part of the studies that aim to bring explanations for the phenomenon and its potential solutions. However, behavioural studies alone could not bring an ultimate explanation or solution to the problem and hence more research has been required in the area and still is. In line with this, Dr Brent Moritz and his colleagues have expanded the field by investigating the application of dual process theories in the bullwhip phenomenon (*e.g.*, Moritz et al., 2013; 2014; 2020; Narayanan and Moritz, 2015; Ovchinnikov et al., 2015). Adding to these studies Carter et al. (2017) recommend further research regarding the interplay between intuition and rationality dimensions in the SC context, therefore this research has produced further explanations for the role of dual thinking systems in the occurrence and mitigation of the BWE by including the impacts of the decision-making environment. Preceding studies have already put forward that the cognitive propensity of an SC manager for either thinking system has an impact on the management of bullwhip. It was argued that managers with a propensity for the rational thinking system (System 2 managers) perform better than intuitive managers (System 1 managers) in dealing with the bullwhip. This research, however, enhances that argument by demonstrating the context-dependent variations in this comparative performance. While the validity of this argument has also been confirmed within this study for normal environments, it does not remain so when the nature of the decision-making environment differs. Time pressure in the environment may increase the performance of System 1 managers while it may aggravate that of System 2 managers. Problem complexity works vice versa and boosts System 2 managers' performance while decreasing System 1 managers' performance. Additional demand information in the environment, on the other hand, helps both manager groups mitigate the bullwhip; however, that positive impact cannot be sustained when the time pressure or problem complexity increases in the environment. Therefore, this study has additionally shown that access to information is not the final solution as the environment should be suitable for the proper utilisation of provided information.

Secondly, this research has contributed to the adopted theory. Experiments help identify any violations against the rules of theories and then modify those theories accordingly. This has been valid for dual-process theories as well. Although the separation of cognitive thinking systems dates to the middle of the previous century, it has taken time and further research to understand the interaction between those systems. The recent school of thought that looks for further diversification of thinking systems is also a result of this violation-led theory development. This study did not target the uncovering of a flaw. Nevertheless, results have provided some essential insights regarding the gaps in the existing rules of theory. The first insight is related to information utilisation of cognitive thinking systems. Moving from the extant literature, it was initially thought and proposed that additional information was utilised by the System 2 managers, whereas there would not be any utilisation by System 1 managers. This idea was derived from the intrinsic characteristic of the intuitive thinking system that utilises the information coming from prior experience. However, System 1 managers utilised the provided additional information in this study just like System 2 managers. This may be caused by the fact that a propensity for either thinking system does not mean that decisions are purely made by that thinking system. Instead, the interaction between the system maintains that individuals can still utilise the other system for a decision, although they are inclined to use their preferred one.

In this study, the observation regarding the successful information utilisation may also be caused by this interaction. While this study affirms this interaction, it opens a research direction to investigate the information-related variances for intuitive systems. Investigation of potential information attributes such as the type, quantity or presentation of information may further help reveal the potential impacts of dual thinking systems on managers' order management capabilities as well as the other SC tasks.

Besides, this study has extended the application of dual-process theories in the SCM area. Although the aforementioned studies had applied the theory in inventory management decisions, this research provided further explanation by adding environmental variations that can be commonly seen in SCs. Since SCs are composed of complex networks with various decisions, it has been an important contribution to simulate these SC environments, which may apply to other SC activities as well.

Lastly, this study has also contributed to the methodology. The experiment tool utilised was composed of two main sections: Cognitive Reflection Test (CRT) and Order Management Game (OMG). The CRT has continuously gained popularity and been used by researchers extensively; however, it has some drawbacks related to multiple exposure and measurement capability. This study has reviewed 46 questions that had been used in the CRT and went through an elimination process via recommendations of the previous works and a pilot study. As a result, five questions were chosen for the customised CRT of this study. Results have shown that they have tested the cognitive reflection of participants properly and produced nearly equal numbers for the cognitive propensity of groups in this study. Moreover, it was noted that selected questions were seen by only a quarter of the participants on average. Considering that this experiment was conducted in an online platform where multiple exposure is an even a higher concern as participants are coming across with various CRT versions, the selection process of the questions can be seen as successful. This can be utilised and further developed by future researchers.

The OMG was also a customised application which was devised by mixing the settings of two popular inventory management experiments: Newsvendor Problem (NP) and Beer Distribution Game (BDG). One of the reasons for combining them related to the measurement technique used in previous studies. Studies using the BDG have conventionally measured inventory and order management together by using Sterman (1989)'s formulae and statistically predicting all unknown parameters. This study, however, has operated only the demand-related parameter which was defined as crucial for the occurrence and mitigation of the bullwhip. This decision was made considering the core characteristic of experimental methods that focuses on pure causal relationships, therefore, pure impacts of demand variation on the bullwhip occurrence were observed. This application should be a guide for experiments in the field in order to configure the settings of the conventional methods and gather more precise results.

7.4 Practical Implications

In line with the research balance framework of Stentoft and Rajkumar (2018), the practical implications of this study were also considered throughout the research process. The importance of this aspect was also highlighted by Corley and Gioia (2011) who recommend studies to improve the current managerial practices, as well as advancements in the theoretical knowledge.

The first managerial implication is directly related to the findings of this study. As observed, demand tracking and the associated bullwhip mitigation performance of managers from opposite cognitive propensity groups differ in relation to the decision-making environment of the SC. More specifically, System 2 managers (managers with a cognitive propensity for rational thinking system) overall track the demand changes better than System 1 managers in all three environmental settings tested in this study: higher information presence, time pressure and problem complexity. However, the results also demonstrated that time pressure may potentially increase the performance of System 1 managers.

These findings also have indirect implications for the management of SCs from various aspects. Firstly, considering the nature of the SC in association with the sector or relationships with SC partners, organisations can include the cognitive propensity in their recruitment criteria. Pan et al. (2020) mention that psychometric tests are often used in the recruitment processes of both public and private sectors and also recommend the utilisation of cognitive tests. In line with those authors, according to the findings of this study, an organisation with a complex SC network, with multiple material and demand flows, may be a more suitable place for a System 2 candidate than a System 1 to manage operations. On the other hand, a company operating in a more dynamic SC, in which last minute changes and decisions are frequently seen, may prefer recruiting either type of manager. Likewise, organisations from sectors with a stable demand flow can also hire from either managerial group. This differentiated allocation proposition also brings a solution to the concern of Pan et al. (2020) who mention the potential problem of hiring stereotypical candidates is that it may violate diversity. By recruiting different manager groups to roles with different characteristics or organisations will allow a more diverse workforce in organisations and sectors.

Moving on from this mindset, to increase the productivity and performance in the organisation, cognitive evaluation may also apply to the existing employees. According to their cognitive propensity, SC managers may be reassigned to a different role or department that will better suit their cognitive propensity. Profiling existing employees' cognitive propensities may also be helpful for the identification of the deficiencies in the process and consequently results in a better allocation of resources. For example, if a low performance or productivity is caused by the manager's cognitive propensity and the context of the decision or decision-making environment, then investing in operational solutions would not yield the expected return on investment. Instead, customised training programmes can be applied for distinct managerial groups. While System 1 managers can be trained for the common biases related to demand-tracking in OM, System 2 managers can have training for preserving decision quality levels when the time pressure increases.

In sum, a revised recruitment and reallocation strategy would benefit the performance of both managerial groups. With a better order management and hence the creation of a smoother bullwhip, the performance of the company and SC would increase as a consequence.

7.5 Limitations and Future Research Directions

Even though this research has provided multiple contributions from both theoretical and practical aspects, it also has a number of limitations, as anticipated. The main limitation of this study is related to the significance level of the statistical findings. As mentioned in the Results and Discussions chapters, some hypotheses' statistical confidence value approximated but did not meet the statistical threshold of 0.05. However, they were still indicating a substantial difference among the compared groups depending on the hypothesis and their potential insights were discussed where required. This decision can be based on two main reasons. This is the first study in this research direction, and it was considered that potential findings may provide an explorative foundation for further studies. Because of the same reason, the study was designed to cover multiple moderating conditions, which led ro its considerably simple experiment construction. Therefore, further studies are required in this domain to research and consolidate the findings of this study by focusing their works on the moderators individually.

Second limitation comes from the applied method, i.e. the experiment. As already explained, in spite of bringing a high level of internal validity to investigate the causal relationships, experimental methods lack in providing external validity. This study has applied three main strategies to enable the required external validity and generalisability: the application of an SC scenario in the OMG, development of the experiment tool with the consultation of SC experts, and lastly the selection of participants that are familiar with the field. However, it would still be beneficial to observe the reflection of this research's findings on studies that utilise research methods with higher external validity, such as case study or field experiment.

Thirdly, this study has employed a between-subjects experiment design. Although it is better than within-subjects design in terms of comparing differences among the cases, it may create a problem related to the individual differences. Although we believe that observed differences between the test groups derive from the cognitive differences of participants, other characteristics may interfere with the results. Although the within-subject design eliminates all contamination related to potential interpersonal differences, it was considered as unsuitable for the OMG stage of this study as the subsequent exposure of the same test with different configurations would lose its validity for the configurations seen at later stages. Therefore, future studies can devise another experiment setting that will deal with the drawbacks of the within-subjects design to reach more precise results. This would also reveal any possible misinterpretation made because of the between-subjects experiment design of this study.

Despite the selection of the between-subjects design, the learning curve in the OMG has been a concern and consequently a control variable for this study; however, any learning behaviour from either managerial group was not observed. In contrast, it was seen that System 2 managers' performance decreased throughout the 15-period game. This was considered to be related to the characteristic of the rational thinking system to assert high effort in decision-making. As the high effort consumption does not last long, the effective performance of participants might have been attenuated during the game. However, the length of the task in the real-life situation and in the simulation can be another research aspect for future studies.

Another limitation of this study was related to its application platform. For the several reasons already explained, this study was applied online. Since it decreased the control capability of the experimenter, the experiment was designed to be as simple and comprehensible as possible. One of the reasons for choosing an online platform was the requirement for a high number of participants, which was originally caused by the high number of variables. Therefore, future studies can both divide the components of this study into a series of works and conduct laboratory experiments. As well as a higher experimenter control helping to increase the reality of the experiment scenario, creating the scenario in a laboratory might add to participants' contribution.

It is also important to mention the limitation about the simplicity of the devised OMG. As explained in Appendices B (*p. 130*), the development of the game started in a BDG form, however, the first pilot study showed that there are length and clarity problems regarding the application of the game in that form. Therefore, considering the feedbacks of participants and field experts, the game was simplified and the existing OMG version was prepared. The simplicity of the game can be grouped under these three aspects: simple demand profile, single variable information and single-echelon decision-maker and the reasons were explained in Subsection 4.3.2 (b-IV) (*p. 63*). The potential threat of this simplicity could have been its operability and measurability. In other words, it must have been ensured that the designed experiment could measure the intended performance of participants in various configurations. This concern was checked in the second pilot run where the newly designed experiment was tested and its operational and measurement capabilities were affirmed. Therefore, it was concluded that the required simplicity of this OMG does not invalidate the validity of this experiment.

Besides further research requirements in relation to the limitations of this study, there are some aspects that can enhance the field from the points that his study has made. Firstly, the bullwhip mitigation was measured by participants' demand tracking performance in this study. Future studies may also isolate the other parameters that are commonly used in the field (e.g., biases related to supply line inventory, in-stock inventory and expected inventory) and measure their variations by using managers' cognitive propensity. Likewise, alternative methods to the OMG can be applied in further studies to minimise OMG-related limitations. This game was designed as a mixed derivative of BDG and NP experiments which were extensively utilised to understand the fundamentals of the inventory and order management issues in the OM field. Even though NP is not directly designed for BWE observation, the conventional BDG could have been a suitable alternative for the OMG. The reason for the development of OMG was the specific requirements of this experiment which were not sufficiently met by these other potential game-based experimental techniques (*see* Subsection 4.3.2 [b-II, III, *p. 61*). However, further studies on the topic can bring various revisions to the OMG using the dynamics of multi-echelon inventory management models according to their requirements, such as changing the demand trend, information type, manipulated information or the number of players per game.

Next, two more valuable extensions can be made from the behavioural aspect of this study. Depending on their correct and intuitively incorrect responses to the CRT, participants were categorised into two managerial groups in this study. However, there was another response type: non-intuitive incorrect. These are the answers that are not correct, but also not the expected impulsive answers. In line with the developments in the dual-thinking theories, further categorisation can be made utilising the non-intuitive incorrect answers and further interpretation can be gathered in relation to the role of cognition in BWE mitigation. The other extension direction, on the other hand, is related to the methodology that was used to select the five CRT questions used in this study. Since multiple exposure and the measurement power of the CRT questions are matters of concern for studies employing this test, there is a need for a systematic procedure for creating a customised CRT. Considering the low exposure and high measurement power of the selected five questions in this study, a future study can enhance the procedure that this study has followed and produce a systematic CRT creation method.

7.6 Chapter Summary

In this final chapter of the thesis, research questions that had been asked at the beginning were revisited. The findings of the study were matched by the questions to show that the aims of this study were realised.

Next, the theoretical contributions of the findings were presented from three aspects: contributions to the literatures of BSCM and BWE, an extension of dual-process theory in different SC decision-making contexts, and the devised experiment tool for the measurement of this study's variables.

The chapter continued by explaining the practical reflections of this study. The findings brought implications about the cognitive propensity-dependent performance of supply chain managers that can be considered in recruitment and allocation of human resources.

Lastly, limitations of this study were mentioned together with the related future research directions.

References

Ackerman, R., 2014. The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General*, *143*(3), p.1349. doi: 10.1037/a0035098

Allais, M., 1953. Violations of the betweenness axiom and nonlinearity in probability. *Econometrica*, *21*, pp.503-546. doi:10.2307/1907921

Alós-Ferrer, C., Garagnani, M. and Hügelschäfer, S., 2016. Cognitive reflection, decision biases, and response times. *Frontiers in Psychology*, *7*, p.1402. doi: 10.3389/fpsyg.2016.01402.

Ancarani, A., Di Mauro, C. and D'Urso, D., 2013. A human experiment on inventory decisions under supply uncertainty. *International Journal of Production Economics*, *142*(1), pp.61-73. doi: 10.1016/j.ijpe.2012.09.001.

Ayal, S., Rusou, Z., Zakay, D. and Hochman, G., 2015. Determinants of judgment and decision making quality: the interplay between information processing style and situational factors. *Frontiers in Psychology*, *6*, p.1088. doi: 10.3389/fpsyg.2015.01088

Bachrach, D.G. and Bendoly, E., 2011. Rigor in behavioral experiments: A basic primer for supply chain management researchers. *Journal of Supply Chain Management*, *47*(3), pp.5-8. doi: 10.1111/j.1745-493X.2011.03230.x.

Baganha, M.P. and Cohen, M.A., 1998. The stabilizing effect of inventory in supply chains. *Operations Research*, *46*(3-supplement-3), pp.S72-S83.

Baihaqi, I. and Beaumont, N., 2006. Information sharing in supply chains: a literature review and research agenda, *Department of Management, Monash University*. Available at: http://www.buseco.monash.edu.au/mgt/research/working-papers/2006/wp24-06.pdf.

Banerjee, A. and Burton, J.S., 1994. Coordinated vs. independent inventory replenishment policies for a vendor and multiple buyers. *International Journal of Production Economics*, 35(1-3), pp.215-222. doi: 10.1016/0925-5273(94)90084-1

Baron, J., Scott, S., Fincher, K. and Metz, S.E., 2015. Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition*, *4*(3), pp.265-284. doi: 10.1016/j.jarmac.2014.09.003

Barr, N., Pennycook, G., Stolz, J.A. and Fugelsang, J.A., 2015. The brain in your pocket: Evidence that Smartphones are used to supplant thinking. *Computers in Human Behavior*, *48*, pp.473-480. doi: 10.1016/j.chb.2015.02.029

Barratt, M. and Oke, A., 2007. Antecedents of supply chain visibility in retail supply chains: a resource-based theory perspective. *Journal of Operations Management*, 25(6), pp.1217-1233. doi: 10.1016/j.jom.2007.01.003

Bell, E. and Bryman, A., 2007. The ethics of management research: an exploratory content analysis. *British Journal of Management*, *18*(1), pp.63-77. doi: 10.1111/j.1467-8551.2006.00487.x

Bendoly, E., Donohue, K. and Schultz, K.L., 2006. Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of Operations Management*, *24*(6), pp.737-752.doi: j.jom.2005.10.001

Bendoly, E. and Swink, M., 2007. Moderating effects of information access on project management behavior, performance and perceptions. *Journal of Operations Management*, *25*(3), pp.604-622. doi: 10.1016/j.jom.2006.02.009.

