Exploring the Contribution of Individual Differences and Planning Policy Parameters to Demand Planning Performance

by

Alexander Alexandrovitch Kharlamov

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

University of Warwick, WMG

October 2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Thank you single life, caffeine, Google, Wikipedia and my supervisors, Jan and Ganna for making this possible.

In addition, all the exquisite ladies (who shall remain anonymous) contributing to my artistic endeavours and keeping my mind away from work.

DECLARATION

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

ABSTRACT

Demand planning (DP) is important for business performance. DP depends both on managers and on supporting systems. Managers are known to increase uncertainty by systematically overriding the systems and making unnecessary judgemental adjustments. This is a behavioural problem. Systems are assumed to be represented by different policies and individual differences by measurable traits and characteristics. The contribution of individual differences and policy parameters to DP performance is not clear.

A framework is proposed based on the cumulative prospect theory (CPT) and myopic loss aversion (MLA). Methodology of decision making experiment based on the newsvendor is used. Individual differences are collected using previously validated psychometric scales and demographic questions. The sample (N=339) includes three main groups: professional planners (N=84), naïve students (N=166), logistics and supply chain management (L&SCM) students (N=56).

The MLA hypothesis is supported. Longer planning horizons (less frequent decisions) outperforms short planning horizons. Regarding individual differences, only experience/knowledge and naïve interventionism are significant predictors of DP performance. L&SCM students with theoretical knowledge but without practical experience perform the best. No significant difference in performance is found between professional planners and naïve students. Naïve interventionism (plan instability) contributes negatively to DP performance. Personality (Big Five), impulsiveness, propensity to plan, decision-making style or demographics (e.g. age, sex, and years of experience or managerial level) are not significant for DP performance.

The view that there is a 'right' mind-set (personality) to be a good planner is challenged. DP policy can offset individual differences. A MLA informed policy can reduce uncertainty introduced by behaviour. System restrictiveness (binding policy for long commitment) outperforms decisional guidance (non-binding policy for optional commitment). This is one of the first applications of CPT and MLA to DP decisions.

ABBREVIATIONS

ACT	Adaptive Control of Thought theory
AI	Artificial Intelligence
APS	Advanced Planning Systems
BIS	Barratt Impulsiveness Scale
BOR	Behavioural Operations Research
CPT	Cumulative Prospect Theory
DM	Decision Making
DP	Demand Planning
DS	Decision Support
DSS	Decision Support Systems
EPM	Exposure-Performance Matrix
EPO	Elaboration on Potential Outcomes
ERP	Enterprise Resource Planning
ETO	Engineer-to-Order
EUT	Expected Utility Theory
EV	Expected Value theory
GBP	Great Britain Pound
GDMS	Global Decision Making Style
GUI	Graphical User Interface
HTML	Hyper Text Mark-up Language
IS	Information Systems
IT	Information Technology
JS	Java Script
MBA	Masters of Business and Administration
mIPIP	Mini International Personality Item Pool
MIS	Management Information Systems
MLA	Myopic Loss Aversion
MPC	Manufacturing Planning and Control
MRP	Materials Requirement Planning
MRP II	Manufacturing Resource Planning
MSc	Masters of Science
MTF	Make-to-Forecast (manufacture-to-forecast)
MTO	Make-to-Order (manufacture-to-order)
MTS	Make-to-Stock (manufacture-to-stock)
MWW	Mann-Whitney-Wilcoxon
N (n)	Number
NI	Naïve Interventionism
O&SCM	Operations and Supply Chain Management
OLS	Ordinary Least Squares
OM	Operations Management
OR	Operations Research
PCA	Principal Component Analysis
RDU	Rank-Dependent Utility
SC	Supply Chain (SCs for Supply Chains)
SCIP	Supply Chains In Practice (research group)
SCM	Supply Chain Management
S&OP	Sales and Operations Planning

1 Introduction

1.1 Research rationale

"A bad system will beat a good person every time"

William Edwards Deming[1]

On 13th of October 2012, two pilots on the Sriwijaya Air flight SJ-21 ignored the cockpit navigation system that they assumed to be faulty. They made a 'visual approach' landing an Indonesian passenger plane carrying 96 people relying on their own navigational skills – only to discover they were at the wrong airport. The incident report[2] was only released 4 years later, on the 12th of October 2016. This was not the first time this mistake had happened and preventive measures were in place. Pilots flying in the region must carry an information plate with a map and a chart with instructions containing a warning that the airport of Tabing can be mistaken for Minangkabau. They were operating in a repetitive and mostly controlled system with state of the art navigation technology enabling visibility and comprehensive failure prevention systems. Regardless, the SJ-21 pilots decided they knew better and landed at the wrong airport. Such level of control, technological sophistication, and skill requirements are unthinkable in most business activities, yet, mistakes caused by human judgement still happen.

Deming (1986) suggested that people's best efforts can be destructive when carried out without knowledge, understanding of variation or when the system is broken. Unnecessary actions, regardless of good intentions, are damaging. Businesses are complex systems of exchange with supply of and demand for goods, services, or both (Simon 1979; Deming 1986). Businesses are supervised by humans which are complex systems themselves (Jung

---

[1] The quote by W. Edwards Deming originally appeared on a Deming Four Day seminar in Phoenix (Arizona) in February 1993

[2] The report with reference KNKT.12.10.21.04 can be downloaded at: http://knkt.dephub.go.id/knkt/ntsc_aviation/baru/Final%20Report%20PK-CJT.pdf

1951; Deming 1986; Buss & Hawley 2010). Hence, when businesses fail, it could be because of faulty systems within but it could also be because of humans who design and operate these systems (Deming 1986). Deming (1986) warns that management can create the best systems, understand variation and possess knowledge but still fail if there is no understanding of psychology (people).

The attention to the importance of understanding both the system and the human simultaneously to ensure business success has been growing (e.g., Blattberg et al. 1990; Fildes et al. 2006; Kremer et al. 2011; Moritz et al. 2013). This research considers both the human and the system in a particular type of decision-making task – demand planning (DP) decisions – seeking to contribute to the debate on the human versus systems.

1.1.1 Imperfect decision makers

As the human decision making ability is not perfect, humans sometimes require additional support in order to make good decisions *(e.g.*, Kahneman & Tversky 1972; Tversky & Kahneman 1991). People make decisions constrained by both their cognitive resources and the task environment, a concept known as bounded rationality (Simon 1955; Simon 1956; Simon 1990). As a result of these limitations, people might systematically fail to cope with uncertainty, e.g., while making predictions or judging probabilities (e.g., Tversky & Kahneman 1974; Einhorn & Hogarth 1981; Benartzi & Thaler 1995). Uncertainty is common to most decision-making and is what makes it often very challenging. Failures of judgement or reasoning have been repeatedly demonstrated in different studies (e.g., Tversky & Kahneman 1974; Kleinmuntz 1990; Ouwersloot et al. 1998; Bazerman 2005; Lin et al. 2014). As a result, it is commonly accepted that human judgement alone is often insufficient and sometimes is prone to errors.

Since the 1970's research on errors has been gaining significant momentum. There has been a preference to study what people do wrong rather than what people do right. This has resulted in debate regarding over-citation biased towards negative performance (Crandall 1984; Evans 1984; Krueger & Funder 2004). One possible explanation is summarised by Crandall (1984, p.1499) as *'…mistakes are fun! Errors in judgement make humorous anecdotes, but good performance does not.'* As a result, there is a generalised prejudice against the human ability to make decisions. After reviewing the literature on decision making, Lopes (1991, p.65) also observed the over-citation of publications reporting on human errors and noted that the literature provided *'widely published claims that human judgement abilities are poor'*. However, the alternative view takes into account what type of tasks are demanded from human judgement. Therefore, it is incorrect to assume that human judgement is generally poor. Humans are simply expected to perform well in inappropriate circumstances. Human judgement excels in many other circumstances (see Dreyfus 1992).

Studying errors provides insight into how human judgement (in particular) and how cognitive systems (in general) work (Funder 1987). Many of these errors can be explained using rather simple behavioural biases. The notion of bias comes from the deterministic approach to choice, the Deterministic Theory (DT). DT is where the preference is predicted using a normative model and *systematic deviations* to the prediction constitute the bias. Biases are usually stable and often resistant to training (e.g., Bolton & Katok 2004). The degree to which an individual is prone to suffer from biases potentially depends on individual differences (e.g., Stanovich & West 2000; Shiloh et al. 2002; Oreg & Bayazit 2009).

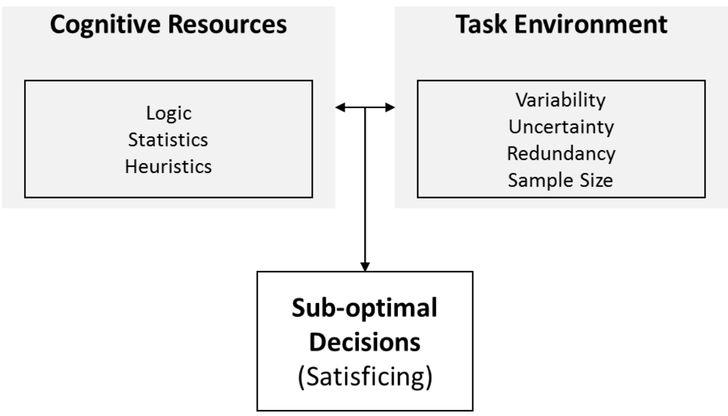


Figure 1 Bounded Rationality (based on Simon, 1955, 1956, 1990)

Managers (as human beings) can be considered boundedly rational (Simon 1979). Therefore, management relies on decision rules or frameworks (heuristics) to make decisions. These decision rules / frameworks are strategies to deal with human limitations and the complexity of the real world. Such a view of decision-making is largely explanatory in nature and does not enable reliable predictions of human behaviour. Bounded rationality is built on two main elements (depicted on Figure 1), the task environment and actors' cognitive ability (or cognitive resources).

Considering the task environment, the world is far too complex to be seized or perfectly understood by the human mind. Most real-world problems may trigger behaviour which cannot be explained by a standard 'rational' decision-theoretic model, where utility-maximizing agents are assumed to select optimal responses (Simon 1955; Simon 1979). Behaviour is shaped by its environment (e.g., Simon 1969; Gigerenzer 2001; Todd & Gigerenzer 2007; Wilke & Todd 2010; Gigerenzer & Gaissmaier 2011). Todd et al. (2012) suggest that environmental structures include:

1. Uncertainty: how well a criterion can be predicted
2. Redundancy: the correlation between cues[3]
3. Sample size: number of observations (relative to number of cues)
4. Variability in weights: the distribution of the cue weights (e.g., skewed or uniform)

Taking each of these points in turn. First, usually uncertainty determines the decision making approach. For example, the simpler the decision making approach is, the more robust it is. An example is the elementary *hiatus heuristic*, 'a one-reason' heuristic used to determine active and inactive customers. It is known to outperform more complex models with more information mostly because the decision making environment is highly uncertain (Hogarth & Karelaia 2007; Wübben & Wangenheim 2008). This can potentially explain why simple management frameworks are so widely used. Second, redundancy also tends to benefit the accuracy of inference strategies. For example, simple managerial heuristics tend to perform as well as strategies that integrate all available information in moderate to high

[3] Cues are information used to make a decision

information redundancy while the opposite situation benefits the integration of more information (Dieckmann & Rieskamp 2007). Third, as is suggested by the Law of Large Numbers (von Mises 1957), sample size generally tends to have a positive correlation with the accuracy of predictive models (Sedlmeier & Gigerenzer 2000). Consequently, it is beneficial to use more robust (simpler) models for smaller samples (Hogarth & Karelaia 2007). Finally, variability in weights introduces the idea of uniformity and skewness, where simple heuristics fit decisions in environments of moderate to high variability better than more complex models (Hogarth & Karelaia 2007).

Considering the decision-maker's cognitive ability, the actor can use logic, statistics (Tversky & Kahneman 1983), or heuristics (Gigerenzer & Gaissmaier 2011) to make decisions in complex environments. Gigerenzer and Gaissmaier (2011, p.454) define heuristics as *"…a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods."* Heuristics are developed naturally or artificially to simplify the cognitive strain and to enable not perfect, but often 'good enough' (*satisficing*) decisions (Shah & Oppenheimer 2008). Simon (1990) considers the limited computational capabilities of an agent as being recognition, memory and reaction. Models of cognition can get extremely complex and detailed, hence, for the sake of simplicity, it will be assumed that human behaviour is limited on the three basic levels (Simon 1969; 1987; 1990).:

(i) The number of items they can memorise is relatively small, i.e., humans have limited short-term working memory;

(ii) Recognition is powerful and quick but not instantaneous and is dependent on previous knowledge recall;

(iii) Reaction is not instantaneous.

1.1.2 Working around human limitations

One of the main limitations of the human mind is memory and processing ability (Simon 1987). Artificial systems and developed technology can compensate for these limitations. They can enhance human abilities by allowing high frequency sequential and parallel processing. They also enable relatively quick access to vast amounts of working memory. The ability to rely on external tools in order to process intensive repetitive tasks dramatically changed the '*accuracy-effort trade-off curve*'. The accuracy-effort trade-off is a power distribution type curve relationship between the quality of the decision (accuracy) and cost (effort), i.e., the greater the effort, the greater the accuracy (Payne et al. 1993; Kruglanski & Gigerenzer 2011).

Artificial systems have the potential to automate a number of tasks. Uptake is limited by both technological limitations as well as user preference. The idea of relying solely on the artificial system '*…for making important decisions that do not depend heavily on human inputs seems unlikely as well as unattractive*' (Edwards & Fasolo 2001, p. 588). People (in general) and experts (in particular) have a long history of resistance to '*machines*' taking over decisions and judgements (Meehl 1954).

Technology and more specifically Decision Support Systems (DSS), should not focus on unilateral control of the decision-making process but rather enhance the human decision making processes. This can take four forms (Larrick et al. 2004, p. 330):

i) Basic normative algorithms that are known to be unnatural and hard to implement or remember for humans can be assigned to DSS.
ii) Analytical decision tools and decision algorithms that are otherwise intimidating and hard to understand can be incorporated in a user-friendly DSS.
iii) The thankless task of consistency checks, such as criteria weights or probabilities, can be made less intrusive, faster and effortless if performed by a DSS.
iv) DSS can perform sensitivity analysis.

Systems must be designed around human limitations, allowing a symbiosis between human intellectual ability enhanced by the power of high frequency processing and working memory that can be incorporated into the systems. As Silver (1991, p.106) sets 'to establish a unified approach, one that recognises the importance of both technological and behavioural issues'. The directive is *'…how our mind's design, relying on decision mechanisms tuned to specific environments, should be taken into account in our technology's design, creating environments that can enable better decisions'* (Todd 2007, p.1317).

1.2 Demand planning

1.2.1 The importance of demand planning

Planning is one of the most important processes in business and operations management (OM) (Lee 2004). It relies heavily on decision-making and decision support. The performance of planning activities in a demand-supply network is highly dependent on the DP outputs (Chen et al. 2007). This makes DP central to the concept of supply chain (SC) (Christopher 2011, p.13) because DP is essential for balancing supply and demand. It allows the company to reduce its purchasing, production and logistics costs as well minimizing inventory necessary to buffer for uncertainty.