Bialek, M. and Domurat, A., 2018. Cognitive abilities, analytic cognitive style and overconfidence: a commentary on Duttle (2016). *Bulletin of Economic Research*, *70*(1), pp.E119-E125. doi: 10.1111/boer.12117

Bialek, M. and Pennycook, G., 2018. The cognitive reflection test is robust to multiple exposures. *Behavior Research Methods*, *50*(5), pp.1953-1959.doi: 10.3758/s13428-017-0963-x.

Blacksmith, N., Yang, Y., Behrend, T.S. and Ruark, G.A., 2019. Assessing the validity of inferences from scores on the cognitive reflection test. *Journal of Behavioral Decision Making*, *32*(5), pp.599-612. doi: 10.1002/bdm.2133.

Bloomfield, R.J. and Kulp, S.L., 2013. Durability, transit lags, and optimality of inventory management decisions. *Production and Operations Management*, *22*(4), pp.826-842. doi: 10.1111/poms.12017

Bolton, G.E., Ockenfels, A. and Thonemann, U.W., 2012. Managers and students as newsvendors. *Management Science*, *58*(12), pp.2225-2233. doi: 10.1287/mnsc.1120.1550.

Bonabeau, E., 2003. Don't trust your gut. *Harvard business review*, *81*(5), pp.116-23. Available at: https://hbr.org/2003/05/dont-trust-your-gut

Boyer, K.K. and Swink, M.L., 2008. Empirical elephants—why multiple methods are essential to quality research in operations and supply chain management. *Journal of Operations Management*, *26(3)*, pp. 338–344. doi: 10.4324/9781315572413.

Brañas-Garza, P., Kujal, P. and Lenkei, B., 2015. Cognitive Reflection Test: whom, how, when. *Journal of Behavioural and Experimental Economics*, *82*, pp. 1–51. doi: 10.1016/j.socec.2019.101455.

Bray, R.L. and Mendelson, H., 2012. Information transmission and the bullwhip effect: An empirical investigation. *Management Science*, *58*(5), pp.860-875. doi: 10.1287/mnsc.1110.1467

Bustinza, O.F., Parry, G.C. and Vendrell-Herrero, F., 2013. Supply and demand chain management: The effect of adding services to product offerings. *Supply Chain Management: An International Journal*, *18(6)*, pp. 618–629. doi: 10.1108/SCM-05-2013-0149.

Cachon, G.P., 1999. Managing supply chain demand variability with scheduled ordering policies. *Management science*, *45*(6), pp.843-856. doi: 10.1287/mnsc.45.6.843

Cachon, G.P. and Fisher, M., 2000. Supply chain inventory management and the value of shared information. *Management science*, *46*(8), pp.1032-1048. doi: 10.1287/mnsc.46.8.1032.12029

Cachon, G.P. and Lariviere, M.A., 1999. Capacity choice and allocation: Strategic behavior and supply chain performance. *Management science*, *45*(8), pp.1091-1108. doi:10.1287/mnsc.45.8.1091

Cachon, G.P., Randall, T. and Schmidt, G.M., 2007. In search of the bullwhip effect. *Manufacturing & Service Operations Management*, *9*(4), pp.457-479. doi: 10.1287/msom.1060.0149

Campbell, D.J., 1988. ask complexity: A review and analysis. *Academy of Management Review*, 13(1), pp.40-52. doi: 10.5465/amr.1988.4306775

Campitelli, G. and Labollita, M., 2010. Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making*, *5*(3), pp.182-191. doi: 10.1037/t66183-000

Cantor, D.E. and Macdonald, J.R., 2009. Decision-making in the supply chain: examining problem solving approaches and information availability. *Journal of Operations Management*, *27*(3), pp.220-232. doi: 10.1016/j.jom.2008.09.002

Cantor, D.E. and Katok, E., 2012. Production smoothing in a serial supply chain: A laboratory investigation. *Transportation Research Part E: Logistics and Transportation Review*, *48*(4), pp.781-

794. doi: 10.1016/j.tre.2012.01.005

Cantor, D.E., Blackhurst, J.V. and Cortes, J.D., 2014. The clock is ticking: The role of uncertainty, regulatory focus, and level of risk on supply chain disruption decision making behavior. *Transportation Research Part E: Logistics and Transportation Review*, *72*, pp.159-172. doi: 10.1016/j.tre.2014.10.007.

Caridi, M., Crippa, L., Perego, A., Sianesi, A. and Tumino, A., 2010. Do virtuality and complexity affect supply chain visibility? *International Journal of Production Economics*, *127*(2), pp.372-383. doi: 10.1016/j.ijpe.2009.08.016.

Caridi, M., Moretto, A., Perego, A. and Tumino, A., 2014. The benefits of supply chain visibility: A value assessment model. *International Journal of Production Economics*, *151*, pp.1-19. doi: 10.1016/j.ijpe.2013.12.025.

Carter, C.R., Kaufmann, L. and Michel, A., 2007. Behavioral supply management: a taxonomy of judgment and decision-making biases. *International Journal of Physical Distribution & Logistics Management*, 37(8), pp. 631–669. doi: 10.1108/09600030710825694.

Carter, C.R., Kaufmann, L. and Wagner, C.M., 2017. Reconceptualizing intuition in supply chain management. *Journal of Business Logistics*, *38*(2), pp.80-95. doi: 10.1111/jbl.12154.

Carter, C.R., Rogers, D.S. and Choi, T.Y., 2015. Toward the theory of the supply chain. *Journal of Supply Chain Management*, *51*(2), pp.89-97.doi: 10.1111/jscm.12073

Castañeda, J.A., 2019. Behavioral Experiments in Single-Agent Operational Problems. In *Decision-making in Humanitarian Operations* (pp. 127-146). Palgrave Macmillan, Cham. doi: 10.1007/978-3-319-91509-8.

Chaiken, S. and Trope, Y. (Eds.), 1999. *Dual-process theories in social psychology*. Guilford Press.

Chater, N., Felin, T., Funder, D.C., Gigerenzer, G., Koenderink, J.J., Krueger, J.I., Noble, D., Nordli, S.A., Oaksford, M., Schwartz, B. and Stanovich, K.E., 2018. Mind, rationality, and cognition: An interdisciplinary debate. *Psychonomic Bulletin & Review*, *25*(2), pp.793-826. doi: 10.3758/s13423-017-1333-5

Chatfield, D.C., Kim, J.G., Harrison, T.P. and Hayya, J.C., 2004. The bullwhip effect—impact of stochastic lead time, information quality, and information sharing: a simulation study. *Production and operations management*, *13*(4), pp.340-353. doi: 10.1111/j.1937-5956.2004.tb00222.x

Chen, F., 1998. Echelon reorder points, installation reorder points, and the value of centralized demand information. *Management Science*, *44*(12-part-2), pp.S221-S234. doi: 10.1287/mnsc.44.12.S221

Chen, F., 1999. Decentralized supply chains subject to information delays. *Management Science*, *45*(8), pp.1076-1090. doi: 10.1287/mnsc.45.8.1076

Chen, F., Drezner, Z., Ryan, J.K. and Simchi-Levi, D., 2000. Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science*, *46*(3), pp.436-443. doi: 10.1287/mnsc.46.3.436.12069.

Chen, Y.F. and Disney, S.M., 2007. The myopic order-up-to policy with a proportional feedback controller. *International Journal of Production Research*, *45*(2), pp.351-368. doi: 10.1080/00207540600579532.

Chen, Y., Su, X. and Zhao, X., 2012. Modeling bounded rationality in capacity allocation games with the quantal response equilibrium. *Management Science*, *58*(10), pp.1952-1962. doi: 10.1080/00207540600579532

Cheyne, J.A. and Pennycook, G., 2013. Sleep paralysis postepisode distress: Modeling potential effects of episode characteristics, general psychological distress, beliefs, and cognitive style. *Clinical Psychological Science*, *1*(2), pp.135-148. doi: 10.1177/2167702612466656

Christopher, M.I., 2017. *Logistics & supply chain management*. Pearson Education Limited.

Christopher, M. and Ryals, L.J., 2014. The supply chain becomes the demand chain. *Journal of Business Logistics*, *35*(1), pp.29-35. doi: 10.1111/jbl.12037

Cokely, E.T. and Kelley, C.M., 2009. Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision making, 4*(1), pp.20–33.

Collis, J. and Hussey, R., 2013. *Business research: A practical guide for undergraduate and postgraduate students*. Macmillan International Higher Education.

Coppini, M., Rossignoli, C., Rossi, T. and Strozzi, F., 2010. Bullwhip effect and inventory oscillations analysis using the beer game model. *International journal of Production Research*, *48*(13), pp.3943-3956. doi: 10.1080/00207540902896204.

Corgnet, B., Espín, A.M. and Hernán-González, R., 2015. The cognitive basis of social behavior: cognitive reflection overrides antisocial but not always prosocial motives. *Frontiers in Behavioral Neuroscience*, *9*, p.287. doi: 10.3389/fnbeh.2015.00287

Corley, K.G. and Gioia, D.A., 2011. Building theory about theory building: what constitutes a theoretical contribution?. *Academy of Management Review*, *36*(1), pp.12-32. doi: 10.5465/amr.2009.0486

Creane, A., 2008. Input suppliers, differential pricing, and information sharing agreements. *Journal of Economics & Management Strategy*, *17*(4), pp.865-893. doi: 10.1111/j.1530-9134.2008.00198.x

Croson, R. and Donohue, K., 2006. Behavioral causes of the bullwhip effect and the observed value of inventory information. *Management Science*, *52*(3), pp.323-336. doi: 10.1287/mnsc.1050.0436.

Croson, R., Donohue, K., Katok, E. and Sterman, J., 2014. Order stability in supply chains: Coordination risk and the role of coordination stock. *Production and Operations Management*, *23*(2), pp.176-196. doi: 10.1111/j.1937-5956.2012.01422.x.

Cueva, C., Iturbe-Ormaetxe, I., Mata-Pérez, E., Ponti, G., Sartarelli, M., Yu, H. and Zhukova, V., 2016. Cognitive (ir) reflection: New experimental evidence. *Journal of Behavioral and Experimental Economics*, *64*, pp.81-93. doi: 10.1016/j.socec.2015.09.002

Cui, R., Allon, G., Bassamboo, A. and Van Mieghem, J.A., 2015. Information sharing in supply chains: An empirical and theoretical valuation. *Management Science*, *61*(11), pp.2803-2824. doi: 10.1287/mnsc.2014.2132.

Damnjanović, K., Novković, V., Pavlović, I., Ilić, S. and Pantelić, S., 2019. A cue for rational reasoning: Introducing a reference point in cognitive reflection tasks. *Europe's Journal of Psychology*, *15*(1), pp. 25–40. doi: 10.5964/ejop.v15i1.1701.

Dane, E. and Pratt, M.G., 2007. Exploring intuition and its role in managerial decision making. *Academy of Management Review*, *32*(1), pp.33-54. doi: 10.2307/20159279

Dejonckheere, J., Disney, S.M., Lambrecht, M.R. and Towill, D.R., 2004. The impact of information enrichment on the bullwhip effect in supply chains: A control engineering perspective. *European Journal of Operational Research*, *153*(3), pp.727-750. doi: 10.1016/S0377-2217(02)00808-1

De La Fuente, D. and Lozano, J., 2007. Application of distributed intelligence to reduce the bullwhip effect. *International Journal of Production Research*, *45*(8), pp.1815-1833. doi: 10.1080/00207540500456021

Del Missier, F., Mäntylä, T. and Bruine de Bruin, W., 2012. Decision-making competence, executive functioning, and general cognitive abilities. *Journal of Behavioral Decision Making*, *25*(4), pp.331-351. doi: 10.1002/bdm.731

De Neys, W. (Ed.), 2017. *Dual process theory 2.0*. Routledge.

De Neys, W. and Glumicic, T., 2008. Conflict monitoring in dual process theories of thinking. *Cognition*, *106*(3), pp.1248-1299. doi: 10.1016/j.cognition.2007.06.002

De Neys, W., Rossi, S. and Houdé, O., 2013. Bats, balls, and substitution sensitivity: Cognitive misers are no happy fools. *Psychonomic bulletin & review*, *20*(2), pp.269-273. doi: 10.3758/s13423-013-0384-5

Diamond, J.M., 1983. Ecology: Laboratory, field and natural experiments. *Nature*, *304*(5927), pp.586-587.Disney, S.M. and Towill, D.R., 2003. On the bullwhip and inventory variance produced by an ordering policy. *Omega*, *31*(3), pp.157-167. doi: 10.1016/S0305-0483(03)00028-8.

Drucker, P. F., 1973. *Management: Tasks, Responsibilities, Practice*. Harper & Row. New York.

Easterby-Smith, M., Thorpe, R. and Jackson, P.R., 2012. *Management research*. Sage.

Edali, M. and Yasarcan, H., 2014. A mathematical model of the beer game. *Journal of Artificial Societies and Social Simulation*, *17*(4), p.2. doi: 10.18564/jasss.2555.

Epstein, R.A., Patai, E.Z., Julian, J.B. and Spiers, H.J., 2017. The cognitive map in humans: spatial navigation and beyond. *Nature Neuroscience*, *20*(11), p.1504. doi: 10.1038/nn.4656

Epstein, S., 1994. Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*(8), p.709. doi: 10.1037/0003-066X.49.8.709

Erceg, N. and Bubić, A., 2017. One test, five scoring procedures: different ways of approaching the cognitive reflection test. *Journal of Cognitive Psychology*, *29*(3), pp.381-392. doi: 10.1080/20445911.2016.1278004.

Erceg, N., Galić, Z. and Ružojčić, M., 2020. A reflection on cognitive reflection-testing convergent/divergent validity of two measures of cognitive reflection. *Judgment & Decision Making*, *15*(5), pp. 1-32. doi: 10.31234/osf.io/ewrtq

Eriksson, P. and Kovalainen, A., 2015. *Qualitative methods in business research: A practical guide to social research*. Sage.

Ettlie, J.E., Groves, K.S., Vance, C.M. and Hess, G.L., 2014. Cognitive style and innovation in organizations. *European Journal of Innovation Management*. doi: 10.1108/EJIM-11-2012-0101

Evans, J.S.B., 2003. In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), pp.454-459. doi: 10.1016/j.tics.2003.08.012.

Evans, J.S.B., 2007. On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, *13*(4), pp.321-339. doi: 10.1080/13546780601008825

Evans, J.S.B., 2008. Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Reviews of Psychology*, *59*, pp.255-278. doi: 10.1146/annurev.psych.59.103006.093629

Evans, J.S.B., 2010. Intuition and reasoning: A dual-process perspective. *Psychological Inquiry*, *21*(4), pp.313-326. doi: 10.1080/1047840X.2010.521057

Evans, J.S.B. and Over, D.E., 2013. *Rationality and reasoning*. Psychology Press.

Evans, J.S.B. and Stanovich, K.E., 2013. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, *8*(3), pp.223-241. doi: 10.1177/1745691612460685

Evans, J.S.B., Handley, S.J. and Bacon, A.M., 2009. Reasoning under time pressure: A study of causal conditional inference. *Experimental Psychology*, *56*(2), pp.77-83. doi: 10.1027/1618-3169.56.2.77.

Fan, C., Olorunniwo, F.O., Jolayemi, J. and Li, X., 2013. A characterization of lower-tier supplier visibility practices in supplier relationship management. *Supply Chain Forum: An International Journal*, *14*(1), pp.2–14. doi: 10.1080/16258312.2013.11517306.

Fawcett, S.E., Magnan, G.M. and McCarter, M.W., 2008. Benefits, barriers, and bridges to effective supply chain management. *Supply Chain Management: An International Journal*, *13*(1), pp.35–48. doi: 10.1108/13598540810850300.

Fisher, M.L., Hammond, J.H., Obermeyer, W.R. and Raman, A., 1994. Making supply meet demand in an uncertain world. *Harvard Business Review*, *72*, pp.83-83.

Forrester, J.W., 1958. Industrial Dynamics. A major breakthrough for decision makers. *Harvard Business Review*, *36*(4), pp.37-66.doi: 10.1225/58404

Forrester, J.W., 1961. Industrial dynamics. Cambridge, Boston, MA: MIT Press.

Franken, I.H. and Muris, P., 2005. Individual differences in decision-making. *Personality and Individual Differences*, *39*(5), pp.991-998. doi: 10.1016/j.paid.2005.04.004

Fraser-Mackenzie, P.A. and Dror, I.E., 2011. Dynamic reasoning and time pressure: Transition from analytical operations to experiential responses. *Theory and Decision*, *71*(2), pp.211-225. doi: 10.1007/s11238-009-9181-z

Frederick, S., 2005. Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), pp.25-42. doi: 10.1257/089533005775196732.

Fugate, B.S., Thomas, R.W. and Golicic, S.L., 2012. The impact of coping with time pressure on boundary spanner collaborative behaviors. *International Journal of Physical Distribution & Logistics Management, 42*(7), pp.697-715. doi: 10.1108/09600031211258156

Furlan, S., Agnoli, F. and Reyna, V.F., 2016. Intuition and analytic processes in probabilistic reasoning: The role of time pressure. *Learning and Individual Differences*, *45*, pp.1-10. doi: 10.1016/j.lindif.2015.11.006.

Gavirneni, S. and Isen, A.M., 2010. Anatomy of a newsvendor decision: Observations from a verbal protocol analysis. *Production and Operations Management*, *19*(4), pp.453-462. doi: 10.1111/j.1937-5956.2009.01110.x

Gavirneni, S., Kapuscinski, R. and Tayur, S., 1999. Value of information in capacitated supply chains. *Management Science*, *45*(1), pp.16-24. doi: 10.1287/mnsc.45.1.16.

Geary, S., Disney, S.M. and Towill, D.R., 2006. On bullwhip in supply chains—historical review, present practice and expected future impact. *International Journal of Production Economics*, *101*(1), pp.2-18. doi: 10.1016/j.ijpe.2005.05.009

Gervais, W.M. and Norenzayan, A., 2012. Analytic thinking promotes religious disbelief. *Science*, *336*(6080), pp.493-496. doi: 10.1126/science.1215647

Gigerenzer, G., 2003. *Reckoning with risk: learning to live with uncertainty*. Penguin UK.

Gino, F. and Ariely, D., 2012. The dark side of creativity: original thinkers can be more dishonest. *Journal of Personality and Social Psychology*, *102*(3), p.445. doi: 10.1037/a0026406

Gino, F. and Pisano, G., 2008. Toward a theory of behavioral operations. *Manufacturing & Service*

Operations Management, *10*(4), pp.676-691. doi: 10.1287/msom.1070.0205.

Gladwell, M., 2005. *Blink: The Power of Thinking without Thinking*. Little Brown and Company, New York, NY.

Goldberg, P., 1983. *The Intuitive Edge*. New York: Putnam.

Goldberg, P., 1990. 'The many faces of intuition' in *Intuition in Organizations*, W.H. Agor (Ed.) 62-77. Newbury Park, CA: Sage Publications.

Grant, A.M. and Wall, T.D., 2009. The neglected science and art of quasi-experimentation: Why-to, when-to, and how-to advice for organizational researchers. *Organizational Research Methods*, *12*(4), pp.653-686. doi: 10.1177/1094428108320737

Grayot, J.D., 2020. Dual process theories in behavioural economics and neuroeconomics: a critical review. *Review of Philosophy and Psychology*, *11*(1), pp.105-136. doi: 10.1007/s13164-019-00446-9

Gurnani, H., Ramachandran, K., Ray, S. and Xia, Y., 2014. Ordering behavior under supply risk: An experimental investigation. *Manufacturing & Service Operations Management*, *16*(1), pp.61-75. doi: 10.1287/msom.2013.0453

Haigh, M., 2016. Has the standard cognitive reflection test become a victim of its own success? *Advances in Cognitive Psychology*, *12*(3), p.145-149. doi: 10.5709/acp-0193-5

Hammond, K.R. and Brehmer, B., 1973. Quasi-Rationality and Distrust: Implications for International Conflict. *Human Judgment and Social Interaction*, p.338.

Healey, M.P., Vuori, T. and Hodgkinson, G.P., 2015. When teams agree while disagreeing: Reflexion and reflection in shared cognition. *Academy of Management Review*, *40*(3), pp.399-422. doi: 10.5465/amr.2013.0154

Heikkilä, J., 2002. From supply to demand chain management: efficiency and customer satisfaction. *Journal of Operations Management*, *20*(6), pp.747-767. doi: 10.1016/S0272-6963(02)00038-4

Helfat, C.E. and Peteraf, M.A., 2015. Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strategic Management Journal*, *36*(6), pp.831-850. doi: 10.1002/smj.2247.

Hoppe, E.I. and Kusterer, D.J., 2011. Behavioral biases and cognitive reflection. *Economics Letters*, *110*(2), pp.97-100. doi: 10.1016/j.econlet.2010.11.015

Huang, T., Allon, G. and Bassamboo, A., 2013. Bounded rationality in service systems. *Manufacturing & Service Operations Management*, *15*(2), pp.263-279. doi:10.1287/msom.1120.0417

Hwang, M.I., 1994. Decision making under time pressure: a model for information systems research. *Information & Management*, *27*(4), pp.197-203. doi: 10.1016/0378-7206(94)90048-5

Jankelová, N., Jankurová, A. and Rašovec, O., 2018. Psychological aspects of managerial decision-making. *Вестник Марийского государственного университета. Серия «Сельскохозяйственные науки. Экономические науки»*, *4*(2 (14)). doi: 10.30914/2411-9687-2018-4-2-134-144

Jüttner, U., Christopher, M. and Baker, S., 2007. Demand chain management-integrating marketing and supply chain management. *Industrial Marketing Management*, *36*(3), pp.377-392. doi: 10.1016/j.indmarman.2005.10.003

Kahn, J.A., 1987. Inventories and the volatility of production. *The American Economic Review*, pp.667-

679.

Kahneman, D., 2003. A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, *58*(9), p.697-720. doi: 10.1037/0003-066X.58.9.697

Kahneman, D., 2011. *Thinking, fast and slow*. Macmillan.

Kahneman, D. and Frederick, S., 2002. Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The Psychology of Intuitive Judgment*, *49*, p.81. doi: 10.1017/CBO9780511808098.004

Kahneman, D. and Klein, G., 2009. Conditions for intuitive expertise: a failure to disagree. *American Psychologist*, *64*(6), p.515. doi: doi.org/10.1037/a0016755

Kalkancı, B., Chen, K.Y. and Erhun, F., 2014. Complexity as a contract design factor: A human-to-human experimental study. *Production and Operations Management*, *23*(2), pp.269-284. doi: 10.1111/poms.12067

Kaminsky, P. and Simchi-Levi, D., 1998. A new computerized beer game: A tool for teaching the value of integrated supply chain management. *Global Supply Chain and Technology Management*, *1*(1), pp.216-225. Available at: https://kaminsky.ieor.berkeley.edu/Reprints/PK_DSL_98b.pdf

Katok, E., 2011a. Laboratory experiments in operations management. In *Transforming Research into Action* (pp. 15-35). INFORMS. doi: 10.1287/educ.1110.0090

Katok, E., 2011b. Using laboratory experiments to build better operations management models. *Foundations and Trends in Technology, Information and Operations Management*, *5*(1), pp.1–86. doi: 10.1561/0200000022.

Kaufmann, L., Carter, C.R. and Rauer, J., 2016. The coevolution of relationship dominant logic and supply risk mitigation strategies. *Journal of Business Logistics*, *37*(2), pp.87-106. doi: 10.1111/jbl.12126

Kaufmann, L., Michel, A. and Carter, C.R., 2009. Debiasing strategies in supply management decision-making. *Journal of Business Logistics*, *30*(1), pp.85-106. doi: 10.1002/j.2158-1592.2009.tb00100.x

Kaufmann, L., Wagner, C.M. and Carter, C.R., 2017. Individual modes and patterns of rational and intuitive decision-making by purchasing managers. *Journal of Purchasing and Supply Management*, *23*(2), pp.82-93. doi: 10.1016/j.pursup.2016.09.001.

Kiss, H.J., Rodriguez-Lara, I. and Rosa-García, A., 2016. Think twice before running! Bank runs and cognitive abilities. *Journal of Behavioral and Experimental Economics*, *64*, pp.12-19. doi: 10.1016/j.socec.2015.01.006

Kleinmuntz, D.N., 1993. Information processing and misperceptions of the implications of feedback in dynamic decision making. *System Dynamics Review*, *9*(3), pp.223-237. doi: 10.1002/sdr.4260090302

Klueber, R. and O'Keefe, R.M., 2013. Defining and assessing requisite supply chain visibility in regulated industries. *Journal of Enterprise Information Management*, *26*(3), pp. 295-315. doi: 10.1108/17410391311325243.

Knemeyer, A.M. and Walker Naylor, R., 2011. Using behavioral experiments to expand our horizons and deepen our understanding of logistics and supply chain decision making. *Journal of Business Logistics*, *32*(4), pp.296-302. doi: 10.1111/j.0000-0000.2011.01025.x.

Koehler, D.J. and James, G., 2010. Probability matching and strategy availability. *Memory & cognition*, *38*(6), pp.667-676. doi: 10.3758/MC.38.6.667

de Kok, T., Grob, C., Laumanns, M., Minner, S., Rambau, J. and Schade, K., 2018. A typology and literature review on stochastic multi-echelon inventory models. *European Journal of Operational Research*, 269(3), pp.955-983. doi: 10.1016/j.ejor.2018.02.047

Kosmol, T., Reimann, F. and Kaufmann, L., 2018. Co-alignment of supplier quality management practices and cognitive maps–A neo-configurational perspective. *Journal of Purchasing and Supply Management*, *24*(1), pp.1-20. doi: 10.1016/j.pursup.2017.11.002.

Kraft, T., Valdés, L. and Zheng, Y., 2018. Supply chain visibility and social responsibility: Investigating consumers' behaviors and motives. *Manufacturing & Service Operations Management*, *20*(4), pp.617-636. doi: 10.1287/msom.2017.0685

Lee, H.L., Padmanabhan, V. and Whang, S., 1997a. Information distortion in a supply chain: The bullwhip effect. *Management Science*, *43*(4), pp.546-558. doi: 10.1287/mnsc.43.4.546

Lee, H.L., Padmanabhan, V. and Whang, S., 1997b. The bullwhip effect in supply chains. *Sloan Management Review*, *38*, pp.93-102. doi: 10.1007/1-4020-0612-8_108

Lee, H.L., Padmanabhan, V. and Whang, S., 2004. Comments on "Information distortion in a supply chain: The bullwhip effect." *Management Science, 50*(12), pp.1887-1893. doi: 10.1287/mnsc.1040.0305

Lee, H.L., So, K.C. and Tang, C.S., 2000. The value of information sharing in a two-level supply chain. *Management Science*, *46*(5), pp.626-643. doi: 10.1287/mnsc.46.5.626.12047

Lewin, K., 1936. Pyschology of success and failure. *Occupations: The Vocational Guidance Journal, 14*(9), pp.926-930. doi: 10.1002/j.2164-5892.1936.tb00275.x

Liberali, J.M., Reyna, V.F., Furlan, S., Stein, L.M. and Pardo, S.T., 2012. Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making*, *25*(4), pp.361-381. doi: 10.1002/bdm.752

Maitland, E. and Sammartino, A., 2015. Managerial cognition and internationalization. *Journal of International Business Studies*, *46*(7), pp.733-760. doi: 10.1057/jibs.2015.9.

Mata, A., Schubert, A.L. and Ferreira, M.B., 2014. The role of language comprehension in reasoning: How "good-enough" representations induce biases. *Cognition*, *133*(2), pp.457-463. doi: 10.1016/j.cognition.2014.07.011

Mauboussin, M. J. 2009. *Think Twice*. Harvard Business Press, Cambridge, MA.

Maule, A.J., Hockey, G.R.J. and Bdzola, L., 2000. Effects of time-pressure on decision-making under uncertainty: changes in affective state and information processing strategy. *Acta Psychologica*, *104*(3), pp.283-301. doi: 10.1016/S0001-6918(00)00033-0

Mentzer, J.T. and Esper, T.L., 2010. Supply Chain Management Strategy. In J. Sheth and N. Malhotra (Eds.) *Wiley international encyclopaedia of marketing*, (pp.1-7). doi: 10.1002/9781444316568.wiem01056

Metters, R., 1997. Quantifying the bullwhip effect in supply chains. *Journal of Operations Management*, *15*(2), pp.89-100.doi: 10.1016/S0272-6963(96)00098-8

Moinzadeh, K., 2002. A multi-echelon inventory system with information exchange. *Management Science*, 48(3), pp.414-426. doi: 10.1287/mnsc.48.3.414.7730

Moritz, B.B., Hill, A.V. and Donohue, K.L., 2013. Individual differences in the newsvendor problem: Behavior and cognitive reflection. *Journal of Operations Management*, *31*(1-2), pp.72-85. 10.1016/j.jom.2012.11.006.

Moritz, B.B., Narayanan, A. and Parker, C., 2020. Unraveling behavioral ordering: Relative costs and the bullwhip effect. doi: 10.2139/ssrn.3195282

Moritz, B.B., Siemsen, E. and Kremer, M., 2014. Judgmental forecasting: Cognitive reflection and decision speed. *Production and Operations Management*, *23*(7), pp.1146-1160. doi: 10.1111/poms.12105

Nagarajan, M. and Shechter, S., 2014. Prospect theory and the newsvendor problem. *Management Science*, *60*(4), pp.1057-1062. doi: 10.1287/mnsc.2013.1804

Narayanan, A. and Moritz, B.B., 2015. Decision making and cognition in multi-echelon supply chains: An experimental study. *Production and Operations Management*, *24*(8), pp.1216-1234. doi: 10.1111/poms.12343.

Niranjan, T.T., Wagner, S.M. and Bode, C., 2011. An alternative theoretical explanation and empirical insights into overordering behavior in supply chains. *Decision Sciences*, *42*(4), pp.859-888. doi: 10.1111/j.1540-5915.2011.00334.x

Nisula, J., 2016. Does smart mean rational? A study of cognitive dispositions, heuristics and rationality. Available at: http://epub.lib.aalto.fi/en/ethesis/pdf/14766/hse_ethesis_14766.pdf

Nooraie, S.V. and Mellat Parast, M., 2015. A multi-objective approach to supply chain risk management: Integrating visibility with supply and demand risk. *International Journal of Production Economics*, *161*, pp.192-200. doi: 10.1016/j.ijpe.2014.12.024.

Obrecht, N.A., Chapman, G.B. and Gelman, R., 2009. An encounter frequency account of how experience affects likelihood estimation. *Memory & Cognition*, *37*(5), pp.632-643. doi: 10.3758/MC.37.5.632

O'Donnell, T., Maguire, L., McIvor, R. and Humphreys, P., 2006. Minimizing the bullwhip effect in a supply chain using genetic algorithms. *International Journal of Production Research*, *44*(8), pp.1523-1543. doi: 10.1080/00207540500431347

Oechssler, J., Roider, A. and Schmitz, P.W., 2009. Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, *72*(1), pp.147-152. doi: 10.1016/j.jebo.2009.04.018

Oldrati, V., Patricelli, J., Colombo, B. and Antonietti, A., 2016. The role of dorsolateral prefrontal cortex in inhibition mechanism: A study on cognitive reflection test and similar tasks through neuromodulation. *Neuropsychologia*, *91*, pp.499-508. doi: 10.1016/j.neuropsychologia.2016.09.010

Oliva, R. and Gonçalves, P., 2005. Behavioural causes of the bullwhip effect: "Satisficing" policies with limited information cues. *Proceedings of the 23rd International Conference of the System Dynamics Society*, pp.118-119.

Ordonez, L. and Benson III, L., 1997. Decisions under time pressure: How time constraint affects risky decision making. *Organizational Behavior and Human Decision Processes*, *71*(2), pp.121-140. doi: 10.1006/obhd.1997.2717

Ovchinnikov, A., Moritz, B. and Quiroga, B.F., 2015. How to compete against a behavioural newsvendor. *Production and Operations Management*, *24*(11), pp.1783-1793. doi: 10.1111/poms.12361

Özer, Ö., Zheng, Y. and Ren, Y., 2014. Trust, trustworthiness, and information sharing in supply chains bridging China and the United States. *Management Science*, *60*(10), pp.2435-2460. doi: 10.2139/ssrn.1961774.

Pan, J., Shachat, J. and Wei, S., 2020. Cognitive reflection and economic order quantity inventory management: An experimental investigation. *Managerial and Decision Economics*, *41*(6), pp.998-1009. doi: 10.1002/mde.3154

Patel, N., 2017. *The Cognitive Reflection Test: a measure of intuition/reflection, numeracy, and insight problem solving, and the implications for understanding real-world judgments and beliefs* (Doctoral dissertation, University of Missouri--Columbia). doi: 10.31234/osf.io/xeyj8

Pennycook, G., Cheyne, J.A., Koehler, D.J. and Fugelsang, J.A., 2016. Is the cognitive reflection test a measure of both reflection and intuition? *Behavior Research Methods*, *48*(1), pp.341-348. doi: 10.3758/s13428-015-0576-1.