To understand the importance of DP it is important to consider its implications on the supply and demand mismatch (Stadtler et al. 2015). One of the most common means of addressing uncertainty is by holding inventory that buffers variation. Inventory costs are commonly estimated to be 20% on average across different industries globally, however, this is often an underestimation.[4] These costs come in many forms. For example, inventory costs cash and the interest on that cash, insurance and taxes on inventory, labour to handle the inventory, warehouse rent costs to store the inventory, are just some of the examples. On the extreme end, failure to anticipate demand can result in complete inability to satisfy the customer and to do business (Stadtler et al. 2015). This could result in a disruption and usually has implications on the whole SC: *'Disruptions occur here from a mismatch between a company's projections and actual demand as well as from poor supply chain coordination. Consequence of which are costly shortages, obsolescence, and inefficient capacity utilisation. An important issue in this context, affecting forecast quality and therefore demand-side disruptions, is the bullwhip effect,*

[4] Source: Forbes.com

which is characterised by an amplification of demand volatility in the upstream direction of the supply chain' (Wagner & Bode 2006, p.304).

In practice, it is extremely difficult to accurately calculate the total cost of the supply and demand mismatch as such calculations need to often involve subjective values such as (potential or real) opportunity cost or reputation. For example, in 2013, poor inventory management cost Walmart $3 billion. The situation was contradictory because its inventory was growing faster than its sales but the merchandise was not on the shelves for customers to buy. This seriously damaged the Walmart's reputation.[5]

DP is the starting point for SC planning and its quality will affect all subsequent planning activities (Chen et al. 2007). Central to DP is forecasting (Stadtler et al. 2015). Forecasts are critical for OM and integrated part of functions of scheduling, resource planning, and marketing (Fildes et al. 2006). Generally, the forecast within a plan combines managerial judgement with statistical forecasts within a support system. This means that forecasts incorporate decisions under uncertainty involving humans and systems.

1.2.2 Dealing with demand uncertainty

One of the main challenges of DP is dealing with the demand uncertainty (Lee et al. 1997; Geary et al. 2006). Demand uncertainty makes the demand signal one of the most unreliable sources of information in the SC (Geary et al. 2006). In general, the further upstream in a supply-demand chain, the greater the demand uncertainty as it is propagated and magnified (Lee et al. 1997; Geary et al. 2006). This makes the DP performance one of the greatest challenges in management and its improvement is a priority for business success.

Demand uncertainty has different sources, uncertainty can be natural but often much is caused both by the planning systems as well as by managers themselves (Lee et al. 1997). There is no clarity about what the different contributions to uncertainty are and how they can be addressed, especially

[5] Sources: Forbes.com and Bloomberg.com

regarding managers. For example, Wu and Katok (2006) show that training can improve individual knowledge and understanding of the system, which consequently reduces behavioural issues causing uncertainty. However, Sterman and Dogan (2015) conclude that irrational behaviour still persists even with training and perfect visibility across the chain – managers could not resist the urge to introduce uncertainty under conditions of certainty. This highlights an important problem with DP that must be better understood.

1.2.3 The problem with demand planning

The DP task, especially forecasting under uncertainty, is hard. Contrary to the belief that greater effort leads to better results, sometimes unnecessary effort can have negative consequences and lead to worse results (e.g., see Katsikopoulos 2010). Childerhouse et al. (2003, p.135) highlight that *'…much uncertainty is induced by "players" [managers] within the system as opposed to being introduced by the marketplace.'* In practice, managers show relatively little adherence to the original plan (Harrison, 1997). After the plan is developed, managers often intervene and make changes due to mistrust, second-guessing, over-reactions, and fear of losing sales (Niranjan et al., 2009). Unnecessary interventions with the plan often have negative effects on the whole SC (Niranjan et al., 2009). Examples of unnecessary interventions include hoarding and phantom ordering, which are often triggered by emotional impulses (Sterman & Dogan 2015).

Much of the irrational behaviour in DP seems to be resistant to additional knowledge, training or experience (e.g., Schweitzer et al. 2000; Sterman & Dogan 2015). In an experiment, Sterman and Dogan (2015) demonstrate that even with perfect visibility of demand (which was known to be constant) as well as perfect knowledge of the orders at each instance of the chain, managers cannot resist the urge to hoard, ending up destabilising the whole SC. On a larger scale, the behavioural effect is often cumulative and can throw an efficient SC out of balance (Mason-Jones & Towill 2000; Croson & Donohue 2006). Such behaviour causes amplification of small variations (over-reactions and distorted information) e.g., the Forrester effect (Forrester, 1958) also known as bullwhip effect (Lee et al., 1997). This results in extra costs, e.g., inventory, markdowns, stock-outs or obsolescence (e.g., Niranjan et al., 2009).

Managerial interventions are especially problematic for forecasting in DP. There is strong evidence that combining managerial judgement with statistical forecasts in support systems negatively affects accuracy (Fildes et al. 2006) and, consequently, has an adverse effect on DP performance. Part of the issue comes from people's preference towards their own intuition ('gut') as opposed to the artificial rationale in the form of formulas, statistical or mechanical procedures (Meehl 1954; Kleinmuntz 1990; Dane et al. 2012), etc. While intuition and judgement can be extremely powerful in some situations (e.g., see Syntetos et al. 2016), it can also lead to judgement errors and biases (e.g., Tversky & Kahneman 1974). Business management is no exception to this problem (Lawrence et al. 1986; Lim & O 'connor 1995; Bazerman 2005; Fildes & Goodwin 2007).

Following their intuition, managers systematically disregard the existing forecasts in general (even though these forecasts are often quite accurate). Anecdotally, this is referred to as the *'we know best'* syndrome (Mason-Jones & Towill, 1998, p.19) and often attributed to overconfidence (Brenner et al. 1996; Lawrence & Sim 1999; Moore & Cain 2007). Statistical forecasts are often completely ignored as managers try to incorporate known special events (e.g., promotions) while making adjustments. This happens even when the statistical forecast accurately describes the underlying predictable pattern (Goodwin & Fildes 1999). In cases when the outputs from statistical models are not completely ignored, managers are prone to make frequent adjustments (Fildes et al., 2009) suffering from overconfidence when it comes to the quality of their judgement (Brenner et al. 1996; Lawrence & Sim 1999; Moore & Cain 2007). The quality of the adjustments to the forecast is likely to be linked to heuristics and biases (Goodwin 2002). When it comes to selecting the statistical model, managers perform poorly (Lawrence et al. 2002) resorting to default parameters and sub-optimal models (Fildes & Beard 1992). In an attempt to improve the sub-optimality of the chosen models and parameters, managers make exaggerated judgemental adjustments to the statistical forecast, which are still not 'good enough' when compared to better alternative statistical models (Goodwin et al. 2007).

1.3 The research gap and the subsequent research question

To date, literature on decision making in planning tasks primarily focused either on systems, management or on individual behaviour separately. On the one hand, decision making literature (behavioural economics and psychology) offers many different explanations to some sub-optimal performance in the real world (e.g., Tversky & Kahneman 1974; Schwenk 1988; Bazerman 2005). This sub-optimal performance is observed in the form of systematic deviations (biases) from normative expectations and theoretical optima. Moreover, trait theory from psychology literature suggests that individual differences play a significant role in explaining group heterogeneity and differences in decision making performance (Weber & Milliman 1997; John et al. 2008; Fleeson & Jayawickreme 2015). On the other hand, decision support literature has looked into improving the way systems support human decision-makers (Leighton 1981; Silver 1991; Burstein & Holsapple 2008; Goodwin et al. 2011). Finally, OM and operations research (OR) literature describes a wide variety of the challenges in business, many of which are caused by both management systems as well as individuals, highlighting behavioural issues in the context of operations and supply chain management (O&SCM) (e.g., Lee et al. 1997; Geary et al. 2006; Bendoly 2006; Carter et al. 2007; Niranjan et al. 2009; Kaufmann et al. 2010).

Although the problem of managerial judgement under uncertainty has been observed from many different perspectives (Tversky & Kahneman 1974; Silver 1991; Lee et al. 1997; Bazerman 2005; Geary et al. 2006; Bendoly 2006; Niranjan et al. 2009), the causes behind this persistent behaviour are not clear and there is still lack of research providing cross-field solutions. Recent advances in O&SCM literature offer a progressively interdisciplinary view on the planning problem incorporating insights from behavioural economics and psychology. For example, there is some empirical evidence suggesting that individual differences play a significant role in decisions similar to the ones made in DP (Franken & Muris 2005; Lapide 2007; Moritz et al. 2009). Similarly, how decision-makers interact with DSS to make DP decisions received significant attention (O'Connor et al. 1993; Fildes & Hastings 1994; Webby & O'Connor 1996; Parackal et al. 2007).

Following the call for theory in O&SCM (Carter 2011) and current absence of solid explanations behind some clearly observed behavioural issues in the literature (Bolton & Katok 2008; Gans & Croson 2008; Croson, Schultz, et al. 2013), it is necessary to borrow theory from other relevant fields to seek an explanation. Without theoretical understanding of the behavioural mechanisms governing the ill behaviour of managers, it is hard to progress in understanding and solving the problem of sub-optimal decisions in DP.

Although problems related to human behaviour have been highlighted in previous OM and OR research, theoretical foundations explaining the observed behaviour require further development. The focus of the research on DP to date has been on the average behaviour of a large number of individuals (representative agent behaviour) which is different from concentrating on individual heterogeneity in DP tasks (individual differences). While there is certainly great value in observing the average of a large sample, in order to derive the basic behavioural principles in various environments, typically, DP decisions are made by individuals rather than groups of people and individual heterogeneity among planners in their propensity to generate successful plans is important. Therefore, it is important to explore individual differences in the context of OM (Croson & Donohue 2002; Gans & Croson 2008; Croson, Schultz, et al. 2013) as so far contributions relating to this aspect of DP decision making are limited (Zmud 1979; Strohhecker & Größler 2013; Moritz et al. 2013). It is still unclear why individuals behave in a certain way when making DP decisions and, hence, necessary to understand the relative contribution of individual factors and systems parameters to DP which lies at the core of this thesis.

The main research question is as follows: *What is the contribution of individual differences and planning policy parameters to demand planning performance?*

The aim of this research is to develop and test a theoretical framework drawing on theory from behavioural economics and psychology to identify planning policy parameters and individual traits that can be used to predict DP performance. The main contributions of this thesis can be split across four fields: Engineering, Management, Psychology, and Economics (Figure 2).

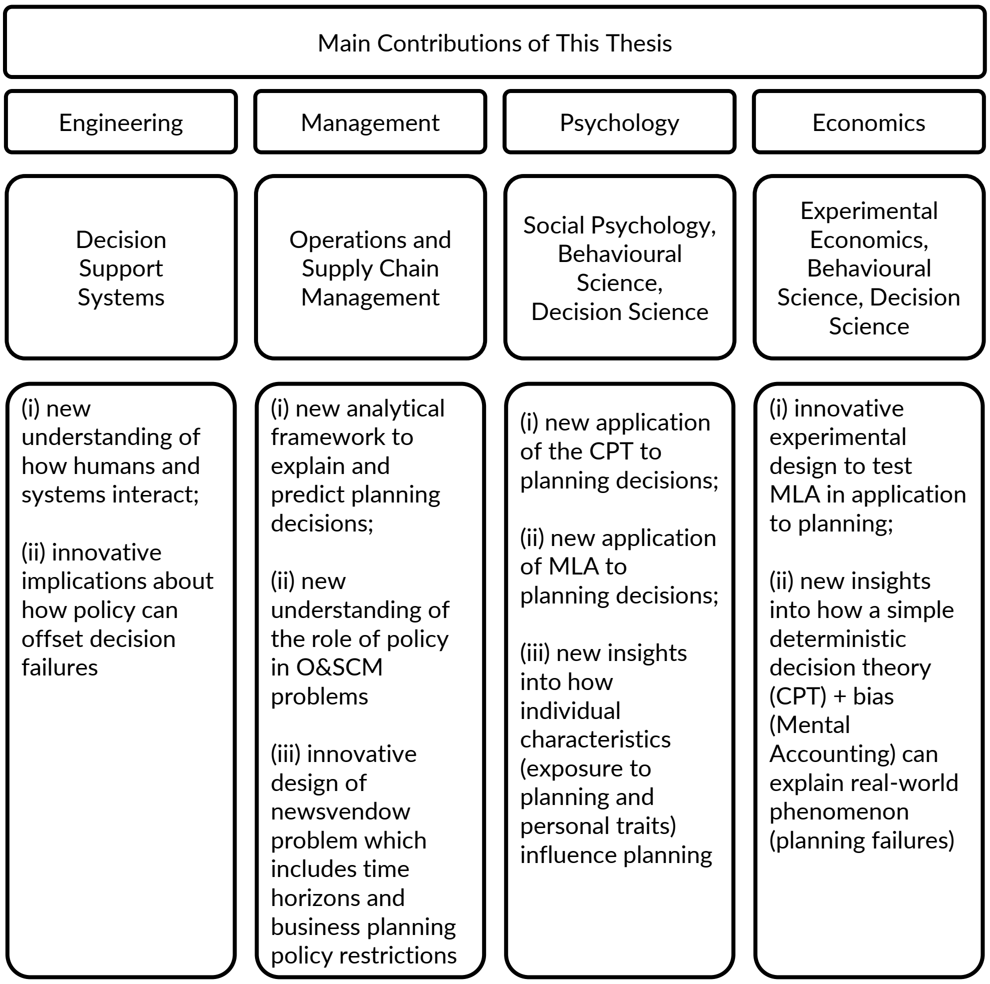


Figure 2 Summary of the main contributions of this thesis

For DSS, it is provided a new understanding of how humans and systems interact together with innovative implications of how policy can reduce decision failures.

For O&SCM it is suggested a new analytical framework explaining and predicting DP decisions. Moreover, it is provided new understanding of the important role of policy in O&SCM. Finally, the contribution to O&SCM is an innovative design of the newsvendor problem that includes time horizons and DP policy restrictions.

For social psychology, behavioural science and decision science it is provided a new application of the Cumulative Prospect Theory (CPT) and Myopic Loss Aversion (MLA) to DP decisions. Additionally, a new insight is provided on how individual characteristics (e.g., exposure to DP and personality traits) influence DP performance.

Finally, the contribution to experimental economics, behavioural science and decision science is an innovative experimental design to test MLA in application to DP. A new insight is provided into how a simple deterministic theory (CPT) with the mental accounting bias (as part of the MLA) can explain real-world phenomena (sub-optimal DP decisions).

1.4 Thesis structure

The thesis is structured is outlined in Table 1. Following this introduction, Chapter 2 reviews the literature around the task, performance and support of DP decisions. Chapter 3 expands the literature into theory potentially explaining the previously described behavioural problem, exploring models of human behaviour and individual differences. This is used to develop a conceptual framework is developed together with a set of hypothesis. Chapter 4 presents the research design to test the framework and hypothesis; it presents the adopted epistemology, reviews behavioural experiment methodology and outlining the experimental design used to collect data to test the hypothesis. Chapter 5 presents experimental results testing the previously developed hypothesis. Finally, Chapter 7 summarises the research, focusing on findings, contributions, limitations, and suggests further research directions.