Pennycook, G., Cheyne, J.A., Seli, P., Koehler, D.J. and Fugelsang, J.A., 2012. Analytic cognitive style predicts religious and paranormal belief. *Cognition*, *123*(3), pp.335-346. doi: 10.1016/j.cognition.2012.03.003

Pennycook, G., Fugelsang, J.A. and Koehler, D.J., 2015. What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, pp.34-72.

Petty, R.E., Briñol, P., Loersch, C. and McCaslin, M.J., 2009. The need for cognition. In M.R. Leary, M.R. and R.H. Hoyle (Eds.) *Handbook of individual differences in social behavior*, (pp.318-329), New York: Guilford Press.

Peysakhovich, A. and Rand, D.G., 2016. Habits of virtue: Creating norms of cooperation and defection in the laboratory. *Management Science*, *62*(3), pp.631-647. doi: 10.1287/mnsc.2015.2168

Ponti, G. and Rodriguez-Lara, I., 2015. Social preferences and cognitive reflection: evidence from a dictator game experiment. *Frontiers in behavioral neuroscience*, *9*, p.146. doi: 10.3389/fnbeh.2015.00146

Porter, M.E., 1985. *The Competitive Advantage: Creating and Sustaining Superior Performance*, Free Press, New York.

Primi, C., Morsanyi, K., Chiesi, F., Donati, M.A. and Hamilton, J., 2016. The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making*, *29*(5), pp.453-469. doi: 10.1002/bdm.1883

Remenyi, D., Williams, B., Money, A. and Swartz, E., 1998. *Doing research in business and management: an introduction to process and method*. Sage.

Rokeach, M. and Kliejunas, P., 1972. Behavior as a function of attitude-toward-object and attitude-toward-situation. *Journal of Personality and Social Psychology, 22*(2), pp.194–201. doi: 10.1037/h0032614

Saunders, M., Lewis, P. and Thornhill, A., 2009. *Research methods for business students*. Pearson Education.

Schorsch, T., Wallenburg, C.M. and Wieland, A., 2017. The human factor in SCM: Introducing a meta-theory of behavioral supply chain management. *International Journal of Physical Distribution & Logistics Management*, *47*(4), pp. 238–262. doi: 10.1108/ijpdlm-10-2015-0268.

Schweitzer, M.E. and Cachon, G.P., 2000. Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science*, *46*(3), pp.404-420. doi: 10.1287/mnsc.46.3.404.12070.

Sekaran, U. and Bougie, R., 2016. *Research methods for business: A skill building approach*. John Wiley & Sons.

Shan, J., Yang, S., Yang, S. and Zhang, J., 2014. An empirical study of the bullwhip effect in

China. *Production and Operations Management*, *23*(4), pp.537-551. doi: 10.1111/poms.12034

Shenhav, A., Rand, D.G. and Greene, J.D., 2012. Divine intuition: Cognitive style influences belief in God. *Journal of Experimental Psychology: General*, *141*(3), p.423. doi: 10.1037/a0025391

Shoda, Y., 2004. Individual Differences in Social Psychology: Understanding Situations to Understand People, Understanding People to Understand Situations. In C. Sansone, C.C. Morf, & A.T. Panter (Eds.), *The Sage handbook of methods in social psychology* (p. 117–141). Sage Publications, Inc.

Shtulman, A. and McCallum, K., 2014. Cognitive reflection predicts science understanding. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36, No. 36).

Siemsen, E., 2011. The usefulness of behavioral laboratory experiments in supply chain management research. *Journal of Supply Chain Management*, *47*(3), pp.17-18. doi: 10.1111/j.1745-493X.2011.03227.x.

Simon, H.A., 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), pp.99-118.doi: 10.2307/1884852

Simon, H.A., 1956. Rational choice and the structure of the environment. *Psychological Review*, *63*(2), p.129. doi: 10.1037/h0042769

Simon, H.A., 1957. *Models of Man: Social and Rational*. John Wiley & Sons, New York

Simon, H.A., 1978. Rationality as process and as product of thought. *The American Economic Review*, *68*(2), pp.1-16. doi: 10.1017/cbo9780511598951.005.

Simon, H.A., 1987. Making management decisions: The role of intuition and emotion. *Academy of Management Perspectives*, *1*(1), pp.57-64. doi: 10.5465/ame.1987.4275905.

Sloman, S.A., 1996. The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), p.3. doi: 10.1037/0033-2909.119.1.3

Sogomonian, A.G. and Tang, C.S., 1993. A modeling framework for coordinating promotion and production decisions within a firm. *Management Science*, *39*(2), pp.191-203. doi: 10.1287/mnsc.39.2.191.

Šrol, J., 2018. These problems sound familiar to me: Previous exposure, cognitive reflection test, and the moderating role of analytic thinking. *Studia Psychologica*, *60*(3), pp.195-208. doi: 10.21909/sp.2018.03.762.

Stalk Jr, G. and Hout, T.M., 1990. Competing against time. *Research-Technology Management*, *33*(2), pp.19-24. doi: 10.1080/08956308.1990.11670646

Stanovich, K., 2011. *Rationality and the reflective mind*. Oxford University Press.

Stanovich, K.E., 1999. *Who is rational?: Studies of individual differences in reasoning*. Psychology Press.

Stanovich, K.E., 2009. Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In J.S.B.T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55-88). Oxford University Press. doi: 10.1093/acprof:oso/9780199230167.003.0003

Stanovich, K.E., 2016. The comprehensive assessment of rational thinking. *Educational Psychologist*, *51*(1), pp.23-34. doi: 10.1080/00461520.2015.1125787.

Stanovich, K.E. and West, R.F., 2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*(5), pp.645-665. doi: 10.1017/S0140525X00003435

Steckel, J.H., Gupta, S. and Banerji, A., 2004. Supply chain decision making: Will shorter cycle times and shared point-of-sale information necessarily help? *Management Science*, *50*(4), pp.458-464. doi: 10.1287/mnsc.1030.0169.

Stentoft, J. and Rajkumar, C., 2018. Balancing theoretical and practical relevance in supply chain management research. *International Journal of Physical Distribution & Logistics Management*, *48*(5), pp. 504–523. doi: 10.1108/IJPDLM-01-2018-0020.

Sterman, J.D., 1989a. Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes*, *43*(3), pp.301-335. doi: 10.1016/0749-5978(89)90041-1

Sterman, J.D., 1989b. Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*, *35*(3), pp.321-339. doi: 10.1287/mnsc.35.3.321

Sterman, J.D., 1994. Learning in and about complex systems. *System Dynamics Review*, *10*(2-3), pp.291-330. doi: 10.1002/sdr.4260100214

Sterman, J.D., 2000. Business Dynamics: System Thinking and Modeling for a Complex World Irwin McGraw-Hill. *Massachusetts Institute of Technology, Engineering Systems Division: Cambridge, MA, USA*.

Stevens, C.K., 2011. Questions to consider when selecting student samples. *Journal of Supply Chain Management*, *47*(3), pp.19-21. doi: 10.1111/j.1745-493X.2011.03233.x.

Stieger, S. and Reips, U.D., 2016. A limitation of the Cognitive Reflection Test: familiarity. *PeerJ 4:e2395*. doi: 10.7717/peerj.2395.

Strozzi, F., Bosch, J. and Zaldívar, J.M., 2007. Beer game order policy optimization under changing customer demand. *Decision Support Systems*, *42*(4), pp.2153-2163. doi: 10.1016/j.dss.2006.06.001.

Stupple, E.J., Pitchford, M., Ball, L.J., Hunt, T.E. and Steel, R., 2017. Slower is not always better: Response-time evidence clarifies the limited role of miserly information processing in the Cognitive Reflection Test. *PloS One*, 12(11). doi: 10.1371/journal.pone.0186404

Szaszi, B., Szollosi, A., Palfi, B. and Aczel, B., 2017. The cognitive reflection test revisited: exploring the ways individuals solve the test. *Thinking & Reasoning*, *23*(3), pp.207-234. doi: 10.1080/13546783.2017.1292954.

Tan, K.C., 2001. A framework of supply chain management literature. *European Journal of Purchasing & Supply Management*, *7*(1), pp.39-48. doi: 10.1016/S0969-7012(00)00020-4.

Terwiesch, C., Ren, Z.J., Ho, T.H. and Cohen, M.A., 2005. An empirical analysis of forecast sharing in the semiconductor equipment supply chain. *Management Science*, *51*(2), pp.208-220.

Thomas, K.A. and Clifford, S., 2017. Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior*, *77*, pp.184-197. doi: 10.1016/j.chb.2017.08.038.

Thomas, R.W., Esper, T.L. and Stank, T.P., 2010. Testing the negative effects of time pressure in retail supply chain relationships. *Journal of Retailing*, *86*(4), pp.386-400. doi: 10.1016/j.jretai.2010.07.011.

Thomas, R.W., Esper, T.L. and Stank, T.P., 2011a. Coping with time pressure in interfirm supply chain relationships. *Industrial Marketing Management*, *40*(3), pp.414-423. doi: 10.1016/j.indmarman.2010.04.008

Thomas, R.W., Fugate, B.S. and Koukova, N.T., 2011b. Coping with time pressure and knowledge sharing in buyer–supplier relationships. *Journal of Supply Chain Management*, *47*(3), pp.22-42. doi: 10.1111/j.1745-493X.2011.03229.x

Thomas, R.W., Davis-Sramek, B., Esper, T.L. and Murfield, M.L., 2014. Attribution effects of time pressure in retail supply chain relationships: Moving from "what" to "why". *Journal of Business Logistics*, *35*(3), pp.225-240. doi: 10.1111/jbl.12056

Thomson, K.S. and Oppenheimer, D.M., 2016. Investigating an alternate form of the cognitive reflection test. *Judgment and Decision making*, *11*(1), p.99-113.

Tokar, T., Aloysius, J.A. and Waller, M.A., 2012. Supply chain inventory replenishment: The debiasing effect of declarative knowledge. *Decision Sciences*, *43*(3), pp.525-546. doi: 10.1111/j.1540-5915.2012.00355.x

Toplak, M.E. and Stanovich, K.E., 2002. The domain specificity and generality of disjunctive reasoning: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, *94*(1), p.197. doi: 10.1037/0022-0663.94.1.197

Toplak, M.E., West, R.F. and Stanovich, K.E., 2011. The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*(7), pp.1275-1289. doi: 10.3758/s13421-011-0104-1

Toplak, M.E., West, R.F. and Stanovich, K.E., 2014. Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*(2), pp.147-168. doi: 10.1080/13546783.2013.844729.

Travers, E., Rolison, J.J. and Feeney, A., 2016. The time course of conflict on the Cognitive Reflection Test. *Cognition*, *150*, pp.109-118. doi: 10.1016/j.cognition.2016.01.015

Trémolière, B. and De Neys, W., 2014. When intuitions are helpful: Prior beliefs can support reasoning in the bat-and-ball problem. *Journal of Cognitive Psychology*, *26*(4), pp.486-490. doi: 10.1080/20445911.2014.899238

Trochim, W.M. and Donnelly, J.P., 2001. *Research methods knowledge base* (Vol. 2). Atomic Dog Pub.

Tsanos, C.S. and Zografos, K.G., 2016. The effects of behavioural supply chain relationship antecedents on integration and performance. *Supply Chain Management: An International Journal*, *21*(6), pp. 678–693. doi: 10.1108/SCM-06-2016-0211.

Tversky, A. and Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), pp.1124-1131. doi: 10.1126/science.185.4157.1124.

Tversky, A. and Kahneman, D., 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), p.293. doi: 10.1037/0033-295X.90.4.293

Vandendriessche, T., 2017. *Three novel games of information and competition: exploring human strategic reasoning* (Doctoral dissertation, University of Warwick).

Van Riel, A.C., Ouwersloot, J. and Lemmink, J.G.A.M., 2003. Antecedents of effective decision making: A cognitive approach. Available at: https://core.ac.uk/download/pdf/231396541.pdf

Wan, X. and Evers, P.T., 2011. Supply chain networks with multiple retailers: A test of the emerging theory on inventories, stockouts, and bullwhips. *Journal of Business Logistics*, *32*(1), pp.27-39. doi: 10.1111/j.2158-1592.2011.01003.x

Wang, Q., Yang, S., Liu, M., Cao, Z. and Ma, Q., 2014. An eye-tracking study of website complexity from cognitive load perspective. *Decision Support Systems*, 62, pp.1-10. doi: 10.1016/j.dss.2014.02.007

Wason, P.C. and Evans, J.S.B., 1975. Dual processes in reasoning? *Cognition*, *3*(2), pp.141-154. doi: dx.doi.org/10.1016/0010-0277(74)90017-1

Weller, J.A., Dieckmann, N.F., Tusler, M., Mertz, C.K., Burns, W.J. and Peters, E., 2013. Development and testing of an abbreviated numeracy scale: A Rasch analysis approach. *Journal of Behavioral Decision Making*, *26*(2), pp.198-212. doi: 10.1002/bdm.1751

Welsh, M. and Begg, S., 2017. The Cognitive Reflection Test: familiarity and predictive power in professionals. In *39th Annual Meeting of the Cognitive Science Society*, 1, pp. 3497-3502. Available at: https://mindmodeling.org/cogsci2017/papers/0659/paper0659.pdf.

Wieland, A., Handfield, R.B. and Durach, C.F., 2016. Mapping the landscape of future research themes in supply chain management. *Journal of Business Logistics*, *37*(3), pp.205-212.doi: 10.1111/jbl.12131

Wu, D.Y. and Katok, E., 2006. Learning, communication, and the bullwhip effect. *Journal of Operations Management*, *24*(6), pp.839-850. doi: 10.1016/j.jom.2005.08.006.

Xiong, G. and Helo, P., 2006. An application of cost-effective fuzzy inventory controller to counteract demand fluctuation caused by bullwhip effect. *International Journal of Production Research*, *44*(24), pp.5261-5277. doi: 10.1080/00207540600600114

Yin, R.K., 2009. *Case Study Research - Design and Methods* 4th ed., Sage.

Zhao, Y. and Zhao, X., 2015. On human decision behavior in multi-echelon inventory management. *International Journal of Production Economics*, *161*, pp.116-128. doi: 10.1016/j.ijpe.2014.12.005.

Appendices

A – Pool of Potential CRT Questions

	Question	Source	Notes
1	A postcard and a pen cost 110 cents in total. The postcard costs 100 cents more than the pen. How much does the pen cost? *[correct answer 5 cents; intuitive answer 10 cents]*	Alós-Ferrer et al. (2016)	• Bat and ball derivative
2	If it takes 5 machines 5 minutes to make 5 car tires, how long would it take 100 machines to make 100 car tires? (In minutes) [correct answer = 5] *[correct answer 5 minutes; intuitive answer 100 minutes]*	Alós-Ferrer et al. (2016)	• Widget derivative
3	A rope ladder hangs over the side of a boat with the bottom rung on the surface of the water. The rope ladder has 6 rungs that are 30 cm apart from each other. The tide rises 70 cm. How many rungs will stick out of the water at high tide? *[correct answer 6 rungs; intuitive answer 3 rungs]*	Oldrati (2016)	• Complex
4	There are 12 one-cent stamps in a dozen. How many two-cent stamps are there in a dozen? *[correct answer 12 stamps; intuitive answer 6 stamps]*	Oldrati (2016)	• Tricky
5	A farmer makes 4 piles of hay in one corner of a field and other 5 piles in another corner. If he merges them how many piles will he have? *[correct answer 1 pile; intuitive answer 2 (9) piles]*	Oldrati (2016)	• Tricky • No intuition
6	You are participating in a run. You overtake the second runner in the last meters before the finish line. In what position did you finish? *[correct answer 2nd; intuitive answer 1st]*	Oldrati (2016)	• Potential
7	25 soldiers are standing in a row 3 m from each other. How long is the row? *[correct answer 72m; intuitive answer 75m]*	Oldrati (2016)	• In use
8	A snail starts climbing up a five-meter-high wall in the morning. During day, it climbs 2 m and during the night it slips back 1 m. How many days will it take the snail to reach the top of the wall? *[correct answer 4 days; intuitive answer 5 days]*	Oldrati (2016)	• Potential
9	A brick weighs 1 kg plus half a brick. How much does half a brick weigh? *[correct answer 1 kg; intuitive answer 5 (0.5) kg]*	Oldrati (2016)	• Potential
10	There are 5 white and 5 black socks in Franco's drawer. Franco's room is in the dark. How many socks should Franco take out of the drawer to be sure that he gets a matching pair? *[correct answer 3 socks; no intuitive answer]*	Oldrati (2016)	• Potential
11	You go to bed at eight. You set your old analogue alarm clock to wake you up at nine. How many hours of sleep will you get?	Oldrati (2016)	• Potential

	[correct answer 1h; intuitive answer 13h]		
12	One month has 28 days. How many of the 11 months left have 30 days? *[correct answer 11 months; intuitive answer 4 months]*	Oldrati (2016)	• Tricky
13	If three elves can wrap six toys in half an hour, how many elves are needed to wrap twenty toys in one hour?	Primi et al. (2016)	• Widget derivative • Tricky • No intuition
14	Ellen and Kim are running around a track. They run equally fast but Ellen started later. When Ellen has run 5 laps, Kim has run 15 laps. When Ellen has run 30 laps, how many has Kim run?	Primi et al. (2016)	• Tricky • No intuition
15	An ice cream vendor sells 2/3 of her stock of ice creams on sunny days and1/3 of her stock on cloudy days. Yesterday, it was a sunny day, and she sold300 ice creams. Today is a cloudy day. How many can she expect to sell?	Primi et al. (2016)	• Tricky • No intuition
16	In a class, there are 42 children. There are 12 more girls than boys. How many girls are there in the class?	Primi et al. (2016)	• No intuition
17	In an athletics team, tall members are three times more likely to win a medal than short members. This year, the team has won 60 medals so far. How many of these have been won by short athletes?	Primi et al. (2016)	• No intuition
18	If you are running a race and you pass the person in second place, what place are you in? *[correct answer 2nd; intuitive answer 1st]*	Thomson &Oppenheimer (2016)	• Potential
19	A farmer had 15 sheep and all but 8 died. How many are left? *[correct answer 8; Intuitive answer 7]*	Thomson &Oppenheimer (2016)	• Potential
20	Emily's father had three daughters. The first two are named April and May. What is the third daughter's name? *[correct answer Emily; intuitive answer June]*	Thomson &Oppenheimer (2016)	• Potential • Potential to be seen in social media
21	How many cubic feet of dirt are there in a hole that 3' deep x 3' wide x 3' long? *[correct answer none; intuitive answer 27]*	Thomson &Oppenheimer (2016)	• Complex/hard to understand
22	A cargo hold ship had 500 crates of oranges. At the ship's first stop, 100 crates were unloaded. At the second stop, 200 more were unloaded. How many crates or oranges were left after the second stop? *[correct answer 200 crates]*	Thomson &Oppenheimer (2016)	• Decoy/control items • No intuition
23	Sara, Emma, and Sophia embark on a river trip. Each of them brings one supply item for the trip: a kayak, a cooler of sandwiches, and a bag of apples. Sara brought the apples and Emma didn't bring anything edible. What did Sophia bring? *[correct answer cooler of sandwiches]*	Thomson &Oppenheimer (2016)	• Decoy/control items • No intuition
24	An expedition on a mountain climbing trip was traveling with eleven horse packs. Each horse can carry only three packs. How many horses does the expedition need? *[correct answer 4 horses]*	Thomson &Oppenheimer (2016)	• Decoy/control items • No intuition
25	A mechanic shop had five silver cars, two red cars, and one blue car in the garage. During the day, three silver cars and one red	Thomson &Oppenheimer	• Decoy/control items • No intuition

	car were picked up, and one black car was dropped off. How many silver cars were in the garage at the end of the day? *[correct answer two silver cars]*	(2016)	
26	If it takes 2 nurses 2 minutes to measure the blood pressure of 2 patients, how long would it take 200 nurses to measure the blood pressure of 200 patients?	Baron et al. (2015)	• Arithmetic items with lures • Widget derivative
27	Soup and salad cost $5.50 in total. The soup costs a dollar more than the salad. How much does the salad cost?	Baron et al. (2015)	• Bat-ball derivative • No intuitive
28	Sally is making sun tea. Every hour, the concentration of the tea doubles. IF it takes 6 hours for the tea to be ready, how long would it take for the tea to reach half of the final concentration? (Finucane & Gullion, 2010)	Baron et al. (2015)	• Lilypad derivative
29	A frog fell into a hole 30 meters deep. Every day it climbs up 3 m, but during the night it slides 2 m back down. How many days will it take the frog to climb out of the hole? *[correct answer 28 days; intuitive answer 30 days]*	Ackerman (2014)	• Translated from Hebrew • Potential
30	Apple mash is comprised of 99% water and 1% apple solids. I left 100 kg mash in the sun and some of the water evaporated. Now the water is 98% of the mash. What is the mash weight? *[correct answer 50; intuitive answer 99]*	Ackerman (2014)	• Translated from Hebrew • Found complex by pilot study
31	If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5% what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's signs or symptoms? *[correct answer 2; intuitive answer 50]*	Ackerman (2014)	• Translated from Hebrew • Complex • Long
32	Every day, a bakery sells 400 cookies. When the manager is not there, 20% of the cookies made that day are eaten by the staff. How many additional cookies should be made on the manager's day off to ensure that 400 cookies can be sold? *[correct answer 100; intuitive answer 80]*	Ackerman (2014)	• Translated from Hebrew • Found complex by pilot study
33	Steve was standing in a long line. To amuse himself he counted the people waiting, and saw that he stood 38th from the beginning and 56th from the end of the line. How many people are stood in the line? *[correct answer 93; intuitive answers 92 or 94]*	Ackerman (2014)	• Translated from Hebrew • No intuition • Pure numeracy
34	Ants are walking in a line. A bad-mannered ant cuts in front of the ant waling second. What is the rude ant's place in the line? *[correct answer 2nd; intuitive answer 1st]*	Ackerman (2014)	• Translated from Hebrew • Complex • Not Understandable
35	A house contains a living room and a den that are perfectly square. The living room has 4 times the square footage of the den. If the walls in the den are 10 feet long, how long are the walls in the living room? *[correct answer 20; intuitive answer 40]*	Shtulman & McCallum (2014)	• Potential
36	A storeowner reduced the price of a $100 pair of shoes by another 10 percent. How much do the shoes cost now? *[correct answer 81; intuitive answer 80]*	Shtulman & McCallum (2014)	• Found complex by pilot study

37	If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together? *[correct answer 4 days; intuitive answer 9]*	Toplak et al. (2014)	• Kindly supplied to us by Shane Frederick in personal correspondence • Predicitability verified • Hard to calculate
38	Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class? *[correct answer 29 students; intuitive answer 30]*	Toplak et al. (2014)	• Kindly supplied to us by Shane Frederick in personal correspondence • Predicitability verified • Taken
39	A man buys a pig for $60, sells it for $70, buys it back for $80, and sells it finally for $90. How much has he made? *[correct answer $20; intuitive answer $10]*	Toplak et al. (2014)	• Adapted from Dominowski (1994) (see Gilhooly & Murphy, 2005) • Predicitability verified • Taken
40	Simon decided to invest $8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had pur- chased were down 50%. Fortunately, for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has: a. broken even in the stock market, b. is ahead of where he began, c. has lost money *[correct answer c; intuitive answer b]*	Toplak et al. (2014)	• Created by the authors • Predicitability verified • Complex • Long
41	A Ferrari and a Ford together cost $190,000. The Ferrari costs $100,000 more than the Ford. How much does the Ford cost? *[correct answer $45,000; intuitive answer $90,000]*	Tremoliere & De Neys (2014)	• *Bat-ball derivative* • *Congruent and incongruent versions of the bat-and-ball*
42	A Rolls-Royce and a Ferrari together cost $190,000. The Rolls-Royce costs $100,000 more than the Ferrari. How much does the Ferrari cost? *[correct answer $45,000; no intuitive answer $90,000]*	Tremoliere & De Neys (2014)	• *Bat-ball derivative* • *Congruent and incongruent versions of the bat-and-ball* • *Still has the previous exposure via pilot study*
43	A magazine and a banana together cost $2.90. The magazine costs $2. How much does the banana cost? *[correct answer 90 cents; no intuitive answer]*	De Neys et al. (2013)	• *Control version of the bat-and-ball* • *No intuition*
44	Jack is looking at Anne but Anne is looking George. Jack is married but George is not. Is a married person looking at an unmarried person? a. Yes b. No c. Cannot be determined	Toplak & Stanovich (2002)	• No intuition • There are four more questions in the same style

B – Pilot Studies

"I have not failed. I've just found 10,000 ways that won't work."

THOMAS A. EDISON

This study has applied a staggered process of pilot studies to reach the most suitable version for its aims. At every iteration stage, it was checked for what is working and what is not; potential improvement points were identified; foreseen changes were applied and finally their impacts were measured. Feedbacks coming from colleagues, consulted field experts and participants have been major inputs for this study throughout all these steps.

B-1 The First Version: Beer Game Experiment

The experiment tool of this study was initiated with the beer game version. The version was also composed of the same stages as the final work applied in the main study. It started with the CRT part as is, followed by the demand and inventory management game.

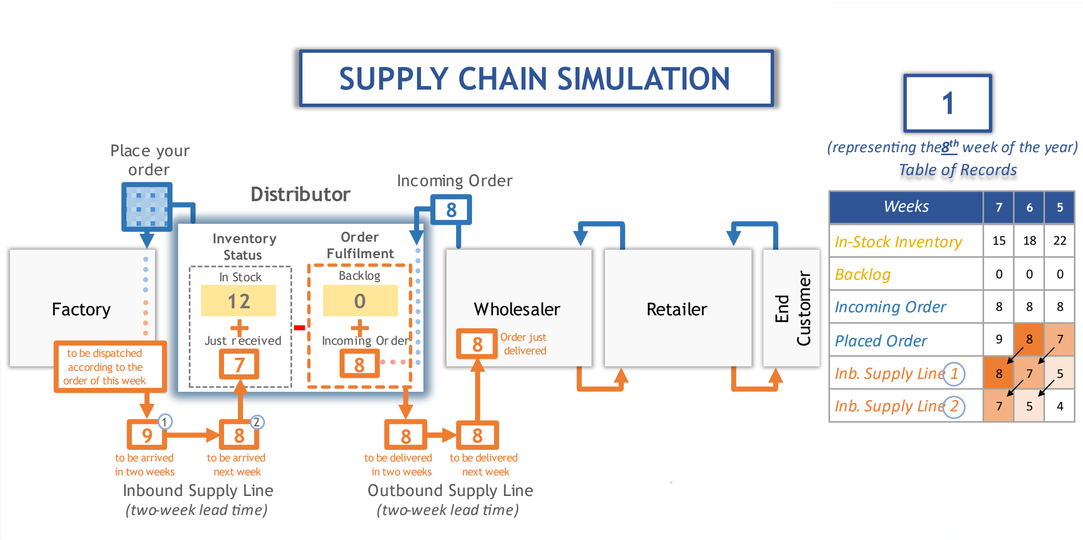
i. Creation story and main constructs of the experiment

In the first version of the study, CRT section was composed of the following 12 questions:

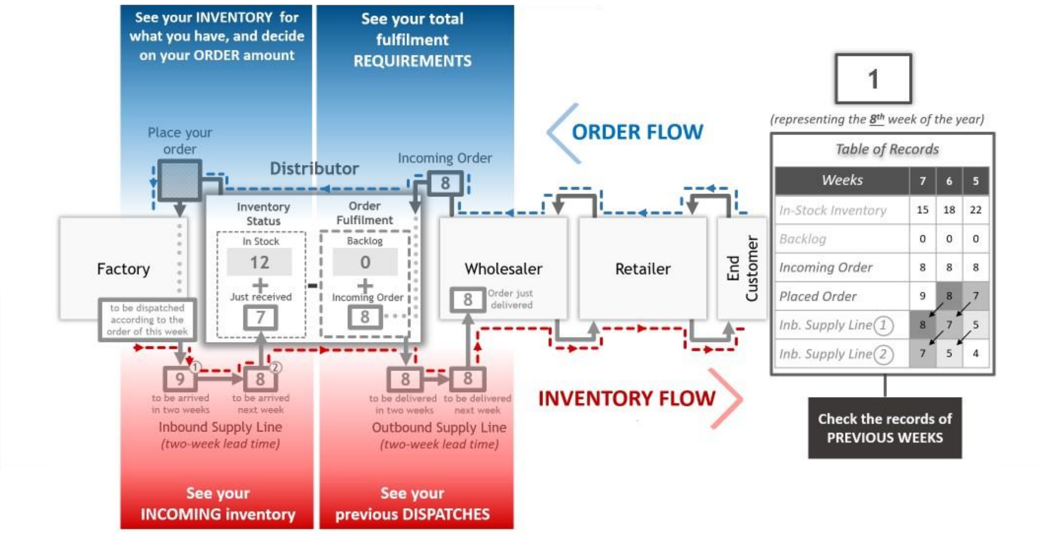
1. 25 soldiers are standing in a row 3 m from each other. How long is the row? (in meters)
2. Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class?
3. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
4. Apple mash is comprised of 99% water and 1% apple solids. I left 100 kg mash in the sun and some of the water evaporated. Now the water is 98% of the mash. What is the mash weight? (kg)
5. A Rolls-Royce and a Ferrari together cost Â£190,000. The Rolls-Royce costs Â£100,000 more than the Ferrari. How much does the Ferrari cost? (£)
6. 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)
7. A store owner reduced the price of a reduced $100 pair of shoes by another 10 percent. How much do the shoes cost now? ($)
8. A man buys a pig for $60, sells it for $70, buys it back for $80, and sells it finally for $90. How much has he made? ($)
9. A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost? ($)
10. Every day, a bakery sells 400 cookies. When the manager is not there, 20% of the cookies made that day are eaten by the staff. How many additional cookies should be made on the manager’s day off to ensure that 400 cookies can be sold?
11. 24-6x(24/6)=?
12. (13*10-13)/9-12=?

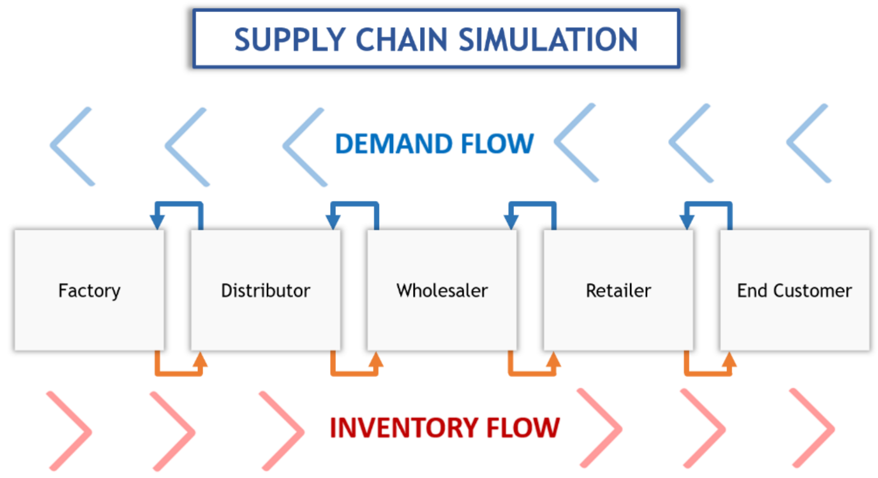
The source of questions varied. The list included the original CRT questions (3), some derivatives of the originals (1), other questions from the accumulated CRT questions pool (6) (*see* Appendix A, *p. 127*) and lastly two numeracy control questions.

In the second stage, the BDG version of the OMG was prepared. It comprised three sections. Firstly, the game was described using a detailed instruction which was supported by an animated slide set that is reached in another browser tab by clicking the link. Following figure demonstrate this slide set, and can be explanatory for the applied game in all treatment conditions:



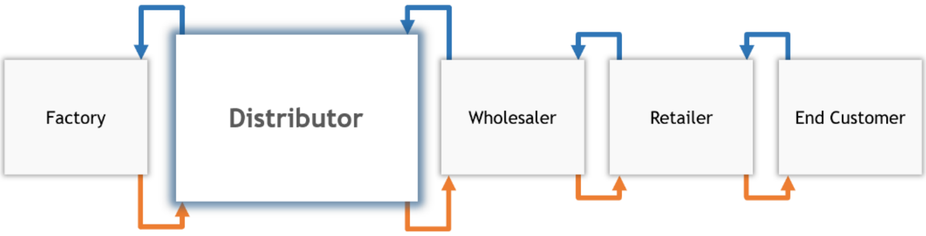
Key elements of the simulation:





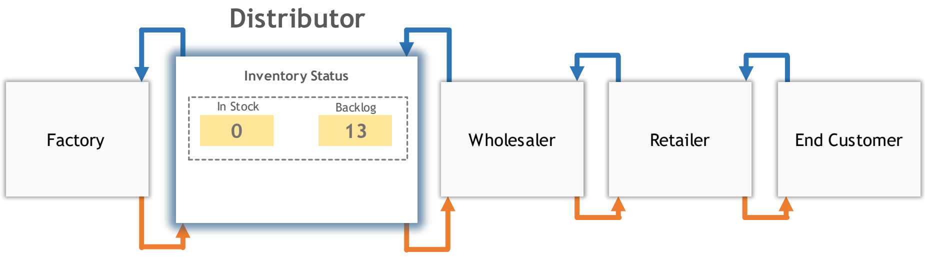
* Supply chains are composed of **upstream demand** and **downstream inventory flow**.