Table 1 Thesis structure

Chapter	Overview
Ch. 1 Introduction	Provides the research rationale, introducing the relevance of DP and the associated problem of mangers' behaviour followed by the research question. Thesis structure is outlined.
Ch. 2 Demand planning: process, performance and support	DP is explored in literature. The task is described along with its context and the types of decisions. The problem of DP performance in form of instability, chaos, nervousness reviewed followed by the relationship between decision-makers and decision supporting systems.
Ch. 3 Underpinning theory and hypothesis	Because the problem of DP is greatly dependent on the individual behaviour, the literature review expands into behavioural economics and psychology to explain the previously identified phenomena of sub-optimal DP performance. Along with the theoretical development relevant hypothesis are derived. The chapter closes with a conceptual framework.
Ch. 4 Research design	The ontological and epistemological positioning is explained followed by the description of behavioural/ decision making experiment used to test the framework and respective hypothesis.
Ch. 5 Results and analysis	The experimental results are analysed using econometric methodology to test the hypothesis.
Ch. 6 Conclusions	The last chapter summarizes the research, includes the review of the findings, contribution to both practice and theory, limitations, and suggest further research.

1.5 Chapter summary

People's best efforts can be destructive and unnecessary actions, regardless of good intentions, are damaging. Mistakes occur because humans are imperfect decision-makers, and it is essential to consider both the decision-maker and the environment (system) in which decisions are made. The quality of decisions depends on both (Figure 3).

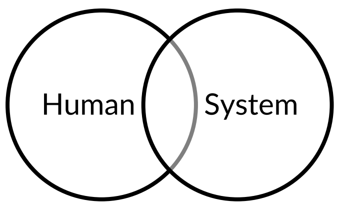


Figure 3 Human and System base framework to consider decision-making

DP is essential for business management. One of the main challenges in DP is making decisions under uncertainty and risk. Managers have a track record of making sub-optimal decisions in such context. Nonetheless, managers still trust their own judgement over decision support and statistical methods.

Managers are known for unnecessary interventions, demonstrating mistrust, and over-reacting to small changes. This sub-optimal behaviour is dependent on both the manager as an individual as well as the system (e.g., the set of policies) in which decisions are made. In what follows, this thesis focuses on the relative contribution of (i) humans (reasoning, biases, and individual differences) and (ii) system (DP policy parameters) on DP performance. The next section reviews the literature around DP, describing the task, performance and DP decision support.

2 Demand planning: process, performance and support

2.1 Introduction

The structure of chapter 2 is outlined in Table 2. Section 2.2 positions DP in the broader context and presents the main concepts around DP that are required for understanding the potential implications of DP decisions. Section 2.3 describes DP performance where the concepts of *system nervousness* (from the system perspective) and *naïve interventionism* (from the human perspective) are introduced. Due to the challenge of DP, DSS are used. However, this comes with resistance and managers often choose to do differently. Decision support can either restrict decision-making or guide it. Its application in DP is discussed. These issues are discussed in section 2.4. The chapter closes with a summary of the key points.

Table 2 Chapter 2 Structure

Section	Overview
2.2 Demand planning process	This section focuses on what is the context and decisions made. The process of DP is broken down into its context of matching supply with demand. Discusses its dependency with the planning horizon and the nature of the available information that is used in the forecasting process.
2.3 Demand planning performance	DP performance is related to nervousness in the SC. Demand plan failures are explained as system nervousness and naïve interventionism. Individual differences are observed as an extension of naïve interventionism.
2.4 Humans and systems: supporting demand planning decisions	The relationship between managers and systems supporting DP decisions is reviewed in terms of how people have been resisting to automation. Follows the two main ways of supporting decisions via either decisional guidance or system restrictiveness. The section closes with an overview of decision support for DP decisions.

2.2 Demand planning process

2.2.1 Balancing supply and demand

To understand the DP process, it is necessary to start by considering its context and reach. Businesses are complex systems of exchange with supply of and demand for goods, services or both (Simon 1979; Deming 1986). These systems form networks which can be broadly referred to as SC. Christopher (2011, p.13) defines SC as a '...*network of organizations that are involved, through upstream and downstream linkages, in the different processes and activities that produce value in the form of products and services in the hands of the ultimate consumer'*. The concept of SC is here used instrumentally as an umbrella term for the broad effects of DP decisions.

DP, even when local or contained within an organisation, has implications beyond firms' boundaries (Stadtler 2005). It is critical to think about DP considering the broader SC since the effects of bad DP will affect the rest of the chain. Failure to plan affects the SC upstream, e.g., in the form of the *bullwhip effect* (Lee et al. 1997), and downstream, e.g., with disruptions (Craighead et al. 2007). *Bullwhip effect* (also known as Forrester effect) is the amplification of demand caused by information distortion as the demand signal passes through the chain (Lee et al. 1997). DP helps to addresses one of the greatest challenges for any business which is minimising supply and demand mismatch (e.g., Vitasek et al. 2003; Christopher 2011).

Uncertain demand presents a particular challenge as it requires the ability to predict or react as quickly as possible (Christopher 2011). Underestimating demand is an issue as demand can then exceed supply, leading to out of stocks and poor customer service. Whilst overestimating demand does not affect customer service, as supply exceeds demand it can lead to the growth of costly inventory buffers.

2.2.2 Demand planning process

DP usually produces forecast data describing demand for products or groups of products by period over a set planning horizon. Hence, DP precedes master planning for demand fulfilment (Figure 4). Master planning's main purpose "...is to synchronise the flow of materials along the entire supply chain" (Stadtler et al. 2015, p.155). Therefore, master planning process creates the plan for the whole SC, including purchasing and production decisions. It generates the plan of supply from external and internal sources. The purpose of DP is threefold: (a) improving forecast accuracy; (b) increasing customer service level; and (c) reducing inventory (Stadtler et al. 2015).

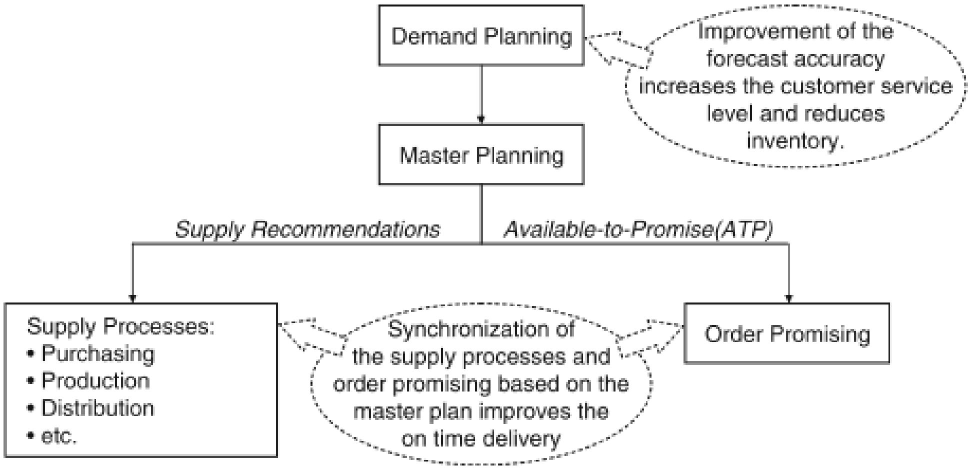


Figure 4 Positioning of demand planning (Stadtler et al. 2015, p.180)

The DP process (Figure 5) is a multistage sequence. The process starts with the preparation of DP structures and historic data, followed by the stage of developing the statistical forecast. The statistical forecasting stage is followed by judgemental forecasting stage where managers analyse the statistical outputs considering additional information that is usually not available historically or is ignored by the forecasting models. Managers working with DP are often responsible for the forecasting, being required to choose the best forecasting methods and adjust the forecasts using their own judgement (Fildes & Goodwin 2007). After adjustments, a consensus stage is reached and planning of dependent demand stage precedes the release of the final forecast into the subsequent processes. It has a feedback loop through the master planning giving the possibility for management to revise the forecasts based on information about capacity constraints (e.g., need to consider how to create additional capacity or manage demand) or surplus (e.g., need to stimulate demand). The impact of the demand plan on the SC is simulated by the master planning process through the *what-if* analysis that is then fed back to managers who judge and adjust if necessary. This feedback loop is essential to understand the effect of business decisions on demand.

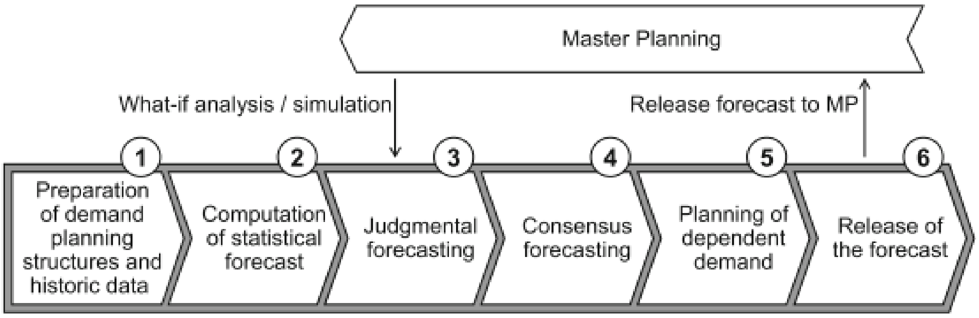


Figure 5 Demand planning process (Stadtler et al., 2015, p.153)

Demand can be greatly influenced by business decisions. Demand is greatly affected by price promotions, new product launches, regional promotions and product bundling. These kind of decisions can be also referred to as demand shaping decisions (Lapide 2013) carried out primarily by marketing and sales managers. Although SC managers are not directly responsible for these decisions, it is critical to ensure that the supply meets the expected future demand as well as to advocate that demand shaping must consider the ability to supply (Lapide 2013).

The process of DP, whilst theoretically comprehensive, can fail in practice. First, the nature of available information that comes from many different sources is imperfect and information streams need to be optimised. Second, the transition between statistical and judgemental forecasting stages requires structuring and it is necessary to understand how it depends on planner's cognitive ability and training. Third, it is important to understand the effect of the forecast horizon and its relationship with uncertainty. The process of DP considers the future. The future can be split as near or far future. Therefore, one of the main considerations to have in DP is the planning horizon.

2.2.3 Planning horizon

The longer the planning horizon, the greater is uncertainty. Generally both statistical and judgemental forecasting become less accurate as the planning horizon increases (Lawrence et al. 1985; Lawrence et al. 1986). Mostly short term, judgemental forecast quality often suffers due to managers' behaviour (e.g., Webby & O'Connor, 1996; Parackal et al., 2007). Although in some cases, intermittent forecasts can be significantly improved when managers provide input on irregularities (Syntetos et al. 2009; Syntetos et al. 2016).

Another important consideration about planning horizon concerns managers. Planning is arguably not a natural act, at least not for everybody. Hey and Knoll (2007, p.8) study experimentally the human propensity to plan in dynamic risky situations and conclude that '*just over half of the subject plan fully, while the rest do not plan ahead at all.*' Bone et al. (2009, p.12) conducted a further experimental study and find that over half of the experimental subjects '*do not appear to be planning ahead; moreover, their ability to plan ahead does not improve with experience*'. This means that those subjects think only about immediate implications of their decisions and this behaviour is resistant to experience.

There is no general consensus on how far ahead in time short term and long term forecasting looks, as it is highly dependent on the context (Taylor & Thomas 1982; Armstrong 2001; Goodwin 2002). It is however generally accepted in SC literature that short term forecast considers the '*next period or periods*' and long term is '*beyond the short term*' (Armstrong 2001). Short-term forecasting is usually achieved with relatively simple procedures. Long term forecasting typically needs more than simple historical data analysis and goes beyond the extrapolation of the trend and seasonality. A common way to work around the issue of lower accuracy for long-term is by planning at a product family level rather than individual stock keeping units.

Finally, time horizon plays a major role in plan instability (Blackburn et al., 1986) and consequently affects business performance. For the purposes of this analysis, the focus is on the **short-term time horizon** as the majority of important DP problems occur within the short time frame.

2.2.4 The nature of available information

Relevant data for forecasting (Table 3) based on Fildes et al. (2006) constitutes time series data, information on customers' activities, information on other relevant variables (e.g., major events, competitors activities, weather forecasts), earlier forecasts from previous periods, other forecasts and information on past forecast errors.

Table 3 Types and source of available information (based on Fields et al. (2006))

Type	Source
Time series data	Usually has different levels of aggregation (product group, pack size, SKU, region, customer, day, week or month). Might require data cleansing (e.g., outlier removal such as special events or promotions). Cleansing decisions are often made based only on judgement. Time series analysis enable powerful insight on the true unpredictability (noise), trends, seasonal patterns and periods.
Information on customers' activities	Information on activities such as discontinued products, promotions and new product development
Other relevant variables	Major events such as festivals or sports leagues, weather forecasts, competitors sales and other activities such as promotions, discontinued products and new product launches
Earlier forecasts from previous periods	One common method is rolling forecasting which updates earlier forecasts when approaching the forecast period
Other forecasts	Different levels in the organisation produce different forecasts, e.g., accounts managers are closer to their customers and can produce more accurate forecasts on the basis of their closer contact
Information on past forecast errors	Knowing how accurate previous forecasts were can provide relevant feedback for the forecasters

Demand patterns will influence DP judgement (Theocharis & Harvey 2016). Accordingly to Wold's theorem (Wold 1954), a time series (i.e., demand over time) consists of two parts, deterministic and stochastic. Alternatively this can be distinguished as regular, irregular and noise parts (Fildes et al. 2006). Following this approach, the deterministic part is assumed as predictable. The time series contains the trend, cycle and season. Trend and cycle can (in some cases) be merged into a common component. This can be considered all regular patterns or relationships. The non-deterministic (stochastic) part contains the random residual also known as the noise component. Importance must be given to outliers forming the irregular components. It is possible to distinguish at least two types of outliers in demand signals. One type of outliers is natural and the other is artificial (often foreseeable). An example of a natural outlier is a natural disaster or a plague, while an artificial outlier could be a significant promotion. While a natural disaster is often unexpected, a promotion is planned, and, therefore, predictable.

Additional sources of data can be greatly relevant in improving forecasting accuracy, in particular judgemental adjustments as irregular components can be considered. This includes information on activities such as discontinued products, price promotions, new product launches, regional promotions, product bundling, and other business decisions that can affect the demand. Additional relevant information can come from events such as major festivals (e.g., Beer Festival) or sports leagues (e.g., Premier League), weather forecasts (e.g., floods, rain seasons, and heat waves), competitors' sales and other activities such as promotions, discontinued products and new product launches.

Earlier forecasts are also relevant information for DP as one common method of formulating new plans is *rolling forecasting* which updates earlier forecasts when approaching the forecast period so it considers historic forecasts. The same organisation usually produces many different forecasts with different levels of aggregation or focus, e.g., accounts managers are typically close to their customers and can produce accurate forecasts because of their close contact. Finally, previous forecast accuracy is relevant as knowing that a given product has been accurately forecasted before or systematically off target, will have implications on the approach and forecasting methods used.