SUPPLY CHAIN SIMULATION



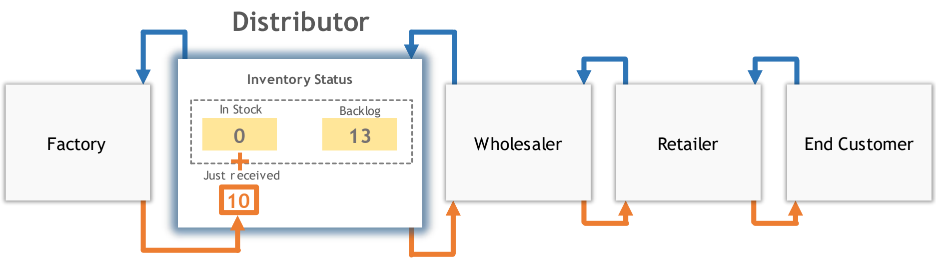
* You are assigned to **Distributor** role in supply chain.

SUPPLY CHAIN SIMULATION



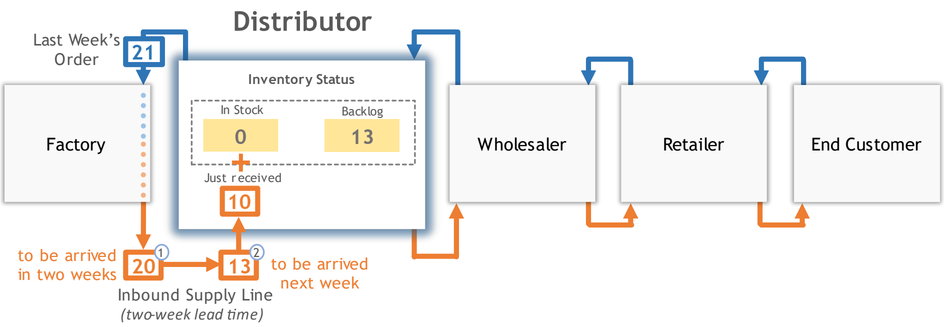
* You manage some **in-stock inventory.** Sometimes you run out of inventory and you miss some customer demand. That is recorded as ***'backlog'***. Unless you meet those demands, you keep having backlogs. Having backlog cost 4 times higher than keeping inventory(£0.50 vs £2.00 per unit).

SUPPLY CHAIN SIMULATION



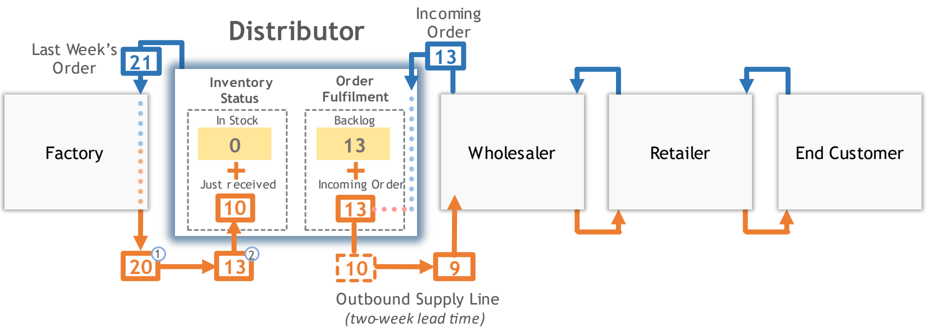
*** On top of your inventory,you get the incoming goods from your inbound supply line.**

SUPPLY CHAIN SIMULATION



*** There is a two-week lead time for delivery of goods. For example, you ordered 21 last week and Factory sent 20 units which will arrive in two weeks. The missing 1 unit shows that factory does not have enough inventory. You can track your incoming deliveries from the inbound supply line.①②**

SUPPLY CHAIN SIMULATION



*** As factory fulfilled your demand, you take the incoming order from your customer Wholesaler and fulfil it together with backlogs if there is any. For example, you need to send 26 units in total, you can only fulfil 10 customer demands with your inventory.**

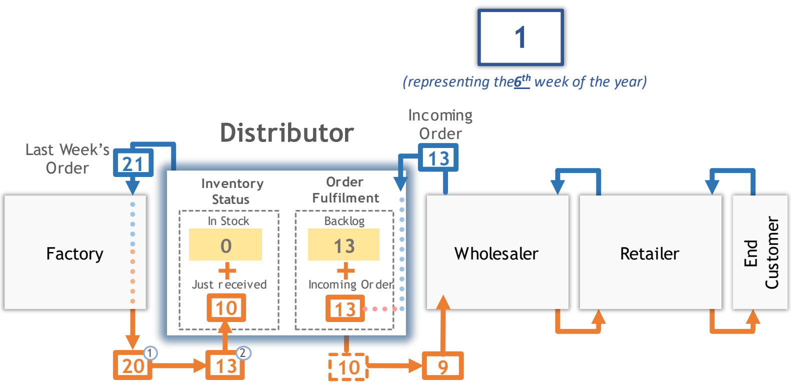


Table of Records

Weeks	5	4	3
Distributor			
In-Stock Inventory	0	0	0
Backlog	13	8	8
Incoming Order	14	12	13
Placed Order	21	18	18
Inb. Supply Line ①	13	10	9
Inb. Supply Line ②	10	9	12

*** You can also get insight from earlier weeks via Table of Records. For example, this question represents the 6th week of the year, and Table of Records provides information for the 5, 4 and 3rd weeks**

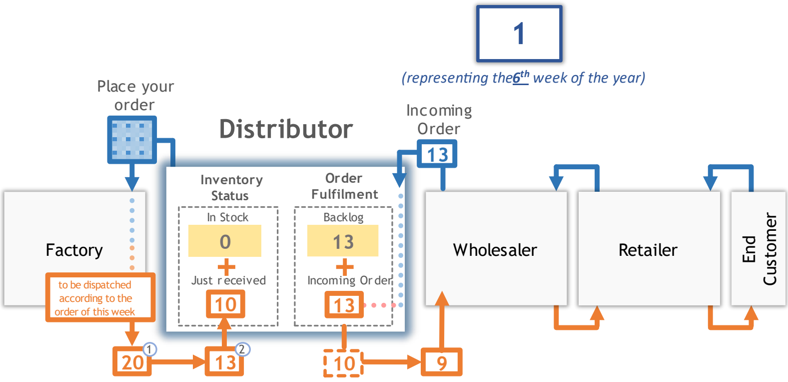


Table of Records

Weeks	5	4	3
Distributor			
In-Stock Inventory	0	0	0
Backlog	13	8	8
Incoming Order	14	12	13
Placed Order	21	18	18
Inb. Supply Line 1	13	10	9
Inb. Supply Line 2	10	9	12

*** Considering all these information, you should decide how much order you will place to the factory in this week**

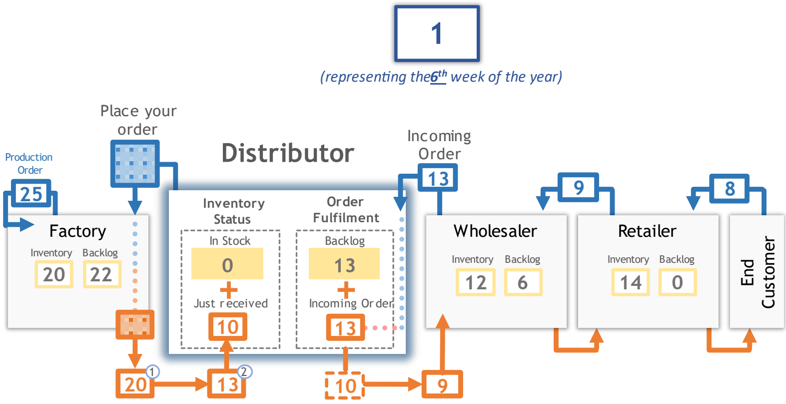


Table of Records

Weeks	5	4	3
Distributor			
My In-Stock Inventory	0	0	0
My Backlog	13	8	8
Incoming Order	14	12	13
Placed Order	21	18	18
Inb. Supply Line 1	13	10	9
Inb. Supply Line 2	10	9	12
Factory			
Inventory	13	10	9
Backlog	17	9	1
Production Order	27	25	20
Wholesaler			
Inventory	10	8	12
Backlog	6	4	6
Incoming Order	9	10	10
Retailer			
Inventory	14	13	8
Backlog	0	1	0
Incoming Order	8	8	8

When making your decision . . .

. . . You may access **further information** from your supply chain partners, which may help you improve your order placement decision.

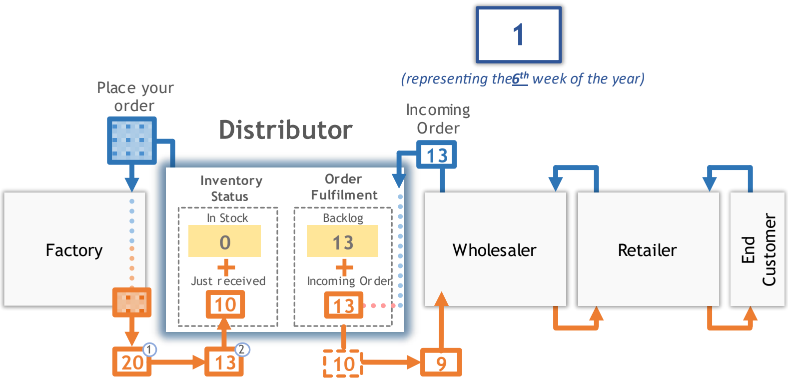


Table of Records

Weeks	5	4	3
Distributor			
In-Stock Inventory	0	0	0
Backlog	13	8	8
Incoming Order	14	12	13
Placed Order	21	18	18
Inb. Supply Line 1	13	10	9
Inb. Supply Line 2	10	9	12

When making your decision . . .

. . . You may have limited time to make your decision.

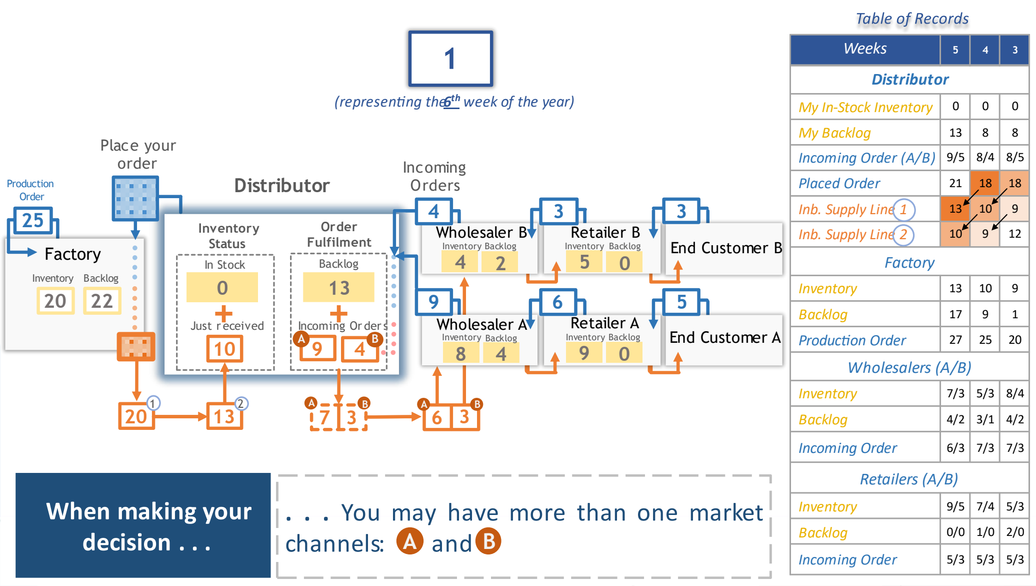


Figure A.1 Exemplary slide sets for the Beer Game version

This version of the study had also the same structure as the main study, however, this version of the experiment required a more detailed decision-making process which led to consider total demand fulfilment requirements, aimed and possessed inventory amount and inventory that had ordered earlier and in transport at that moment. Considering difficulty of comprehending this complexity in short time, each participant went through 5 practice questions that aimed to train participants. Practice questions were designed to give an overall idea about the potential scenarios related to varied demand and inventory situations. These scenarios were stable inventory and demand situation, stable inventory and increasing demand situation, insufficient inventory and high upcoming supply line inventory situation, excessive inventory and low demand situation, and finally stable demand and low upcoming supply line inventory situation.

Following the practice stage where participants had knowledge about the treatment conditions and inventory-demand scenarios that they would face, participants were randomly distributed into six treatment conditions. Actual experiment stage was composed of 24 questions of which the scenario distribution was as follows:

- 5 - Insufficient inventory (Backlog) – Low upcoming supply line inventory
- 5- Backlog – High upcoming supply line inventory
- 5- Stable inventory (Lean) – Low upcoming supply line inventory
- 4- Stable inventory (Lean)– High upcoming supply line inventory
- 5- Demand change (either to increasing or decreasing)

It is also important to mention that 5 of these 24 questions were control questions: 2 to measure participant comprehension by checking the reaction of participants to stable situations, another 2 to check the comprehension by checking the reaction of participants to excessive inventory situations and 1 attention control by asking questions with the same scenario at the beginning and end. To ensure the participant attention, a monetary incentive was also utilised. Top 5% performers of all treatment conditions were offered to earn £5 in addition to their participation fee.

In addition to two main stages, additional questions related to the previous exposure to CRT and any SC simulation was asked. Likewise, feedback from the participants were demanded in relation to the difficulty, clarity and length of the experiment.

ii. Pilot study targets

After preparing the version mentioned above, it was required to control its applicability via a pilot study. Required control points were categorised into 5 groups as follows:

1. Timing
    - Timing of CRTs
    - Timing of simulation
    - Timing of the whole experiment
    - Time related payment amount
    - Time required for the time pressure treatment
2. Instructions
    - Clarity of the instructions
    - Utilisation of the additional slide set
3. Manipulation
    - Applicability and comprehension of SC simulation
    - Measurement capability of the questions
    - Required question number for CRT and SC simulation
4. Control questions
    - Operability of control questions for CRT and SC simulation
5. Operability of practices

iii. Pilot study analysis and implications

The pilot experiment was distributed among the circle of the researcher, including colleagues from academia and industry, previous MSc classmates, MSc students from previous years, researchers met at various academic occasions over PhD years. In total, 237 people were sent the participation link. While only 85 (36%) joined and started the experiment, only 31 of this 85 (37%) completed it. The graph of drop can be seen in the following figure:

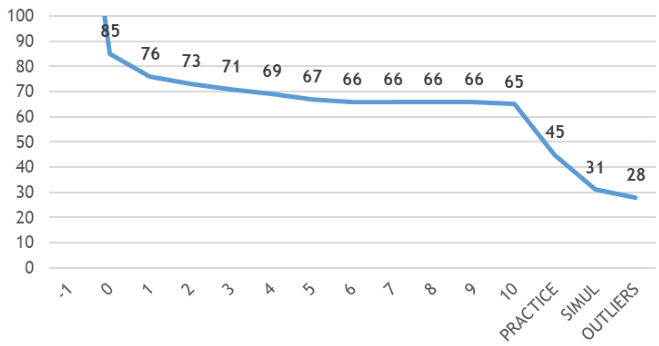


Figure A.2 Drop rate throughout the experiment

The first analysis was made regarding the feedback questions where participants evaluated the difficulty, clarity and length of the tool overall and per stage over a 1(worst)-10(best) scale. The average results for five feedback questions ranged from 4.65 to 5.36. In addition, some participant also provided qualitative feedbacks mentioning various concerns (Table A.1)

Table A.1 Feedback categories

Clarity	• Suppose more info would make tasks more clear • Overall this simulation seemed to be too complex and the unrelated nature of the questions made it hard for the participant to improve their performance … • Overall, I think that the simulation is overly complicated and there need to be a greater level of explanation with a worked example so that the person completing it fully understands the requirements and feels confident to complete the task… Without this, the tendency is to guess… there is nothing to allow us to judge..
Length	• The time required for the entire survey is more than the 30 minutes stated in the introduction • Not enough time to complete the simulation • I feel that overall, it takes too long to complete the survey and the simulation. Thus, you might find it difficult to find participants of your survey.
Tools	• Also, respondents need to be told that the work requires some little calculations. Thus, pen and paper should be handy before starting the survey.
Interface	• Text in the figure was small and would be easier to read if it was bigger. • Picture of simulation should be with zoom in feature.
Live feedback	• …learn from their previous efforts…. • I would also like to have some sort of feedback on my performance.
Incentive	• Personally don't offer me Â£5… Either nothing or Â£20…
Endogeneity	• Also, think about the endogeneity issues associated with it.

Therefore, even without starting a detailed analysis, from the drop rate and feedbacks, it was inferred that the tool must be enhanced to be simpler and shorter to apply the required constructs of this study successfully and realise its targets.

Next, operability of CRT and SC simulation sections were examined respectively. There were 12 CRT questions (including the 2 numeracy control questions) in total. The main of reason for employing high number of questions was to determine and select the most suitable questions to utilise in an online experiment amongst them. Therefore, a relatively high number of questions were selected for the pilot process in comparison to the previous CRT applications and item number recommendations. CRT questions were controlled four aspects: capability to measure the rationality and intuition in the right proportion (approximately 50-50%), previous exposure, time spent on the question and numeracy control.

Starting with the first aspect, it was seen that only 1 question (8th) was close to the target distribution of correct (C)-intuitive incorrect/wrong (IW) answers, while 4th was too hard to solve (considering the high number of non-intuitive incorrect/wrong [NW] answers), 5th and 6th were not tricky enough (low number of intuitive answers) and 10th was too tricky (high intuitive answer amount).

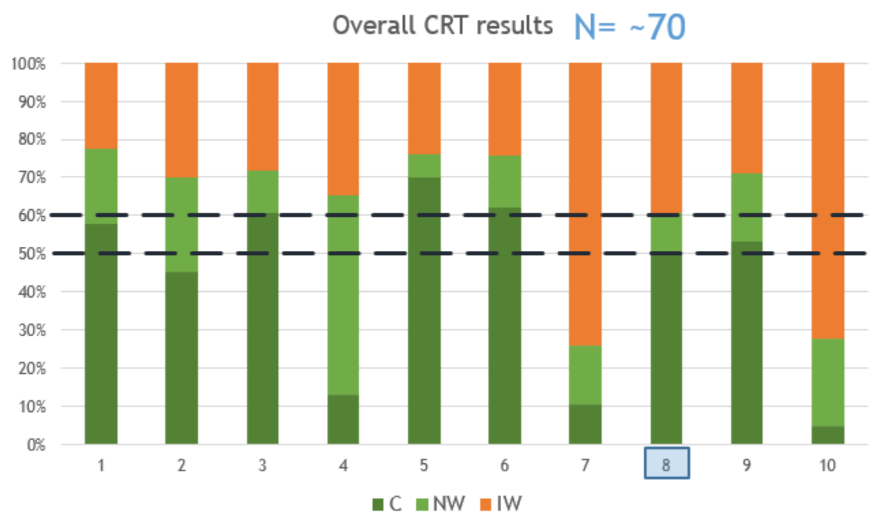


Figure A.3 CRT response ratio analysis

Secondly, the impact of previous exposure on the questions was checked. It was seen that original CRT questions (3, 6, 9) and the only original derivative (5) were seen by nearly half of the participants. High proportion of correct answers was also observed for these questions (Figure A.4).

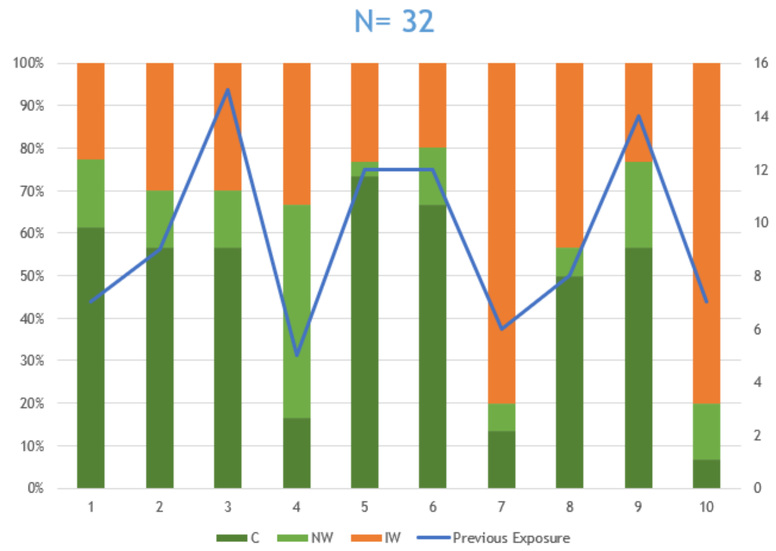


Figure A.4 CRT response ratio – previous exposure interaction

Next, the time that participants spent on questions were analysed to see the too easy and too hard questions (Figure A.5). It was observed that ratio of participants who answer questions 1, 2 and 9 correctly is high in comparison to the average time spent on those questions. On the other side, participants spent nearly average time on questions 7 and 10 but the correct answer ratios of those questions were still too low. This means that the first group of questions was perceived easy to solve, while the second group of question was hard even though a sufficient time was spent on them.

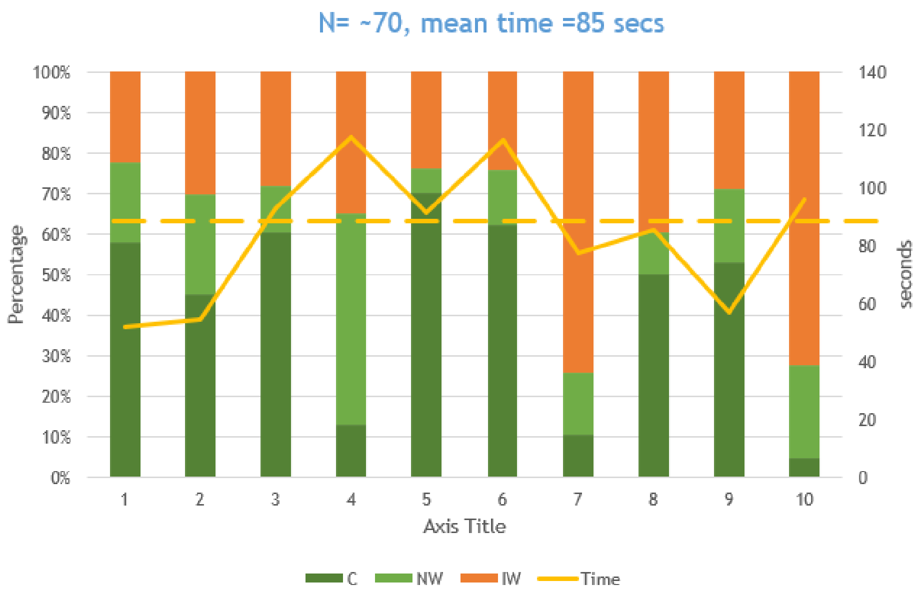
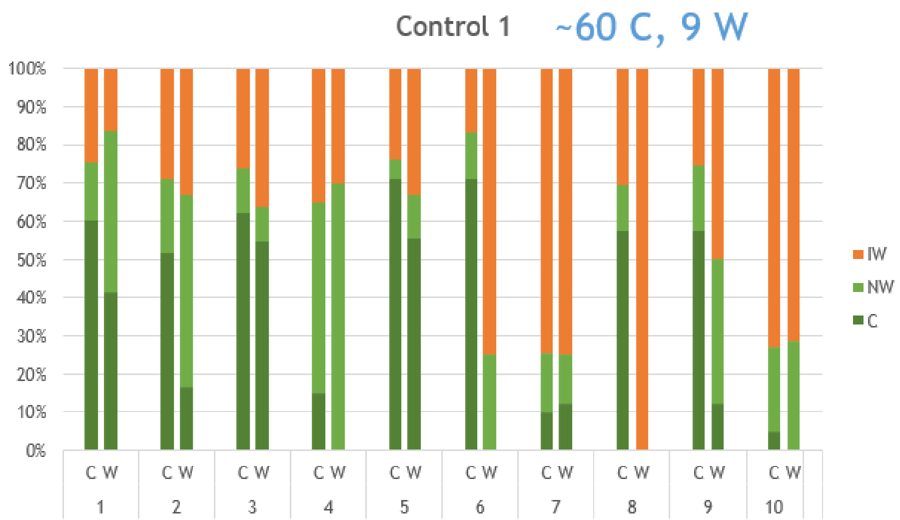


Figure A.5 CRT response ratio – answer time interaction

Lastly, the interaction between participants' CRT performance and numeracy performance via checking their answers in the numeracy control questions. The aim of this control was to determine CRT questions that require least numeracy capability. To understand the impact, CRT response distribution was compared per question between the participants who had given correct and wrong answers to the numeracy control questions. Results have shown that in questions 1, 3, and 5, similar response proportions are produced by both participant groups who reply the numeracy control questions right or wrong. In line with this process, it was seen that while questions 2, 6, 8 and 9 require higher numeracy, questions 4, 7 and 10 were identified as difficult even for high numeracy participants.

In sum, CRT questions were evaluated via this 4-step elimination process (Table A.2) and only 4 of them were selected to be applied in the main study, which are 1, 2, 3 and 8. However, 3 was also eliminated as being one of the original CRT questions. However, recommendations in the literature were suggesting a CRT item number between 5 and 7. Therefore, the remaining three questions were added another two from the CRT questions pool (Appendix A, *p. 127*), which shows the similar characteristics, and the final set of CRT questions were completed to be used in this study.



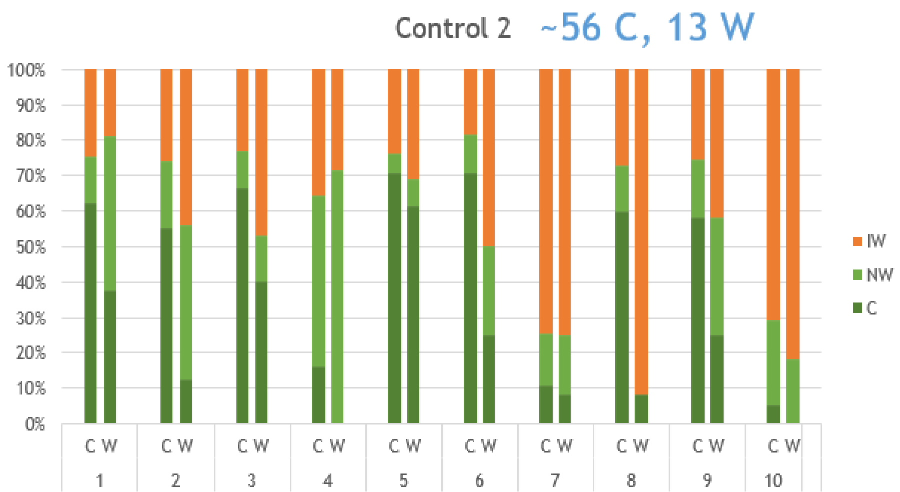


Figure A.6 CRT response ratio – numeracy control interaction

Table A.2 Summary of CRT evaluation

	Proportion	Previous Exposure	Timing	Numeracy
1) **25 soldiers** are standing in a row 3 m from each other. How long is the row? (in meters)	~	+		+
2) Jerry received both the **15th highest and the 15th lowest** mark in the class. How many students are in the class?	~	+		~
3) If it takes 5 machines 5 minutes to make **5 widgets**, how long would it take 100 machines to make 100 widgets?	~	~	+	+
4) **Apple mash** is comprised of 99% water and 1% apple solids. I left 100 kg mash in the sun and some of the water evaporated. Now the water is 98% of the mash. What is the mash weight? (kg)	-		+	-
5) A **Rolls-Royce and a Ferrari** together cost Â£190,000. The Rolls-Royce costs Â£100,000 more than the Ferrari. How much does the Ferrari cost? (£)	-	-	+	+
6) 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)	-	-	+	~
7) A store owner **reduced the price** of a reduced $100 pair of shoes by another 10 percent. How much do the shoes cost now? ($)	-			-
8) A man **buys a pig** for $60, sells it for $70, buys it back for $80, and sells it finally for $90. How much has he made? ($)	+	+	+	~
9) A **bat and a ball** cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost? ($)	~	~		~
10) Every day, a **bakery** sells 400 cookies. When the manager is not there, 20% of the cookies made that day are eaten by the staff. How many additional cookies should be made on the manager's day off to ensure that 400 cookies can be sold?	-			-

Next, operability of the SC simulation was examined from two major aspects: participants' comprehension via control questions and drop rate. Additionally, efficiency of practice stage and the total time required for the test were also considered via pilot study.

As explained, this pilot study employed three types of control questions. The first two types were to measure comprehension: two to measure comprehension of stable demand and inventory situation and two for the excessive inventory situation. These situations were selected as they were the easier and most straightforward scenarios that participants could face. In a stable situation, they receive stable amount of demand from the customer and inventory from the supplier. Therefore, they are expected to maintain this stability without any overreaction. In the 'excessive' scenario, however, customer places a low amount of order, but participants hold much more than they need. Therefore, they are expected to place 0 orders to finish the excessive inventory that they had already had. The results have shown a low performance for these 4 control questions. Except 1 'excessive' situation question, overall half the participants could not provide the expected order responses. The last control question was related to the attention of participants. It was set by a question twice, one at the beginning and one at the end, to observe if participants pay the required attention and analyse the question in the same way on both occasions. It was also considered that there may be some amount of variation between the answers as participants are not expected to calculate each decision literally. It was observed that answers for both questions were in the predefined 10% variation interval, which means that all participants completing the study showed the required effort. However, the problem seen in the comprehension questions was raised a critical question about the applicability of the tool, as comprehension was an essential prerequisite for this study.

Ideally, participants are targeted to feel under the effect of bullwhip where they would require analysing all the information they are provided and decide on how much to order today. However, in practice, it was seen that they cannot understand the game thoroughly unless they have previous knowledge or experience in the beer game. Since the game was not reactive and successive, they could not feel the impact of their inventory decisions, and in return, it did not allow them to feel the bullwhip. They tended to give random orders, which in the end, avoided this study producing robust and logical analysis.

Next, the analysis of collected data was conducted separately for each scenario. As mentioned, the SC simulation part of the experiment was composed of five scenarios regarding the interaction of demand and inventory situations. The reason for creating these scenarios was their effect on participants' order placement decisions. For instance, literature expresses that a manager's reaction to the deficiency of inventory is different from that of excessive inventory. Likewise, the reaction towards an increasing customer demand would be different from the decreasing or steady demand trends. Therefore, to reveal these differences, the analysis was targeted to be conducted separately for each scenario.

The analysis method defined was composed of three stages. Firstly, non-linear least squares technique was decided to be applied via Solver tool in Microsoft Excel. The technique aims to predict unknown parameters of an equation by observing the known variables. As the number of unknown parameters increase, possibility of receiving different solution combinations also increase. Secondly, potential solutions were compared via Genetic Algorithms methods and ideal results for the unknown parameters are determined. Lastly, these defined parameters were compared between the manipulated treatments via Mann-Whitney U-Test.

However, it was seen that to lower the number of potential solutions and have more reliable results, higher number of observations are required per participant. While the length of the version was already a concern as learned from the qualitative and quantitative feedbacks, that version was decided as difficult to continue.

Following the inferences and concerns from the first pilot study, substantial amendments were done in the design of the experiment, which produced the version applied in the main study. In the design of the SC simulation three changes were made. Participants were expected to analyse only demand related information. Reducing amount of information made the interface easier to understand and follow. In relation to that, practice stage was cancelled. The interactivity of the game was enhanced, so that participants could see their performance and improve accordingly.

B-2 Second Pilot Study and the Verification of the Amendments

The new version, which was employed in the main study, was designed to be shorter in time, clearer and more comprehendible in the online application with 5 CRT questions and 15-period OMG. The pilot test of this version was made through a partial participant recruitment over Prolific platform. 5 participants were taken for each treatment condition (30 in total) and the following points were checked:

- Total time spent
- Required time for the time pressure treatment
- CRT manipulation power
- Previous exposure of CRT questions
- Successful completion rate for the OMG

The results have shown the average time participants spent to complete the whole experiment was 12.35 minutes. To provide the sufficient time for all participants and to determine the participation fee, 15 minutes was determined as the approximate completion time and expressed in the initial description of the experiment.

Initially, participants in the time pressure treatment were provided 40 seconds per questions. But it was observed that the average completion time per question was much lower than that with 18 seconds. Therefore, the amount was decided to decrease to 15 seconds to make participants feel the impact of pressure. However, considering that they may need longer time in the initial periods to understand the game, they were provided with 20 seconds initially and it decreased gradually throughout the game. Likewise, at the final periods of the game, participants have higher amount of information cumulated over the previous periods. Therefore, the gradual decrease time was increased back up to 18 seconds by the end of the game.