2.2.5 Statistical and judgemental forecasting process

The forecasting stages of the DP process involve human beings at every stage. Employees and managers who perform the DP in general and the demand forecasting process in particular usually have experience in developing forecasts but have limited theoretical knowledge of forecasting errors and often lack formal training in statistics or statistical methods (Fildes & Hastings 1994). Forecast adjustments based on judgement often happen in managerial meetings under time pressure as well as lack of visual means and flexibility to provide a quick analysis (Fildes et al. 2006).

As illustrated in Figure 5, the DP process includes two types of forecasting processes. Step 2, is the computation of the statistical forecast, and step 3 is the judgmental forecasting process. Consensus is then reached in step 4.

In practice, the demand forecasting process is often divided into two steps: (1) statistical forecast and (2) judgemental adjustment (Figure 6). Demand forecasting usually starts with the creation of a statistical forecast, often conducted automatically (when planners use a template model habitually applied in their organisation) or manually (when planners choose their own 'custom' model). After this, the planners make judgemental adjustments to the statistical forecast taking into account special factors (factors relevant to their organisation) as well as other available information (e.g., relevant externalities). The outcome of this two-step procedure is a set of final forecasts often developed for many different products. This set, in turn, is used to plan SC operations (Fildes & Beard 1992; Fildes et al. 2006). This procedure is often repeated weekly and, sometimes, daily making the demand forecasting process too large to be handled manually and requiring a great degree of automation. This means that planners have very short time frames to apply judgemental adjustments.

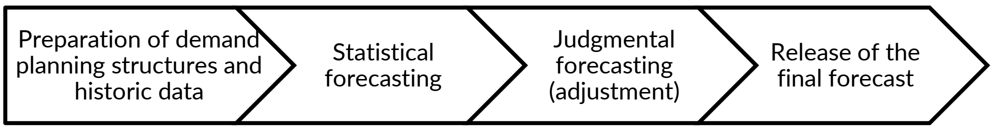


Figure 6 The stages of the forecasting process in the demand planning process (adapted from Stadtler et al., 2015, p.153)

Considering the process and components of the time-series, demand forecasts are set to capture two types of components (e.g., patterns, relationships, events), which, in turn, could be regular and irregular. Regular components are detected and explained by statistical models during statistical forecasting step and irregular (but foreseeable) components by planners in judgemental adjustment step. In principle, evidence (e.g., Goodwin & Fildes 1999; Goodwin et al. 2007; Fildes et al. 2009) suggests that statistical models (automated systems) outperform human judgement in identifying and modelling the regular components. In contrast, automated systems based purely on statistical models often fail to forecast irregular components (Lawrence et al. 1986; Goodwin & Wright 1993; Syntetos et al. 2009).

The combination of statistical forecasting and judgemental adjustment steps can often lead to greater accuracy than each step taken separately (e.g., Lawrence et al. 1986; Blattberg et al. 1990). The final forecast accuracy can be greatly improved especially when the manager adds extra irregular information that is not included (or is naturally ignored) in the statistical model (Mathews & Diamantopoulos 1990; Donihue 1993; Goodwin & Fildes 1999; Fildes & Stekler 2002). However, this is not always true as managers have a tendency to override the statistical forecast of the regular component (Goodwin & Wright 1993; Harvey 1995; Lim & O 'connor 1995; Goodwin & Fildes 1999; Sanders & Ritzman 2001; Sanders & Manrodt 2003). While judgemental adjustment is often necessary and beneficial to the demand forecasting process, it may also harm the forecast creating such problems as excess inventory (Sterman & Dogan 2015), amplification of demand (Lee et al. 1997), to name a few.

2.2.6 Demand planning process summary

DP enables customer service levels to be maintained at a lower cost, as less inventory needs to be held to buffer against the mismatch between demand and supply. DP is the starting point for the subsequent planning processes (e.g., precedes master planning). The DP process includes statistical forecasting followed by judgemental adjustments to consider both the regular and irregular patterns. Judgement can be very powerful to incorporate additional information in the forecasting process (e.g., Lawrence et al. 1986; Mathews & Diamantopoulos 1990; O'Connor et al. 1993). Such information is usually known future irregular events. Irregular events are usually ignored by statistical models. Examples of this are demand shaping business decisions, information about the competition. This is both good (e.g., in case of intermittent demand) and bad when the separation between regular and irregular components is not clear and managers adjust what is highly regular introducing irregularity artificially.

As a result, DP is subject to judgement errors within a system. This constitutes a potential source of DP failure due to a number of reasons related to managers' behaviour. First, many people naturally do not plan ahead in time

(Hey & Knoll 2007; Bone et al. 2009). Second, often managers working with DP lack formal training in statistics or statistical methods for forecasting (Fildes & Hastings 1994). And finally, managers have a general tendency to override the statistical forecast introducing unnecessary adjustments to the statistical forecast (Goodwin & Wright 1993; Harvey 1995; Lim & O 'connor 1995; Goodwin & Fildes 1999; Sanders & Ritzman 2001). This can significantly compromise DP performance.

2.3 Demand planning performance

The performance of DP is not the same as forecasting performance. The quality of a forecast is usually measured using accuracy estimates such as mean absolute deviation (MAD) or mean absolute percentage error (MAPE) while the performance of DP can be assessed in terms of profit, revenue or costs (Stadtler et al. 2015). Many issues can affect DP performance. The following sections focus specifically on DP performance from the system and manager's perspective.

2.3.1 System nervousness in the supply chain

SC complexity and uncertainty forces cause 'chaos' in a SC (Christopher & Lee 2001). Whilst part of complexity and uncertainty is natural, part is artificially induced by people. The base assumption is that the majority of people across the SC want to make good decisions. However, Deming (1986) argued that people's best efforts can be destructive when carried out without knowledge, understanding variation or when the system is broken. This is very likely to happen in conditions of complexity and uncertainty typical to most modern businesses and their SC's. Chaos in SCs makes it impossible to make the right decisions hence it's also impossible to design optimal solutions (Christopher & Lee 2001).

Both people and systems can be the source of the problem. On the one hand, the SC context is far too complex to be perfectly seized and understood. As a result, the part of the chaos comes from managers, in the form of mistrust, unnecessary interventions, distorted information, second guessing and over-reaction across the SC (Christopher & Lee 2001, p.2). On the other hand, managers are not the only source of chaos. Several causes of chaos come from the system. Lee et al. (1997) suggest that information delays and distortion, order batching, shortage gaming, sales promotions, fluctuations of price and rationing contribute to chaos and instability in the SC.

An example of such chaos is the *bullwhip effect* (Lee et al. 1997) where variation greatly increases from downstream to upstream SCs. This effect of amplification of small variations has been earlier described by Forrester (1958) and named the Forrester Effect. Lee et al., (1997) suggests that the bullwhip effect is a consequence of the players' rational behaviour within the SC and that companies wanting to control the bullwhip effect must focus on modifying the infrastructure and policy rather than the decision-makers' behaviour. According to Lee et al., (1997), demand signal processing is one of the main contributors to the bullwhip effect as managers' perceptions and mistrust lead to readjustment of the perceived demand forecast.

Deming (1986) refers to such chaos as 'system nervousness' which affects the SC performance: '*This increased nervousness will of course lead to higher costs and inefficiencies through over-ordering and "squirreling" inventory*' (Christopher & Lee 2001, p.2). DP can either dampen or amplify variation, affecting chaos. Hence, the performance of a SC is dependent on the DP performance.

DP performance depends on both managers and systems. Further the construct of 'nervousness' will be used to designate also chaos in the DP process. The following sections explore nervousness from both human and system perspective in order to understand what the main factors are contributing to DP performance. The following review of literature on nervousness in the SC goes beyond the DP function alone. Processes in a SC are highly interconnected. Therefore, the effects of DP nervousness can be detected in other processes. When the demand signal propagates in the SC, DP nervousness can distort the demand signal and compromise the quality of the master plans that consequently will compromise performance of processes built around them.

Hence, part of the uncertainty comes from the system, however, some authors defend that *'…much uncertainty is induced by "players" [managers] within the system as opposed to being introduced by the marketplace.'* (Childerhouse et al. 2003a, p.135)

2.3.2 Demand planning process failures: system nervousness and naïve interventionism

In practice, there are many factors which may cause the DP process to fail (e.g., Kerkkänen et al. 2009; Stadtler et al. 2015). These failures include but are not limited to the following (see Figure 7).

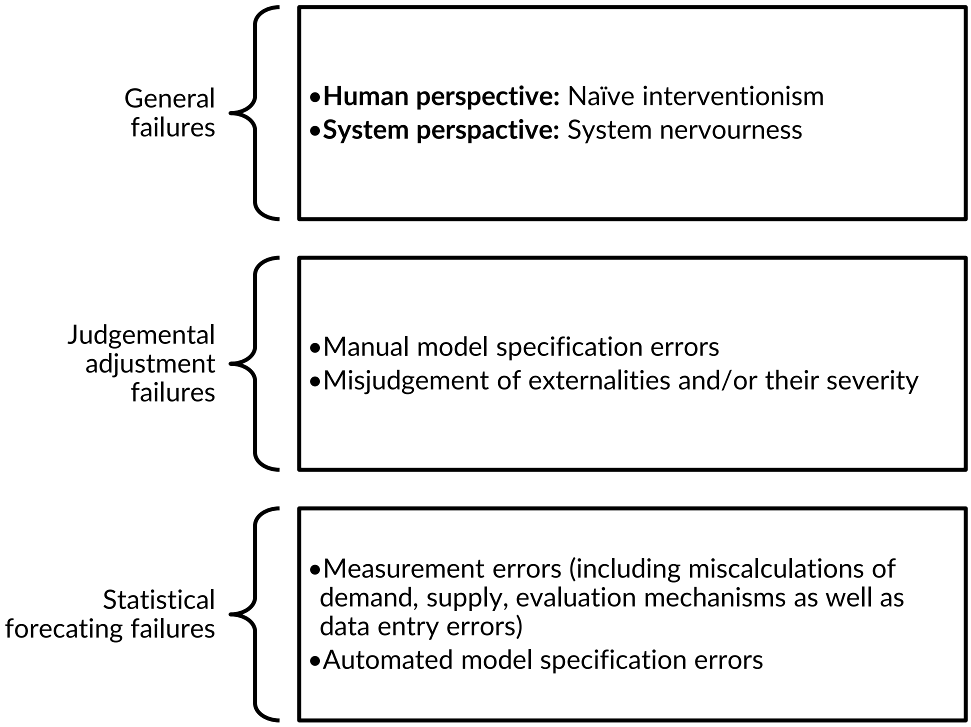


Figure 7 Main factors causing demand-planning failures

Even after considering potential failures of statistical forecasting and judgemental adjustment, there are still general factors that may cause DP to fail. Specifically, the systems in which humans operate are characterised by inherent (endogenous) instability often labelled *system nervousness*. Furthermore, even if the plan is perfectly formulated and accurately takes into account system nervousness, humans may cause DP failures by simply not following the plan, i.e., they are unable to follow the plan and constant desire to alter the plan due to observed small shocks and externalities may cause serious problems (Lim & O ’connor 1995; Goodwin & Fildes 1999). For example, Fildes et al. (2009, p.3) analysed 60,000 forecasts and outcomes of four SCs and observed that not only most of the forecasts were adjusted, the ‘*relatively larger adjustments tended to lead to greater average improvements in accuracy, the smaller adjustments often damaged accuracy.’* This inability to follow the plan and intervene will be called *naïve interventionism*.[6] The rationale behind naïve interventionism lies in the fact that managers have a tendency to intervene with the plan (earlier referred to as best efforts) and naivety comes from the lack of understanding of variation and knowledge. In what follows, the focus is on two general factors: System nervousness and Naïve interventionism.

2.3.3 System nervousness

Historically, nervousness was associated to the material requirement planning (MRP) systems and the two basic causes were uncertainty in supply/demand and variations in lot-sizing decisions (Whybark & Williams 1976). However, both causes and effects of the problem of nervousness is far greater in scope. The flow in SCs is usually managed either based on the demand of the first tier customer downstream (next immediate company in the chain) or on the demand of the end customer (ultimate consumer) in the total SC (Van Donselaar et al. 2000). Nervousness (in this case order instability) is a common problem in planning systems across the whole SC (Blackburn et al. 1985; Blackburn et al. 1986; Kadipasaoglu & Sridharan 1995). The assumption is that the opposite of system nervousness is system stability. System stability depends on plan stability. In planning systems, *‘…plan stability is affected by policy parameters’* (de Kok & Inderfurth 1997, p.55).

Nervousness in the form of order instability causes frequent plan changes leading to adverse effects which propagate through the SC in form of increased cost, reduced productivity, lower service level, and a general state

---

[6] The term ‘naïve interventionism’ is borrowed from Nicholas Nassim Taleb (2012) book *Antifragile*. Due to the nature of the publication, this source is used just as inspiration and not a reference to the argument of this research

of confusion in operations (Campbell 1971; Hayes & Clark 1985; Christopher & Lee 2001). Plan changes contribute to operational confusion and have negative impact on performance. *'Confusion (…) refers to managerial actions that disrupt the stability of the factory's operations'* (Hayes & Clark 1985, p.10). Confusion is also referred to as *'chaos effects'* in SCs (Christopher & Lee 2001, p.2). Lack of planning stability and continual changes to the plan by the system, drive significant short-term and medium-term adjustment efforts and undermine management's confidence in planning systems (Blackburn et al. 1985; de Kok & Inderfurth 1997).

Time horizon plays a major role in planning nervousness (Blackburn et al. 1986). Examining different planning heuristics (policies) in relation to the time horizon, Simpson (1999) identified that time-horizon sensitive logistics exhibit less nervousness than horizon-myopic ones.

Other factors rather than planning horizon is uncertainty in the timing or quantity of demand. Zhao and Lee (1993) looked at freezing the master production schedule under demand uncertainty for parameters such as planning horizon, freezing proportions, freezing methods and replanning periodicity. One of the findings on freezing the planning horizon in Zhao and Lee (1993) suggests that longer planning horizons worsens the performance under demand uncertainty conditions while improves the performance under deterministic demand conditions. However, Kadipasaoglu and Sridharan (1995) identify the freezing method as the most effective to reduce instability. . Comparing freezing different proportions of the planning horizon, it is suggested that freezing the whole planning period reduces instability under demand uncertainty (Zhao & Lee 1993). Considering the inverse relationship between period and frequency, if shorter planning periods increase instability, higher planning frequency contributes to instability as well. *'Higher replanning periodicity results in a lower total cost, schedule instability, and a higher service level under both deterministic and stochastic demand conditions. Less frequent replanning improves system performance'* (Zhao & Lee 1993, p.185).

There are several strategies to deal with system nervousness. No revision of decision within the planning horizon, also known as freezing the plan (Zhao & Lee 1993; Kadipasaoglu & Sridharan 1995). Freezing the whole planning period reduces instability under demand uncertainty (Zhao & Lee, 1993). This is sensitive to stock outs and must be combined with buffer stocks. One of the most effective to reduce changes but at a higher cost are buffer stocks which are also known as end-item safety stocks (Kadipasaoglu & Sridharan 1995). Finally, Ho (1989) suggests that enforcing the distinction between large and small changes affects nervousness.