Next, the manipulation power of CRT section was controlled, and it was seen that 51% of all answers were given intuitive incorrect answers. This meant that the selected questions could provide the targeted 50-50% separation of cognitive propensity groups.

In terms of the previous exposure control, results have shown that only 26% of the questions were previously seen by the participants. This ratio was nearly 50% for the original CRT questions, even though the participants in the first pilot study were not recruited via an online recruitment platform.

Lastly, participants' performance in the OMG was controlled. While 87% of participants successfully completed the experiment, 73% performs higher than the incentive threshold level of £1500. These results were then considered as acceptable to continue.

C – Main Experiment Appendices

C-1 Experiment File

The print view of the experiment tool can be reached via the following link and QR code:

https://drive.google.com/file/d/1uKhmVHzkCo3zJVbWoeX-v4rijVLhdTz-/view?usp=sharing

Moreover, the experiment tool can be tried using the following link:

https://warwickwmg.eu.qualtrics.com/jfe/form/SV_eJpEbV3RsRmpKnz

C-2 Instructions

PROLIFIC:

"Dear participant,

It is a simulation of supply chain dynamics related to different demand and supply conditions. It has two main parts:

1. Preparatory brainstorming questions: You will be asked some questions that aim to prepare you for the simulation by increasing your cognitive alertness.
2. Supply chain simulation: You will participate in a newspaper supply chain simulation, where you will make ordering decisions according to the demand you expect.

Depending on your simulation performance, you may be eligible to join a raffle of £100 (details provided in the study).

In total, it should take approximately 15 minutes to complete the survey."

MAIN:

"Dear participant,

As described, the study has two main parts: 1) Preparatory brainstorming questions, and 2) Supply chain simulation

Required information will be provided at the beginning of both parts. For further information about the study, please read the Participant Information Leaflet.

Thanks for participating. Please proceed to start the questionnaire.

(NOTE: Considering that some visuals in this simulation are detailed, we recommend you using devices with large screen sizes such as computers or tablets, rather than mobile phones.)"

CRT:

"In this stage, you will be asked 5 open-ended questions and you will have 5-7 minutes in total to complete the test. Please only write numbers in your answers, without adding any unit. Good luck!"

OMG:

This instruction was given to the participants who had taken the baseline-low information treatment condition. Other treatment conditions also provided additional information where required and they can be seen in the experiment file provided in the previous appendix

"Thanks for completing the first part! Now you will take a supply chain simulation. This part will last for another 8-10 minutes.

- In a newspaper supply chain, you are assigned the role of NEWSPAPER DISTRIBUTOR. You are expected to meet the demands of your direct customer "wholesaler". To do so, you will purchase newspapers from the "newspaper printing house" by placing purchase orders (see the figure below).

- You will pay SIM£2 (simulation pounds) per newspaper when buying, and sell each at SIM£5. If you buy less newspaper than your customer demands, you will miss the opportunity of making more profit. If you order more than customer demand, at the end of the period, the excess amount of newspapers will be returned to the printing house at a salvage price of SIM£1.

- The simulation will be 15 periods long. YOUR AIM is to make the highest profit possible each period. If you exceed SIM£1500 cap at the end of the simulation, you will be eligible to join the £100 REWARD raffle. After each period, you will see a feedback screen showing the incoming demand and profit related information. So that you will be able to adjust your orders accordingly and follow your profit status.

If you provide random, inattentive answers and end up with negative profit value, your participation will not be accepted in the Prolific system and will not be paid the participation fee of £1.65.

- To help you decide on your order amount, here is some information about your customer demand:

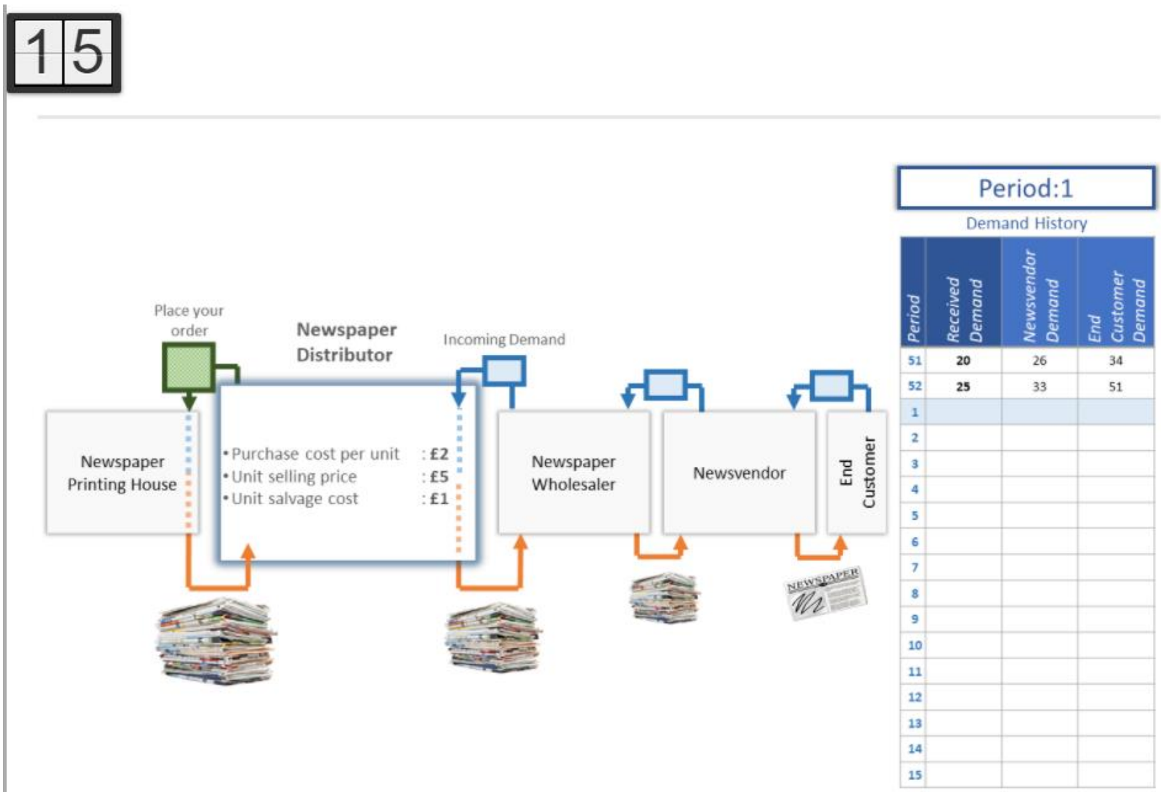
- In the last two periods, you received 20 and 25 units of demands respectively.

- From last year's data, in these 15 periods, you can expect that incoming demand will have an average of 42 and can vary by 18 units on average.

Please proceed to join the simulation."C-3 Selected Examples from Experiment Screenshots

Following two screenshots provide the question and feedback visuals for the OMG, for the first period of time pressure - high information treatment condition. Visuals for other treatment conditions and later periods can also be found in the experiment file provided in the previous appendix.

<u>Question Screen:</u>



Considering all the cost/price and demand related information presented, please write how much order you would place to your newspaper printing house for this period:

Feedback Screen:

- In Period 1, **you ordered newspapers** and **you received** a demand of **32**.
- By ordering that amount, you could not make any profit and lost the opportunity of earning SIM£96. By analysing demand, you can reach better results..
- By being late to make your decision, you lost SIM£10. The penalties will be deducted from your final profit at the end of the simulation.
- Your **accumulated profit** so far is SIM£**0**. You need to earn SIM£**1450 more** to be eligible to join the £100 raffle in the remaining 14 weeks.

Demand and Order History

Period	Incoming Demand	Your Order	Profit (SIM£)	Missed earning
51	20	-	-	-
52	25	-	-	-
1	32		0	96

Mean	25.67
St.Dev.	6.03

Total Profit (SIM£)
0

C-4 Codes Utilised in the Experiment

Codes embedded in the Question screen:

```
Qualtrics.SurveyEngine.addOnload(function()
{
	/*Place your JavaScript here to run when the page loads*/

});

Qualtrics.SurveyEngine.addOnReady(function()
{
	/*Place your JavaScript here to run when the page is fully displayed*/

});

Qualtrics.SurveyEngine.addOnUnload(function()
{
	/*Place your JavaScript here to run when the page is unloaded*/

});

Qualtrics.SurveyEngine.addOnPageSubmit(function()
{
	var actualDe = [32,47,65,67,62,57,35,7,13,30,47,50,45,40,40];
	var quIds = ["QID28"];
	var quId = this.questionId;
	//window.alert(quId);
	var i = quIds.indexOf(quId);
	var order = this.getChoiceAnswerValue();
	if (order <= actualDe[i])
		{var currentPr =  order * 3
		}
```

```
			else
			{currentPr = (actualDe[i] * 5 - order * 2 + (order-actualDe[i]))
			}
var profs = [];
profs[i] = currentPr;
Qualtrics.SurveyEngine.setEmbeddedData('Pr1', profs[0]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr2', profs[1]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr3', profs[2]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr4', profs[3]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr5', profs[4]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr6', profs[5]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr7', profs[6]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr8', profs[7]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr9', profs[8]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr10', profs[9]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr11', profs[10]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr12', profs[11]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr13', profs[12]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr14', profs[13]);
Qualtrics.SurveyEngine.setEmbeddedData('Pr15', profs[14]);

var totalPr = Qualtrics.SurveyEngine.getEmbeddedData('TotalPr');
var newTotal = (Number(currentPr) + Number(totalPr));

/* Assign response to ED */
/* The following line creates (if it doesn't exist) or set (if already exist) an embedded data value */
Qualtrics.SurveyEngine.setEmbeddedData('Order', order) ;
Qualtrics.SurveyEngine.setEmbeddedData('CuPr', currentPr);
Qualtrics.SurveyEngine.setEmbeddedData('TotalPr', newTotal);
var demand = actualDe[i];
Qualtrics.SurveyEngine.setEmbeddedData('Demand', demand);
```

```
if (currentPr == 0)

{var lost = Number(demand) * 3;

 var PrCa1 = " you could not make any profit and lost the opportunity of earning SIM£" + lost + ". By analysing demand, you can reach better results";

 Qualtrics.SurveyEngine.setEmbeddedData('Missing', PrCa1)

}

else if (currentPr < 0)

{lost = Number(demand) * 3;

 var loss = Number(currentPr) * (-1);

 var PrCa2 = " you could not make any profit, instead you lost SIM£" + loss + ". If you had ordered in the right amount, you could have earned SIM£" + lost + ". By analysing demand, you can reach better results" ;

 Qualtrics.SurveyEngine.setEmbeddedData('Missing', PrCa2)

}


else if(order > actualDe[i])

			{

			 lost= Number(order) - Number(demand);

			 var miss1 = " you have made SIM£" + currentPr + " profit. However, ordering more than the received demand, you missed the opportunity of earning SIM£" + lost + " more";

			 Qualtrics.SurveyEngine.setEmbeddedData('Missing', miss1)

			}

			else if (order < actualDe[i])

			{lost= Number(demand) * 3 - Number(currentPr);

			 var miss2 = " you have made SIM£" + currentPr + " profit. However, ordering less than the received demand, you missed the opportunity of earning SIM£" + lost + " more";

			 Qualtrics.SurveyEngine.setEmbeddedData('Missing', miss2)

			}

			else

			{var miss3 = " you have made the highest possible profit, SIM£" + currentPr;

			 Qualtrics.SurveyEngine.setEmbeddedData('Missing', miss3)

			}
```

```
});
```

Utilising the variables in the code, participants earlier performance interactively reflected to the feedback screen as follows:

- In this period, you ordered ${e://Field/Order} newspapers and you received a demand of ${e://Field/Demand}.
- By ordering that amount, ${e://Field/Missing}.
- Your accumulated profit so far is SIM£${e://Field/TotalPr}. You need to earn SIM£$e{1500-e://Field/TotalPr} more to be eligible to join the £100 raffle in the remaining 14 weeks.

Demand and Order History

Period	Incoming Demand	Your Order	Profit (SIM£)	Missed earning
51	20	-	-	-
52	25	-	-	-
1	32	${e://Field/Order}	${e://Field/Pr1}	$e{32 * 3 - e://Field/Pr1}

Mean	25.67		Total Profit (SIM£)	
St.Dev.	6.03		${e://Field/TotalPr}	

Moreover, in the time pressure treatment conditions, to measure the time spent and its impacts on the participant performance, additional codes were added to a hidden empty question as follows:

```
Qualtrics.SurveyEngine.addOnload(function()
{
	/*Place your JavaScript here to run when the page loads */

	//jQuery("#"+this.questionId).hide();
   //this.clickNextButton();

	var subTime = Qualtrics.SurveyEngine.getEmbeddedData('TimeDiff');

	if (subTime > 20000)
	{var penalty = " - By being late to make your decision, you lost SIM£10. The penalties will be deducted from your final profit at the end of the simulation.";
	 Qualtrics.SurveyEngine.setEmbeddedData('Penalty', penalty);
```

```
 var penNum = Qualtrics.SurveyEngine.getEmbeddedData('PenNum');
 var penNum = Number(penNum) +10;
 Qualtrics.SurveyEngine.setEmbeddedData('PenNum', penNum)
}
else
{penalty = "";
 Qualtrics.SurveyEngine.setEmbeddedData('Penalty', penalty)
}


});
Qualtrics.SurveyEngine.addOnReady(function()
{
	/*Place your JavaScript here to run when the page is fully displayed*/
});
Qualtrics.SurveyEngine.addOnUnload(function()
{
	/*Place your JavaScript here to run when the page is unloaded*/
});
```

C-5 Raffle Information Message and Raffle Video

Hello everyone, Firstly I want to thank to all joining my study. I have finished the most part of my analysis and eventually drawn the raffle among 498 eligible cases out of 623 participants in total. The ID of the participant is 599b27be903924000143258b and I sent the bonus payment. Hope to work together in another study and many thanks again. Note: I drew the raffle with Microsoft Excel and recorded its process, you can watch it in the video with the following link:

https://tinyurl.com/ybupmpxf

D – Ethical Appendices

D-1 Participant Information Leaflet (PIL)

Participant Information Leaflet can be reached via the following link and QR code:

https://drive.google.com/file/d/1b3NhyxEdvW_nOxPxNrfSbw08vwspeyuN/view

D-2 Ethical Approval Form

Full ethical approval of BSREC with the reference number BSREC 25/19-20 (Amendment 02) can be reached via the following link and QR code:

https://drive.google.com/file/d/1cKrtFAIY4RC5XGn3hAQdoWOj1XovqqTR/view?usp=sharing

E – Additional Analysis Appendices

E-1 CRT Previous Exposure Analysis

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Seen_count[b]	.	Enter

a. Dependent Variable: #2 (cor)

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std
1	.084[a]	.007	.005	

a. Predictors: (Constant), Seen_count

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8.354	1	8.354	4.421	.036[b]
	Residual	1173.485	621	1.890		
	Total	1181.839	622			

a. Dependent Variable: #2 (cor)

b. Predictors: (Constant), Seen_count

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	1.637	.067		24.304	.000
	Seen_count	-.055	.026	-.084	-2.103	.036

a. Dependent Variable: #2 (cor)

E-2 OMG Learning Curve Analysis

t-test for Equality of Means

	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	95% Confidence Interval of the Difference Upper
LC1.1	-1.993	66	0.05	-3.78819	1.90093	-7.58352	0.00714
LC1.2	-2.84	109	0.005	-4.03704	1.42155	-6.85449	-1.21958
LC2.1	-0.338	89	0.736	-0.48986	1.4501	-3.37117	2.39146
LC2.2	-1.458	92	0.148	-2.63587	1.80794	-6.2266	0.95486
LC3.1	-1.09	118	0.278	-1.60289	1.46999	-4.51387	1.30809
LC3.2	-3.352	104	0.001	-6.58754	1.9652	-10.4846	-2.69048
LC4.1	-1.094	118	0.276	-1.58544	1.44898	-4.45481	1.28393
LC4.2	-2.096	99	0.039	-3.67255	1.75248	-7.14984	-0.19525
LC5.1	-1.4	83	0.165	-2.28682	1.63378	-5.53635	0.96271
LC5.2	-3.33	91	0.001	-6.66096	2.00001	-10.6337	-2.68818
LC6.1	-1.6	91	0.113	-4.42877	2.76721	-9.92549	1.06795
LC6.2	-0.498	81	0.62	-1.24332	2.4951	-6.20778	3.72114