2.3.4 Naïve interventionism

As it was mentioned earlier, Deming (1986) defended that best efforts can be damaging if carried out without understanding. The adopted designation in this thesis for efforts without understanding is naïve interventionism. The lack of confidence in the SC leads to actions and interventions by SC members who believe that they can do better by devising independent actions that undermine the overall performance of the SC (Christopher & Lee 2001). Similarly, unnecessary interventions upstream through the disregard of forecasts (even though they are often accurate) through the *'we know best'* syndrome (Mason-Jones & Towill 1998, p.19) which is one of the identified information flow problems encountered in the practice of SCM. Overconfidence is suggested to be one of the main reasons why decision-makers show strong preference towards their own judgement and the tendency to introduce judgemental adjustments (Kleinmuntz 1990). Moreover, overconfidence can morph into a form of grandiose syndrome, taking an example from OEMs in Childerhouse et al., (2003b, p.141), that all players *'…felt that they were pro-actively leading world-class supply chains and saw no need to change their ways.'*

Furthermore, managers show the tendency to game the system and 'outsmart' the competition in tasks related to DP. Sterman and Dogan (2015) run an experiment based on the Beer Game and found that managers were unable to resist the urge to intervene. They exhibited hoarding and phantom ordering even when they had perfect visibility over the rest of the SC.

Considering how important forecasting is for planning and decision-making, managers are also infamous. *'Forecast errors are directly related to required safety stocks, while frequent adjustments of demand forecasts can lead to dramatic changes in plans (i.e., nervousness)'* (Stadtler et al. 2015, p.582). Managers intervene with their statistical forecasts (Fildes et al., 2009) and are generally over-confident in the accuracy of their adjustment (Brenner et al. 1996; Lawrence & Sim 1999; Moore & Cain 2007). One of the suggested reasons behind tweaking unnecessarily the statistical forecast (Lim & O 'connor 1995; Goodwin & Fildes 1999) as well as preferring judgement over quantitative models (Lim & O'Connor 1996) is incorrect calibration of user's confidence (Fildes et al. 2006).

The separation between statistical and judgemental tasks is poorly implemented (Fildes et al. 2006). The two components are largely confused as planners often disregard the statistical analysis outputs of the regular component as well as mistaking noise for signal perceiving inexistent regularities and apparent patterns (Goodwin & Wright 1993; Harvey 1995; Lim & O 'connor 1995; Goodwin & Fildes 1999; Sanders & Ritzman 2001).

In practice, inputs from judgement adjustments and statistical forecast are hard to separate. For example, Fildes et al. (Fildes et al. 2006) points two instrumental cases in which managers made adjustments to the time series in terms of level, trend or seasonal factors before the statistical method attempting to improve its accuracy. This way the output of the statistical forecast is already greatly altered by judgement.

Some interventions can be positive. Expert judgement can greatly improve poorly performing forecasting models (Franses & Legerstee 2013). For example, expert input can greatly improve intermittent demand forecasting (Syntetos et al. 2009). However, the majority makes many small adjustments to statistical forecasts which leads to waste time and often reduces accuracy (Fildes et al. 2009). This should not be surprising since it is known that planners, as human beings, are prone to decision making biases (Carter et al. 2007; Tokar 2010; Sterman & Dogan 2015)

There are several ways to affect DP performance through policy affecting behaviour. First is necessary to recognise the central role of policy in DP. For example, amongst several techniques used to improve judgmental forecast are training with feedback (Goodwin et al. 2004), taking advice (Goodwin et al. 2012) decomposing the forecasting process and making a separate estimate for each component (Edmundson 1990; Webby et al. 2005) and combining predictions from several forecasters (Clemen 1989).

2.3.5 Demand planning performance and individual differences

So far, the review focused on literature around manager's behaviour as a group. However, it is incorrect to take a one-size fits all approach, since results in performance are never homogeneous (e.g., Moritz et al. 2013; Strohhecker & Größler 2013). It is important to consider the potential relationship between naïve interventionism and individual differences.

Anecdotally, it is suggested that personality types (individual differences) influence planning performance (Lapide 2007). Previous research on dynamic decision making shows performance variations across subjects, e.g., in an experiment by Hey and Knoll (2007, p.8) half of the subjects did not plan ahead, replicated in a follow up experiment (Bone et al. 2009). Considering broadly decision performance which has been repeatedly studied in different disciplines and it is generally accepted that individual differences play a significant role in decision making (Franken & Muris 2005).

Decision patterns and biases have been often observed across the subject sample, variation at the individual level has been often ignored. Individual differences in OM literature have received relatively limited attention, and individual differences in DP decision making is even more rare (e.g., Moritz et al. 2013; Strohhecker & Größler 2013). The few existing studies considering individual differences are mostly focused on a demand-supply balancing problem, e.g., the newsvendor or newsboy problem where participants make decisions facing uncertain demand. Bolton et al. (2008; 2012) as well as Wachtel and Dexter (2010) observed in experimental conditions how professional background (exposure) matters for performance in an DP problem. Traits such as impulsiveness negatively affect demand and supply balancing in experimental conditions (Ockenfels & Selten 2015) and this tendency varies significantly between individuals (Bolton & Katok 2008). Moritz et al. (2013) also run experiments specifically designed to explore individual differences and the task of balancing supply and demand, detecting cognitive reflection as significant positive predictor of performance. De Véricourt et al. (2013) reports on significant gender differences in terms of risk taking showing that male subjects perform better when planning for a high margin product. Finally Strohhecker and Größler (2013) identify high intelligence as the strongest predictor of high performance in a DP type setting. Individual differences in context of work flows (manufacturing processes) are suggested to play a significant role on variability in performance depending on the operating policy (Doerr et al. 2004).

2.4 Humans and systems: supporting demand planning decisions

2.4.1 Resisting the 'machines'

People and in particular experts have a long history of resistance to '*machines*' taking over their decision making processes (e.g., Meehl 1954; Silver 1991). '*Machines*' are used as a broad collection of statistical algorithms, analytical heuristics, decision support, restrictive policies or artificial intelligence. To illustrate, consider the famous opposition between clinicians and (over performing) statistical algorithms for diagnosis (Meehl 1954). A famous example of this opposition is the negative reaction to Apgar's (1953) score. Apgar (1953) proposed a systematic method to evaluate new-born infants using fast and frugal simple heuristic. It is based on a five-item list and three scores. Although it significantly contributed to a lower infant mortality, clinicians showed great criticism. At the time clinicians claimed it was cold, impersonal and too simplistic. A more recent example of how experts reacted to Ashenfelter's (2008) accurate formula to predict the quality and prices of Bordeaux wines. Oenophiles' reactions ranging …*'somewhere between violent and hysterical.'* (Kahneman 2012, p.222).

Regardless of clear resistance to the artificial, DSS developed both in academia and in practice, are becoming essential for business and human activity in general (Power et al. 2015). One of the reasons for such development is the fact that DSS has the '… *ability to relax cognitive, temporal, and economic limits of decision makers – amplifying decision makers' capacities for processing knowledge which is the lifeblood of decision making'* (Burstein & Holsapple 2008, p.9). This brings obvious advantages and from an information systems perspective is essential to support decision making in the modern world.

Management practice, however, despite general adherence and investment in IS and DSS technologies, still shows resistance using their own judgement to override and adjust DSS outputs (e.g., Fildes et al. 2009; Goodwin et al. 2011). Hence, support systems should allow better identification of judgemental intervention opportunities as well as enable managers to intervene when it is most appropriate (Fildes et al. 2006).

2.4.2 System guidance and restrictiveness

Systems can support different aspects of decision making in many forms and via varied modes. Following Silver (1991) a system supporting decision making can be designed to influence decision-makers via *'decisional guidance'* and *'system restrictiveness'* illustrated in Figure 8. The relationship between guidance and restrictiveness can be described as a trade-off and the two are not mutually exclusive. On one hand, a system can offer few decision making processes, which translates as a restricted set of possibilities. Such situation requires minimal guidance because alternatives are restricted *a priori*. On the other hand, unrestricted decision processes offer a wide range of possibilities and combinations, requiring greater levels of guidance. When considering systems to support decision making, it is necessary to decide between guiding the decision making process, restricting it or to do neither (Silver 1991).

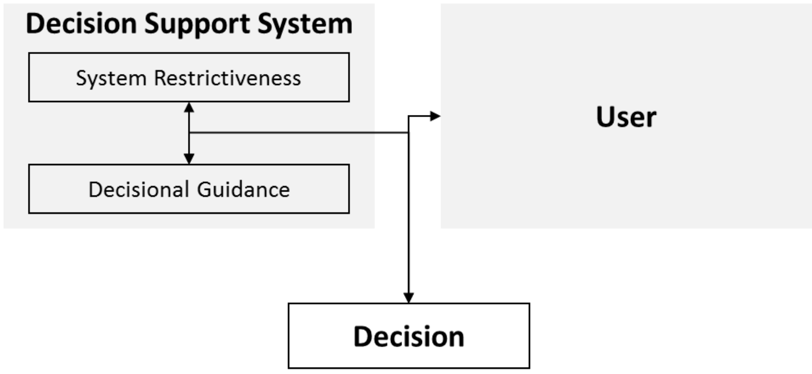


Figure 8 Decision support guidance and restrictiveness (based on Silver 1991)

Decisional guidance

Decisional guidance is defined by Silver (1991, p.107) as *'how a decision support system enlightens or sways its users as they structure and execute their decision making processes – that is, as they choose among and use the systems functional capabilities.'* Decisional guidance is significantly different from simple mechanical meta-support. Typical mechanical meta-support solely helps users with operating the system's features, e.g., which buttons to push when. In contrast, decisional guidance is 'smart' and goes beyond simple mechanical interaction by providing the decision-maker with enhanced information-processing capabilities of the system. To illustrate, mechanical meta-support provides the user with the complete inventory of forecasting methods while decisional guidance will highlight strengths and weaknesses of each method given the situation.

Guidance can influence decision-makers deliberately or inadvertently (Silver 1991; Montazemi et al. 1996; Parikh et al. 2001). Inadvertent guidance happens when the consequences are unintended, as defined by Silver (1991, p.107) *'…is an unintended consequence of the system's design and is not planned by the system designer'*. An example is when the system provides an illustrative example of the task and the decision-maker anchors on the values and results of the illustrative example rather than only considering the relevant information for the actual task (Frederick et al. 2010).

The proposed typology for deliberate decision guidance by Silver (1991) distinguishes three main types: target; form and; mode.

Table 4 Typology of deliberate decision guidance (adapted from Silver, 1991)

Dimensions	Question	Guidance
Target	What is the target? – Aspects of decision making addressed	Structuring the process (Choosing Operators)
		Executing the process (Using Operators)
Forms	What is the form? – What guidance is offered	Suggestive
		Informative
Modes	What is the mode? – How the mechanism for guidance works	Predefined
		Dynamic
		Participative

The main consideration of deliberate decisional guidance is its target, i.e., what aspects of decision making are addressed (Silver 1991). The suggested two target aspects are structuring and execution of the decision making process. Guidance for structuring the decision making process consists of choosing which operators to use and in what order. Guidance for executing the decision-making process focuses on how users interact with the chosen operators, namely how decision-makers perform predictive and evaluative judgements. For example, the decision support targeting process structuring can provide its user with recommended steps to follow and which information to consider to determine what supply strategy is most appropriate for a given product. Further, decision support targeting the execution of the decision making process can provide its user with an evaluation of his choices through what-if simulated scenarios.

Considering the form and what guidance is offered, the two forms of deliberate decisional guidance are suggestive and informative guidance (Silver 1991). Informative form of decisional guidance is characterised by being unbiased, providing information that is relevant in a neutral way, i.e., without suggesting any specific course of action. In contrast, the suggestive form of decisional guidance focuses on suggesting what the decision-maker should follow. To illustrate both forms, informative guidance detects the life-stage of a product while suggestive guidance would make recommendations for what would be the most appropriate forecasting model based on the detected stage.

Regarding how the mechanism for guidance works, three modes are suggested: predefined; dynamic and; participative (Silver 1991). The predefined mode of decisional guidance is created a priori. An example of pre-defined decisional guidance is a checklist, while the user can choose to disregard or simply forget some of the steps to make a decision; it is known in advance that some specific considerations are necessary to make a balanced decision. The dynamic mode of decisional guidance depends on how the user interacts with the system; hence, it is not created a priori. For example, when the user makes adjustments the system can display what has been his performance in the past with similar actions against a what-if scenario when nothing was adjusted. The last mode is participative guidance that is a form of customisable mode of guidance. In the participative mode the user can input his preferences and objectives so the system responds with suggestions (Jiang & Klein 2000).

The most common type of deliberate decisional guidance in current systems supporting DP is predefined informative guidance (Fildes et al. 2006). These systems provide ready-access to time-series data in form of graphs or tables along with KPIs and statistical forecasts and respective errors. The general assumption is that the decision-maker should be presented with rich information. Some less common situations include comparative analysis and benchmarking of different decisions going in some cases further by providing recommendations and respective explanations of most appropriate methods.

System Restrictiveness

Restrictiveness is defined by Silver (1991, p.108) as *'The degree to which, and the manner in which, a DSS limits its users decision making processes'*. One of the possible ways to restrict the user is by displaying only selected information or forcing the user to follow a pre-set process with a strict number of options denying certain actions.

One possible example of restriction is displaying data either as a graph or as a table. In some cases tables work better than graphics as people are more prone to detect false patterns in graphical displays (Carey 1991; Hwang 1995; Harvey & Bolger 1996). In other cases graphics help understanding dimensions better, e.g., reducing the duration neglect bias (Liersch & McKenzie 2009). Alternatively, the order of display and action also plays a significant role on the quality of the judgement (Schkade & Kleinmuntz 1994; Healy 2000; Theocharis & Harvey 2016). For example, people tend to overestimate the relative weight of most recently presented information, i.e., the effect of recency (Arnold et al. 2000; Tan & Ward 2000; Ashton & Kennedy 2002).

Another example is restricting the time-horizon of visible data. Finally, in order to reduce instability under demand uncertainty, freezing the whole planning period so managers cannot make changes to the plan (Zhao & Lee 1993; Kadipasaoglu & Sridharan 1995; Zhao & Lam 1997; Xie et al. 2003)

In some cases, restricting the system's use based on the user's background and ability is desirable. For example, in cases when managers lack training in forecasting it is better to make unavailable the choice of different statistical models or adjustment of its parameters (Fildes & Hastings 1994).

Forcing the user to work on many elementary sub-tasks rather than one complex task takes advantage of decomposition methods and can often lead to more accurate judgement (Edmundson 1990; Srivastava & Raghubir 2002; Abdellaoui et al. 2005). However, Fildes et all., (2006, p.355) warns that *'absolute restrictiveness can be dangerous if it is wrongly applied'* and that too much decomposition can make the judgement task unpractical.

System restrictiveness is not necessary forcing the user to certain processes. It can be implemented by means of making some 'preferred' processes easier to use while leaving other more complicated and bias-prone processes less accessible (Payne et al. 1993; Goodwin et al. 2011). Such user manipulation takes advantage of the accuracy-effort trade-off (Payne et al. 1993; Kruglanski & Gigerenzer 2011). The two possibilities to lead the decision-maker are either reducing the effort for desired paths or increasing it for the undesired ones. By minimising the effort or even automating a desired path, the decision-maker is more likely to follow it (Todd & Benbasat 1999). For example, decision-makers often follow the default as it is the least effortful path, i.e., rely on the 'default heuristic' (Azar 2014). Setting up the system in a way that optimal parameters are pre-set by default is likely to lead to better performance. Similarly, adding extra steps (extra effort) to undesired decision paths in a system can prevent damaging behaviour such as manual interventions and deviations from the system's recommendation. Adding an extra step to the process of making a manual adjustment (e.g., request the change or elicit the reasons to do so) has shown to effectively reduce the number of unnecessary and damaging adjustments (Goodwin 2000). While such procedure has shown to successfully reduce the number of negative interventions, it has not prevented the decision-makers from making adjustments when they were necessary and beneficial.

Table 5 Means of system restrictions (based on Fildes & Beard, 1992; Fildes et al. 2006; Goodwin et al. 2011)

Means	Making it easier	Making it harder
System Defaults	Making systems' recommendations available	Creating extra effort to deviate from the systems' recommendation
Data-quality	Allow easy access to adjustment when data is missing or when exceptional situations are detected	Make access to manual adjustments harder when the data is available and no abnormalities are detected
Product life-stages	Identify the time-series type to encourage adjustments in stages of new product launch, planned promotions, intermittent demand or declining demand.	Identify the time-series type to discourage adjustments in stages of stability such as steady growth and maturity
Signal Monitoring	Report exceptions (e.g., promotions) and abnormal deviations in real time to identify appropriate moments for manual interventions	Report stability and reduce the amount of feedback on small variations in situations of normality to avoid misjudgement (e.g., apply smoothing)
Aggregation level	Allow forecasting both at individual and aggregate levels to drive down the product hierarchy common aggregate effects	Prevent only individual level forecasting
Method comparison	Allow easy comparison of different forecasting methods on a test module	n.a.
Method evaluation	Allow easy data split into estimation and test-data to evaluate the methods' performance	Prevent choosing different models or adjusting its parameters without testing its relative performance
Systems' Menus	Place preferred options more accessible (e.g., first level menus)	Hide undesired options out of immediate reach (e.g., second level menu)

2.4.3 Supporting demand planning decisions

To address decision-maker's overconfidence the system should be designed in such a way that it provides guidance to improve calibration. Better calibration brings closer the perceived accuracy and the real accuracy hence enabling a better choice of strategy and better effort application (Payne et al. 1993).

To observe the broader context of DSS with business and management in a way it includes its use, it is necessary to consider business and management policies beyond DSS (as technology) into systems supporting DP decisions. A policy is a course or principle of action and it defines the scope within which decisions are taken acting as guidelines developed by the organisation. Considering DSS on its own, *'DSSs probably support only a small percentage of all decisions made in organizations'* (Power et al. 2015, p.3). In contrast, business and management policies lie at the heart of management science and by default affect most of decisions made in organisations, including those involving DSSs.

DP policies can both act as guidance or restriction of decision-making processes within an organisation. Restrictive policies will limit while guiding policies will direct decision-makers. For example, restrictive DP policy can partially restrict judgemental adjustments to the statistical forecast if these adjustments are smaller than a certain threshold (*e.g.*, Fildes et al. 2009). Alternatively, a decisional guidance policy could direct managers to acknowledge the possibility of errors and providing them with additional information (e.g., Fildes et al. 2006).

2.5 Chapter summary

The process of DP is of critical importance in the SC. DP is sensitive to planning horizon, subject to data availability and its quality, and finally its process. The process of DP is therefore subject to planning horizon, policy and DSS (Figure 9).

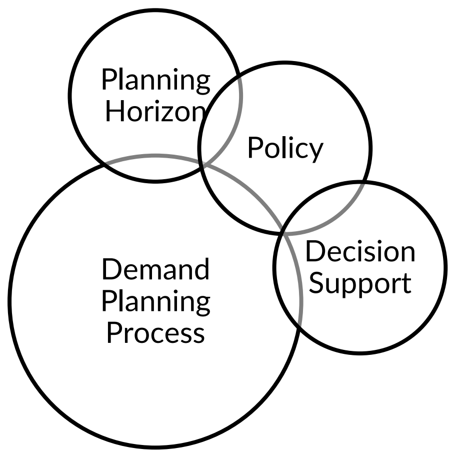


Figure 9 Demand planning process base framework

The process of DP involves two major stages, one is statistical forecasting and the second is the judgemental adjustment of the forecast to derive the final forecast that is fed into the master plan. However, this process is challenging and is subject to several sources of failure, being the two most important the system and the human (manager). System nervousness and naïve interventionism can significantly affect DP performance. Moreover, understanding manager's behaviour is challenging due to individual differences that can also potentially explain heterogeneity in the results.

Supporting DP decisions, the system can either be neutral, restrict or guide the decision-maker. The implementation of the system can be made in form of DP policies and integrated in the DSS. These systems have the ability to compensate for human limitations and improve DP performance.

Despite the fact that both naïve interventionism and system nervousness are acknowledged in many different forms across the OM and OR literature, there is lack theoretical underpinning to explain and predict this behaviour. As a result, the rising field of behavioural operations has been successfully borrowing theory from behavioural economics, psychology and neuroscience (e.g., Bendoly et al. 2009; Katsikopoulos & Gigerenzer 2013). The following chapter focuses on the underpinning theory of decision making which is then used to develop the hypothesis and theoretical framework for this research.

3 Underpinning theory and hypothesis

3.1 Introduction

DP has been described in terms of its process, performance and support with emphasis on the dynamics between managers and systems. Yet, it is also important to consider the underlying mechanisms of DP.

Table 6 Chapter 3 structure

Section	Overview
3.2 Rationality and decision making	This section provides a historical overview of the evolution of rationality and decision-making. An overview of models of human behaviour is provided, explaining the assumption of loss aversion in cumulative prospect theory and the concept of mental accounting which as central to this research.
3.3 Myopic Loss Aversion	Myopic loss aversion is explained in detail along with the main structure of the conceptual framework that relies on mental accounting. The main hypothesis is derived.
3.4 Human and system: individual differences	The interaction of human and systems is where managers are required to make choices. It is expected that individual differences explain the variation in performance in conditions of choice. Following an overview, the focus is on experience and naïve interventionism. Personality is described in terms of the Big Five along with specific constructs. Hypothesis on individual differences are derived.
3.5 Complete conceptual framework, testable hypotheses and predictions	This section joins previously developed components of the framework and summarises the hypothesis.

This chapter (summarised in Table 6) focuses on the development of theoretical framework for explaining planners' behaviour using several important concepts from behavioural science. These concepts (especially, the concepts of myopic loss aversion and individual differences) are rooted in both economics and psychology. In this chapter, models of rationality are outlined and the role of rationality in research on decision-making is summarised with a focus on myopic loss aversion and individual differences. The aim of this chapter is not to discuss rationality, decision-making, and individual differences exhaustively. Instead, these concepts are used instrumentally to inform an overarching analytical framework (developed in this thesis) and to define respective hypotheses in order to explain underperformance in DP. The chapter closes with a summary.

3.2 Rationality and decision making

Rationality has been celebrated as one of the highest achievements of the human species. Traditional view on rationality and decision-making implies that humans rely on logic, statistics and heuristics. One of the first ideals of human reasoning and inference was *logic* as defined by Aristotle (384 BC – 322 BC). Later in the 17th century, *logic* was replaced by the *probability theory* which acknowledged the fundamental uncertainty of human conduct (Daston 1980). Up until mid-19th century, probability theory was considered the ideal way for describing common sense through calculus (Laplace 1902- originally published in 1814). Probability theory enabled the development of normative and descriptive models of decision making (Savage 1954).

Models of decision evolved as different theories of decisions and perspectives on probability developed. The most recognised interpretations of probability are classical, frequentist, logical and subjective (Surowik 2002). This significantly influenced mathematical theories of decision making which are still being used today and are widely adopted by researchers in social and natural sciences (Vranas 2000; Bowers & Davis 2012). The classical view on decision making using probability theory articulated by Daniel Bernoulli and Pierre-Simon Laplace was that the probability of an event is the ratio of the number of favourable cases to the total number of cases being equally weighted. This view was followed by the frequentist idea common to statistical methods of hypotheses testing. The frequentist view is that the probability of an event equals the frequency of its occurrence in repeated trials. Another view on probability was developed by John Maynard Keynes, as *Logical* or *Objective* probability. This probability is connected with statements and can be deduced from the truth-value of the premises of the statement for which it is inferred. Finally, another widely accepted concept of probability adopted by Bruno de Finetti and Leonard Jimmie Savage was the *Subjective* probability. According to this view, probability is a subjective degree of conviction related to a single event or repeated events and measured by psychometric methods (e.g., observation of gambling).

Subjective consideration of probability was introduced by means of one of the greatest contributions to statistics, the Bayes' theorem (Savage 1961). It has been adopted as one of the main models of human reasoning (Chater et al. 2006). This theorem provided a foundation for a number of models, e.g., the Adaptive Control of Thought theory (ACT-R) proposed by Anderson (1996). CAT-R is a cognitive architecture aiming at defining the most basic and irreducible perceptual and cognitive operations of the human brain (Anderson 1996). Yet, empirical tests showed that such a model better describes mathematical and computer programming algorithms rather than human behaviour.

One of the first decision making theories was proposed by Paul Samuelson (1938). He defined utility as desired level of satisfaction obtained from available decision strategies and assumed that an agent's goal is to maximise utility using a rational decision model. This marked the rise of the 'perfectly rational' economic agent or 'homo economicus'. This agent had perfect information and applied principles of rationality to make an optimal decision. An assumption of perfect rationality was important for the development of simple and tractable models of behaviour such as expected utility theory (Neumann & Morgenstern 1947). However, evidence from empirical research led researchers to challenge the concept of rationality by showing that 'well-behaved' axioms and assumptions of theories which had human rationality at its core fail in practice (Simon 1955; Simon 1969; Tversky & Kahneman 1974; Kahneman 2003; Shah & Oppenheimer 2008; Hilbert 2012). One of the main and rather unrealistic assumptions about the classical rational decision-maker was that he has a stable system of preferences as well as possesses advanced computational skills to find the highest possible point (optimal solution) on his preference scale (Simon 1955). However, despite the fact that the extent of the work of Maurice Allais (1953), Daniel Ellsberg (1961), Sarah Lichtenstein and Paul Slovic (1971), as well as Daniel Kahneman and Amos Tversky (1974; 1983; 2003) challenged the classical view on rationality, the normative kernel is still present. The adoption of the classical view, i.e., normative approach to reasoning, is illustrated by the fact that reasoning errors are defined as the deviation from the norm governed by the laws of probability and statistics.

From Aristotle's view on logic (384 BC – 322 BC) to heuristics defended by Gerd Gigerenzer (2001), rationality, behavioural models, and decision making have been hot topics. However, the behavioural science community is still far from converging to one unifying theory of decision making (Gigerenzer 2008). From 'homo economicus' (term suggested by Richard Thaler 2000) to 'homo sapiens', the evolution of how rationality is perceived, changed throughout the development of behavioural science (Thaler 2000). The latest developments suggest that there is no single right view on the subject. Humans rely on all those different ways of thinking depending on the circumstances (decision context), which amounts to an ecological view of rationality.

3.2.1 The unrealistic view of humans as perfect optimisers

An ideal of rationality beyond human abilities dates back to even before the times of John Locke (1690) when the perspective of an omniscient God in a certain and deterministic nature was contrasted with humans living with uncertainties and inconsistencies. God was taken as the ideal of a super-intelligence, which Pierre-Simon Laplace (1814 p.1325) characterised as:

'*...an intelligence which could comprehend all the forces of which nature is animated and the respective situation of the beings who compose it – an intelligence sufficiently vast to submit these data to analysis ... nothing would be uncertain and the future, the past, would be present to its eyes.*' (Laplace 1902 - originally published in 1814, p.1325)

This point of view is still represented today in many decisions models such as Bayesian reasoning or expected utility maximisation, considering that, when given unlimited time, boundless knowledge and unconstrained computational abilities human reasoning is well described under this divine light. Despite the fact that the 'old' view of 'unbounded' (or divine) rationality was dropped in mid-twentieth century due to its relation to the theological doctrine, a similar perspective with rational utility-maximizing human at its core took over. The new perspective labelled *optimisation* (specifically, *constrained optimisation*) assumes that humans can be perfect optimisers. They can do so when the decision context allows them to be rational and when complexity of the decision problem is manageable to make appropriate calculations.

Constraints of the decision environment (or constraints of *decision architecture*) can be understood as, for example, having a finite amount of time, knowledge, attention, resources to spend on a given decision. One main difference between perfect (unbounded) rationality and all other visions of rationality is that under perfect rationality it is assumed that information search can go on endlessly while under bounded rationality this process is limited. The concept of limited information search consequently brings in the need for having a stopping rule, i.e., when to stop looking for information. Optimisation from the optimisation under constraints point of view is now focused on finding the stopping rule that *'optimises search with respect to the time, computation, money and other resources being spent'* (Todd & Gigerenzer 2000, p.729). The main rule holds that the search stops when the costs outweigh benefits, assuming that the mind is able to calculate the benefits and costs of searching for additional pieces of information.

The idea of optimisation under constraints turns out to be even more demanding from agents' computational ability than the classical idea of unbounded rationality (Vriend 1996). Paradoxically, the assumption for a limited search for information is that the mind has unlimited time and knowledge to evaluate the trade-offs of further information search (Todd & Gigerenzer 2000).

3.2.2 The need for small and large worlds

In order to keep using rational models of behaviour which are based on perfect information and work around any informational limitations, Savage (1961) introduced the concept of small worlds (Figure 10). On the one hand, the idea of small worlds enables most of the classical analysis: it described situations where 'optimal' solutions to a problem can be determined because all relevant alternatives, consequences and probabilities are known and where the future is certain. This means that in a small world it is possible to hold perfect knowledge ('god-like' knowledge) and the conditions for rational decision theory are satisfied. On the other hand, a large world (or real world scenario) describes situations of uncertainty that violates the conditions for rational decision theory where part of decision-relevant information is unknown and has to be estimated from smaller samples.

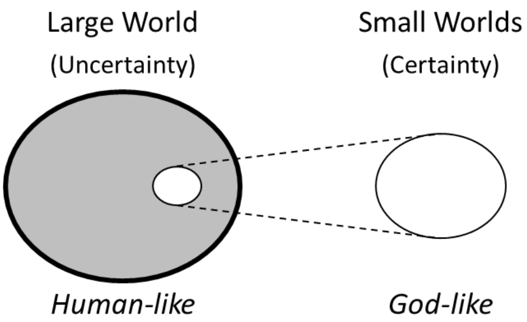


Figure 10 Concept of Small and Large worlds (based on Savage 1961)

Despite an obvious misfit between the idea of humans as perfect optimisers and reality (large world), the view of the decision-maker as the '*homo economicus*' remained accepted within the context of the small world (Figure 11) where everything that does not apply belongs to the large world.

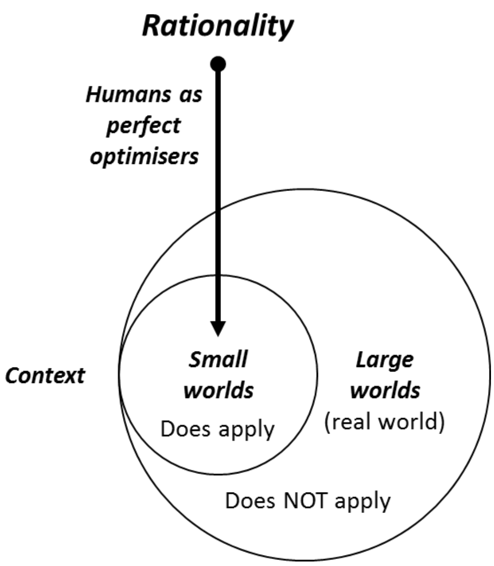


Figure 11 Classical view of rationality reduced to small worlds

It is considered that it is inappropriate to apply small-world norms of optimal reasoning to large worlds (Binmore 2009). Therefore, since conditions for rational decisions are not satisfied in large worlds (real world scenarios), one cannot expect that models of rationality will provide the *'right'* answer and consequently rational expectations theory is not *'taken seriously outside academic circles'* (Soros 2009, p.6).

It is critical to understand the implications of small and large world considerations in practice. Situations where small world theories (economics and behaviour) were applied to large worlds sometimes led to disasters as even slight deviations from the model do matter, e.g., 2008's financial crash where almost perfect information in the form of high volumes of data was assumed to be the same as perfect information (Stiglitz 2010).

The segment of study of rationality in small worlds is essentially the study of constrained optimisation popularized by Gerd Gigerenzer (1991). It introduces the concept of limitations into the study of decision making while still assuming that there exists an 'optimal' solution. This was one of the most widely known attempts of making the 'homo economicus' more human. Stigler (1961) argues that humans do not have all the information necessary to make the perfect decisions available instantly, so they must search for it. This search is not free. There is a resource cost to the decision-maker, e.g., time and money. The ideal of rationality is still present while the main difference from previous decision theories is that the search for more information is stopped when the benefits no longer exceed the cost of further search. In a way, this information/effort trade-off is similar to Simon's (1987) satisficing heuristic which implies that the decision maker looks for 'good enough' solutions when the effort and the decision accuracy are balanced according to the situation. The difference is that Simon (1987) argues that models of rationality should represent actual cognitive capacities of humans, therefore accounting for natural limitations in cognitive capacities such as memory, attention, knowledge.

At first, this idea might sound reasonable, except it has one inconsistency. Humans are expected to be able to calculate the optimal stopping point in the decision making process (similar to the break-even point in economics). While finding a breakeven point in a linear problem might be tangible, most of the real-world scenarios are not linear in their nature. Hence, such operation can easily be more demanding both psychologically and mathematically than assuming that people have unbounded rationality (Vriend 1996). This means that such theory is built on rationality norms, assuming that humans are perfect optimisers, therefore making such approach only applicable to small worlds. The paradox of optimisation under constraints lies in the fact that a limited search for information relies on a mind that has unlimited time and knowledge to evaluate the cost-benefit of further information search (Todd & Gigerenzer 2000).

3.2.3 Model of human behaviour: deterministic versus stochastic theories and Cumulative Prospect Theory

Decision support in many organisations is informed by the developments in decision theory – a field combining methodology from economics, psychology, and neuroscience. Decision theory provides a theoretical underpinning for research of human behaviour. At a very general level (refer to Figure 12), theories and models can be partitioned into three groups: bounded rationality models, deterministic theories, and stochastic theories. Bounded Rationality Models do not allow for a clear prediction of behaviour and are mostly used to explain observed decisions. A deterministic theory makes a prediction about human behaviour and this prediction cannot be altered by the features of decision environment whereas stochastic theory allows for such a possibility (Loomes & Sugden 1982; Hey 2005; Loomes et al. 2008; Starmer et al. 2009; Blavatskyy & Pogrebna 2010).

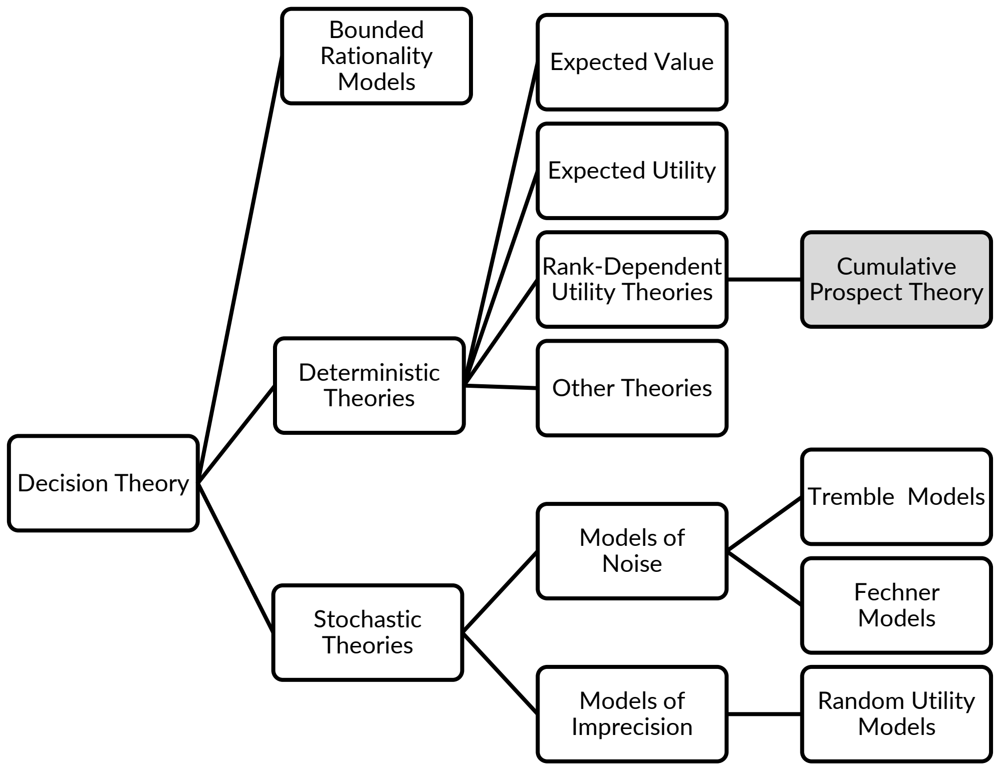


Figure 12 Decision Theory as a building block for Research Methodology

Bounded rationality models

Models of bounded rationality form a separate group in decision theory. Bounded rationality conceptualises empirical evidence that human rationality is often limited by 3 variations of constraints: (i) tractability or complexity of decision problem, (ii) cognitive ability of a decision maker; and (iii) time available for making a decision (Kahneman 2003). Bounded rationality models, therefore, suggest that instead of using utility-based calculations, people apply simple decision rules (such as simple rules of thumb). One of the possible manifestations of bounded rationality is the use of simple heuristics and biases (Tversky & Kahneman 1974), i.e., simple decision rules which allow to significantly decrease the computational burden which may or may not be present in other types of models.

It is important not to confuse bounded rationality as a strand of decision-theoretic modelling and heuristics and biases as concepts. While bounded rationality refers to a particular way of thinking about human behaviour, heuristics and biases as concepts can be stand-alone phenomena. These phenomena could be combined with deterministic or stochastic theories to explain behaviour. For example, one may use Expected Utility Theory (EUT) and 'status quo' bias to explain why in some situations people have a preference to keep to current state of affairs and resist change (Masatlioglu & Ok 2005).

The father of bounded rationality is considered Herbert Simon. He devoted most of his research to investigating how people perform in situations where information is incomplete and the requirements for neoclassical rationality cannot be met. One of the central ideas of bounded rationality as proposed by Simon (1987) is that decision-makers are limited in their cognitive abilities, e.g., limited knowledge, attention, and memory. Simon (1990, p.6) states that:

'Because of the limits on their [computers and the human brain] computing speeds and power, intelligent systems must use approximate methods to handle most tasks. Their rationality is bounded.'

Additionally, to understand decision making it is necessary to look at not only actors' computational ability (cognitive ability) but also at the task context or environment (Simon 1992). Therefore, given that large world situations have constraints such as variability, uncertainty, redundancy, and limited sample sizes (Todd & Gigerenzer 2007) supplemented by the limited resources and limited cognitive abilities of the actor, the resulting decision making rule is based on heuristics. Heuristics, in turn, aims at producing 'good enough' decisions rather than 'perfectly rational' decisions. Shah and Oppenheimer (2008, p.209) defined that heuristics rely on one or more effort reduction methods such as:

1. Examining fewer cues
2. Reducing the difficulty associated with retrieval and storing cue values
3. Simplifying the weighting principles for cues
4. Integrating less information
5. Examining fewer alternatives

Bounded rationality brings the idea of humans using mental shortcuts (heuristics). Heuristics are also a central element of procedural rationality and which Simon (1990, p.11) defines as *'...not optimising techniques, but methods for arriving at satisfactory solutions with modest amounts of computation.'* The 'not optimising' technique and 'satisfactory' solution instantly suggest that heuristics provide solutions that fall below optimal. The terms 'suboptimal' and 'shortcuts' bring the idea that solutions provided by heuristics are always below the optimum (considering that there is one) which is in some cases a misconception when judging heuristics (Gigerenzer 2008). The usefulness of heuristics can be justified by two main perspectives as defended by Gigerenzer and Gaissmaier (2011). First that heuristics are ecologically rational (*i.e.* they fit real-life situations). Second perspective is the accuracy-effort trade-off that is discussed below.

The accuracy-effort trade-off can be better understood when considering effort as the 'cost' of making a decision. Since an individual can exert only a limited amount of effort, one can consequently accept 'good enough' rather than optimal decisions in most circumstances. This relationship between the quality of the decision (accuracy) and cost (effort) usually exhibits a non-linear pattern (a power distribution generally) as illustrated in Figure 13.

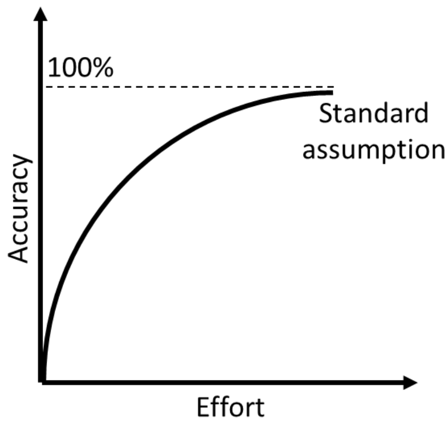


Figure 13 Accuracy-effort trade-off

The positive aspect of this non-linear trade-off relationship is that little effort is necessary to make relatively accurate decisions up until a certain point where the situation inverts and little improvements in accuracy come at great effort expense. This is similar to the Pareto Principle (also known as 80/20 rule) first described by Pareto and Page (1971). Heuristics in this case may help decision-makers to save effort at an expense of accuracy (Payne et al. 1993; Shah & Oppenheimer 2008).

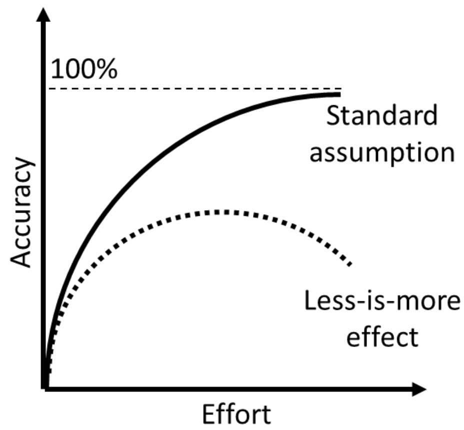


Figure 14 Accuracy-effort trade-off and Less-is-more effect

The accuracy-effort trade-off has been deemed as a universal rule in cognition, where more effort results in more accuracy. However, such relation is not always the case, i.e., simple decision models (heuristics) sometimes outperform complex ones. This is illustrated in Figure 14 with the *Less-is-more* effect (Gigerenzer & Brighton 2009). One of the first examples of the contradictory effect has been observed in an experiment by Gigerenzer and Goldenstein (1996). The experiment consisted of asking students in the U.S. and in Germany about the size of the population of cities in the U.S. and in Germany. Students were more accurate responding about overseas cities than their homeland. The interpretation was that participants using less information to make the inference outperformed others more familiar with the cities who automatically considered a higher range of cues (Gigerenzer & Goldstein 1996; Goldstein & Gigerenzer 2002).

Deterministic decision theories

The main advantage of a deterministic theory is that it provides a clear and easy prediction of decisions. For example, one of the oldest deterministic decision theories is Expected Value theory (EV). In application to binary choice decisions, it says that when people evaluate two risky options A and B they simply choose the option which give them the highest utility (or satisfaction) which is calculated as the probability of each outcome (stake) multiplied by the value of each outcome (stake). An alternative deterministic theory is Expected Utility Theory (EUT). EUT was developed in response to the St. Petersburg Paradox discovered by Daniel Bernoulli (see e.g., Blavatskyy 2005) and later axiomatised by von Neumann and Morgenstern (1947). Bernoulli noticed, that people do not make decisions 'as if' they simply multiply stakes and probabilities in order to calculate the relevant utilities of various options (e.g., that people are not risk neutral). Rather, they modify the stakes by taking a function of the stake (e.g., a function of constant relative risk aversion allows to capture behaviour of risk averse, risk seeking, and risk neutral individuals).

Both EV and EUT are deterministic theories which capture the behaviour of the co-called 'homo economicus' (a perfectly rational agent) who uses simple utility algorithms to formulate decision strategy (e.g., Thaler 2000). The notion of 'homo economicus', was substantially criticised in the economics and psychology literature as many experimental tests (conducted in the laboratory as well as the field) provided strong and robust evidence that human behaviour significantly departs from the predictions of EV and EUT (e.g., Charness et al. 2007). One of the most significant paradoxes was a paradox discovered by Maurice Allais, who showed that one of the fundamental axioms of EUT (Independence Axiom) is violated by individuals (Allais 1953). Further departures from EV and EUT were found (among others) by Amos Tversky and Daniel Kahneman (Kahneman & Tversky 1979). As a result, they proposed Prospect Theory (Kahneman & Tversky 1979) and later Cumulative Prospect Theory (CPT) (Tversky & Kahneman 1992) which allowed to explain many violations of EV and EUT (including violations of the Independence Axiom). Since CPT was proposed, more general formulations of this theory were offered in behavioural science literature giving rise to a whole class of so-called Rank-Dependent Utility (RDU) Models of which CPT is a special case (Wakker 2010). Apart from RDU models, a wide variety of deterministic decision theories that are trying to capture non-'homo economicus' behaviour exist. The numbers of these models run well into several dozen (see, e.g., Starmer, 2000 for a detailed review of many of these theories).

Stochastic decision theories

Stochastic decision theories provide an alternative to deterministic theories. These theories predict that a decision maker when making choices between alternatives A and B on some occasions will go for A and other occasions will choose B. In other words, unlike deterministic theories which predict that an individual will select either A or B with probability $p = 1$, stochastic theories predict that A and B can be chosen with some positive probability $0 < p < 1$. There are two main types of stochastic decision theories: models of noise and models of imprecision. Models of noise (e.g., Tremble model or Fechner models) maintain that people have stable preferences but with some positive probability these preferences are distorted by noise (e.g., Blavatskyy & Pogrebna 2010). These stable preferences could be defined by any deterministic theory (e.g., EV, EUT, CPT, etc.) but then a noise parameter (usually normally distributed with a zero mean) is added to these preferences. This means, e.g., that when choosing between A and B an individual really prefers A to B but with a positive probability this individual will opt for B over A (due to tremble, fatigue, or other type of error). Models of imprecision take a more extreme position. They say that people alternate between A and B simply because they have imprecise preferences. In other words, individuals do not know whether they prefer A over B or B over A and make decisions between the two based on various factors of decision environment (such as context). One of the most widely used theories of this type is Random Utility model (Loomes et al. 2002) which says that each individual has a basket of utilities and, dependent on the situation, draws one utility from this basket which results in an observation that sometimes A is chosen over B and sometimes B over A.

3.2.4 Cumulative Prospect Theory

This dissertation uses one of the most widely accepted deterministic theories – Cumulative Prospect Theory – in order to formulate hypotheses about the performance of planners. The use of Cumulative Prospect Theory is justified as follows.

CPT, as a deterministic theory, has a very clear structure and provides one stable prediction of choice. Since the focus of this study is on DP decisions, deviations from the demand plan can be easily explained by a stochastic choice model. The main drawback of the stochastic approach is that while it can *explain* behaviour, it does not often provide a clear *prediction* of what people are expected to do in a similar situation. Since one of the goals of this dissertation is to draw conclusions for planners in practice, predicting (suboptimal or even surprising) behaviour is more important than finding a perfectly fitting function that may explain the data on planning decisions. Therefore, CPT is well suited to formulate predictions about planning behaviour.

CPT also combines two important features: (1) it attempts to capture real human behaviour, i.e., tries to be sufficiently meaningful to shed light on how humans actually behave in practice and (2) it is quite easy to understand and has several important features that can help explain planning decisions. These features are discussed below. The theoretical approach for this thesis is to use the dominant decision theory (CPT) in combination with a well-known and generally accepted bias (mental accounting) to formulate predictions about human behaviour. Behavioural bias of mental accounting is necessary in order to capture individual attitudes towards time horizons (or time delays) which cannot be conceptualised by CPT alone. This is because CPT is a time-free deterministic theory, i.e., in CPT, all decisions are assumed to be made at a specific (discrete) point in time and preferences towards time are not operationalised. Further subsections show how CPT can be combined with mental accounting bias in order to formulate meaningful predictions for planning decisions.

The Assumption of Loss Aversion in Cumulative Prospect Theory

One of the main assumptions of CPT is that people suffer from loss aversion, i.e., they feel losses more severely than derive satisfaction from equal-sized gains. Specifically, Kahneman and Tversky (1979) conducted a number of hypothetical experiments and discovered that people were much more upset by losses than they were uplifted by gains. Empirically, they calculated that the discrepancy in perceptions between losses and equal-sized gains was $\lambda =$

2.25. In other words, if an individual lost £100, this individual felt 'as if' the actual loss is more than £200 (specifically, -£100*2.25=-£225). In CPT, this phenomenon was labelled *loss aversion* and with λ being is a loss aversion coefficient.

Loss aversion is an inherent part of CPT (its main assumption) which is woven deeply into the CPT's conceptual modelling framework. Specifically, because in CPT people derive satisfaction from *changes in wealth* rather than from *absolute wealth levels*, the utility function in CPT is called *'value function'*. A typical CPT value function is shown on Figure 15.

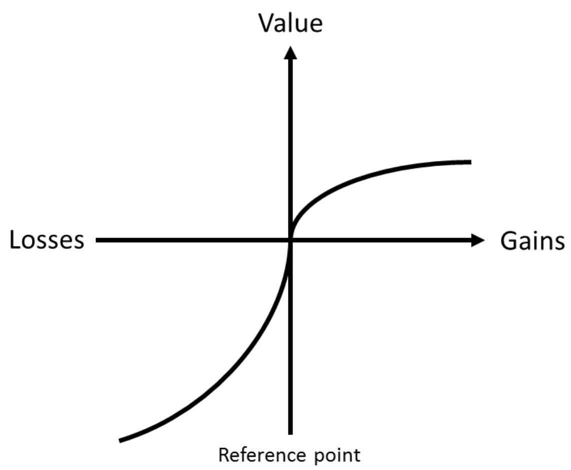


Figure 15 Value function of prospect theory (Kahneman & Tversky 1979)

According to CPT, individuals assess their changes in wealth relative to the so-called *reference point* (in the original paper Kahneman and Tversky (1979) assume that this point is equal to an individual's current wealth position). This is represented by the intercept of axes (coordinates [0,0]) on Figure 15, because a current wealth position can be different for different people but for all of them it represents a state of no gain and no loss. The value function is steeper for losses than for gains due to the assumption of loss aversion. The assumption of loss aversion is important for the argument in this dissertation because it can be combined with *mental accounting* to form *myopic loss aversion* (MLA) which is described in the following sections and which is used to capture planners' behaviour.

Mental Accounting

Mental accounting is a psychological bias coined by Richard Thaler which simply says that people may have different mental accounts for the same kind of resource (Thaler 1999). Thaler (1999, p. 183) defines mental accounting as *'the set of cognitive operations used by individuals and households'* to conduct financial operations. He also maintains that mental accounting has two manifestations: *'[the] first captures how outcomes are perceived and experienced, and how decisions are made and subsequently evaluated [and the second] provides the inputs to be both ex ante and ex post cost-benefit analyses…'* (Thaler, 1999, p. 183). While mental accounting is mostly used in finance literature to explain why people put assets into separate accounts (second manifestation), the first manifestation, specifically, that people perceive different time horizons differently when financial outcomes are at stake is widely used in behavioural science literature to explain important sub-optimal behaviours on financial markets (such as Equity Premium puzzle which is explained further). Mental accounting together with previously described loss aversion assumption of the CPT form MLA.

3.3 Myopic Loss Aversion

Myopic Loss Aversion (MLA) is a combination of the CPT assumption of loss aversion and mental accounting. It is used in behavioural finance in order to explain the Equity Premium Puzzle. This is a financial paradox that shows that while people should always invest in stocks rather than in bonds (due to a much higher profitability of stocks relative to bonds), large amounts of investment are still kept in bonds (see Figure 16). Figure 16 shows that had one invested $1 in small cap stocks in 1926, one would gain $18,106 by 2016. At the same time, by investing $1 into long-term bonds in 1926, one would only gain $110 by 2016. So, why would anyone want to invest into bonds?

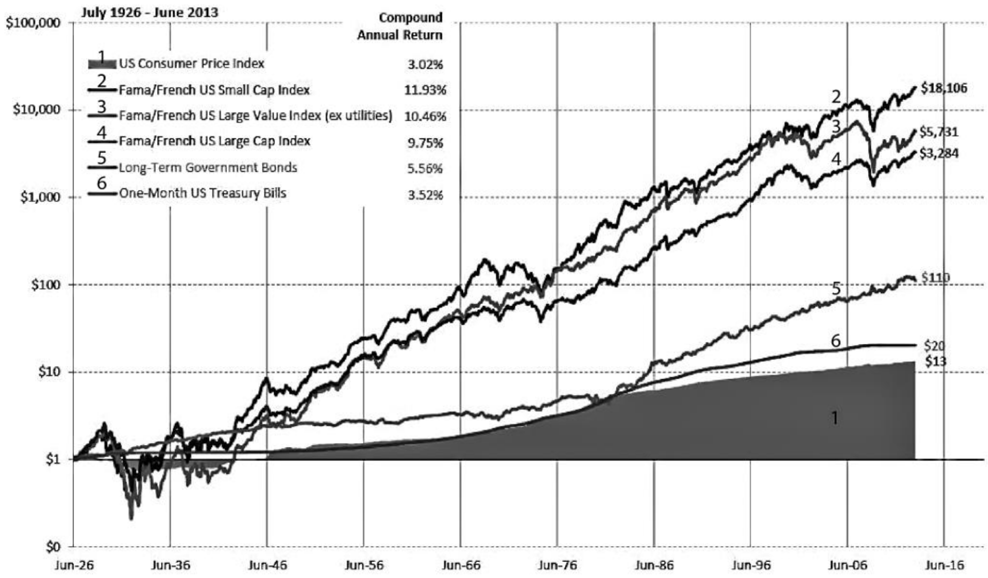


Figure 16 An Illustration of Equity Premium Puzzle (source: Ibbotson Consulting)

To provide an explanation for this Puzzle, MLA was introduced by Benartzi and Thaler (1995). Using both loss aversion (the main feature of CPT) and mental accounting (a well-known behavioural bias) they argued that people invest in bonds because they suffer from MLA. Loss aversion makes a sequence of investments under risk look less attractive in a myopic evaluation. Specifically, when investing in a stock, people are likely to evaluate their financial position frequently (at least once a year). Therefore, they are likely to spot losses more frequently and get more frequently upset and, as a result, pull their investment out due to loss aversion. At the same time, when investing into bonds, people are likely to evaluate their financial position infrequently (once every 5 years or so) which leads to the outcome that they simply do not spot losses frequently and are, therefore, less upset is they discover losses and less likely to pull their investments out.

The concept of MLA for financial decisions was first tested and confirmed experimentally by Uri Gneezy and Jan Potters (Gneezy & Potters 1997) where the degree of myopia systematically influenced the willingness to invest in higher risk alternatives. It was observed that less feedback and greater commitment reduces evaluation myopia making the decision-maker more willing to accept the risk (Gneezy & Potters 1997; Thaler et al. 1997; Haigh & List 2005). Specifically, Gneezy and Potters (1997) designed two experimental treatments: Short evaluation period treatment (Short) and Long evaluation period treatment (Long). Then they asked experimental participants to play 9 periods of the following experimental investment game. At the beginning of each period, each participant received an endowment E and was asked to invest any proportion x of this endowment into a risky asset which yielded $2.5x$ with probability $\frac{1}{3}$ and nothing with probability $\frac{2}{3}$. In treatment Short, participants were allowed to make a new investment decision in each period (irrespective of the outcome, participants would get E anew in each period) and in treatment Long – participants chose investment amount x only once every 3 periods (i.e., in periods 1, 4, and 7) and had to stick to their decision for the subsequent two periods. MLA predicted that in Long people will invest, on average, significantly more than in Short because loss aversion would prevent people from investing more money when they evaluate outcomes frequently (in the Short treatment). This (predicted) behavioural pattern was indeed observed in the experiment.

The original study by Gneezy and Potters (Gneezy & Potters 1997) was further replicated by Haigh and List (Haigh & List 2005) with professional traders (investment experts) where the effect of MLA was even stronger suggesting that professional experience does not reduce the bias. Markets also seem to have little or no influence in reducing the bias as shown by Gneezy et al. (2003) in an experimental market setting less frequent feedback and longer decision binding made prices for risky assets significantly higher. Focusing on markets designed to mitigate MLA, Mayhew and Vitalis (2014) found that MLA persists with inexperienced participants but not with experienced participants. However, experienced participants were unable to transfer this behaviour beyond the specially designed market and exhibited MLA again (Mayhew & Vitalis 2014). Considering individual versus team behaviour, despite an attenuated effect, MLA does persist with team decision making (Sutter 2007).

A refined hypothesis about how myopia affects risk taking was suggested by Langer and Webber (Langer & Weber 2005) which was confirmed in an experiment similar to the original by Gneezy and Potters (1997). Both feedback frequency and commitment have influence on myopic evaluation of assets, but the period of commitment influences myopia more than feedback frequency as demonstrated by Langer and Weber (Langer & Weber 2008).

Lower feedback frequency with longer commitment period decreases myopia and increases willingness to invest in a risky asset (Langer & Weber 2008). Lower frequency feedback delivers information at a more aggregate level and a longer period of commitment leads decision-makers to consider a longer period into the future. Longer commitment period along with more frequent feedback might help to make obvious that occasional losses are outweighed by overall gains (Langer & Weber 2008).

MLA is used as a part of the conceptual framework to test whether and to what extent modifying policy about evaluation periods can change the outcome of planning decisions. The main contribution to the MLA literature is that MLA was never applied to the problems of planning and MLA in application to planning decisions was never tested experimentally.

3.3.1 MLA analytical framework and hypotheses

In what follows, a new analytical framework that uses MLA to explain and predict DP decisions is proposed. Several hypotheses are formulated which are then tested experimentally. MLA as a possible explanation of failures in DP performance aligns with the previous research around DP performance. This is due to the following reasons:

(1) MLA is based on the CPT that offers a good basis for analysing DP failures because of its assumption of loss aversion.
(2) MLA represents a useful combination of a deterministic theory (CPT) and a cognitive bias (mental accounting) which allows not only to incorporate time delays into the analysis of decision-making but also to formulate meaningful predictions of planners' behaviour.
(3) Despite the fact that MLA combines CPT and mental accounting, it is a very simple concept which allows to easily introduce *system versus human* tests and analyse how business planning policy can affect and, possibly, offset the negative influence of human nature (loss aversion, mental accounting) or adverse effect of system nervousness (lack of policy) on planning decisions.

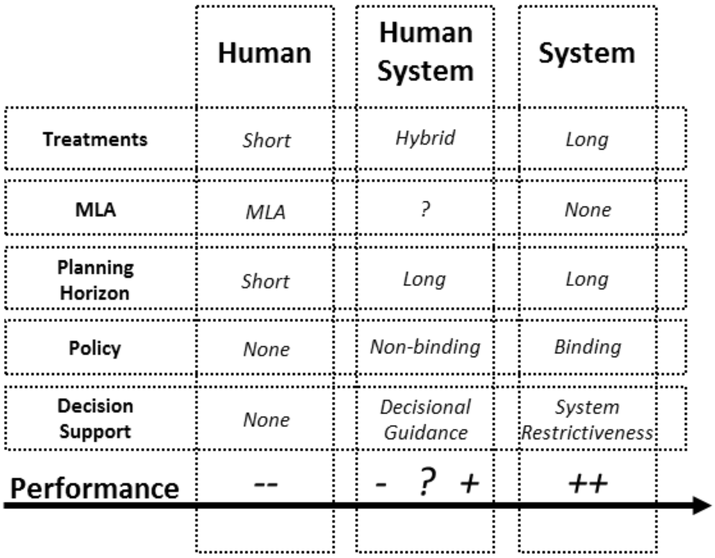


Figure 17 Analytical framework for demand planning decisions

The framework is illustrated in Figure 17. This framework draws on previous research described in the earlier sections of this thesis. Three general variants of determinants of the planning performance are considered (Figure 17) as variants where:

- Human nature is going to prevail over System thinking (Human);
- System thinking is going to offset Human nature (System);
- Both Human and System influences are possible (Human/System).

Following the previous MLA research (e.g., Haigh & List 2005), the framework will also concentrate on two types of planning horizons: Short and Long (see Planning Horizon strand on Figure 17). In the Short planning horizon, planners will be asked to make decisions in every time period and in the long planning horizon, planners will be asked to make decisions once in every few periods. However, in addition to varying planning horizons the framework introduces several important innovations to capture important planning context. Specifically, the focus is on two aspects of the planning problems: Policy and Decision Support (see relevant strands on Figure 17). The Short time horizon planning problem will not be restricted by any policy and will provide no decision support. This situation captures a hypothetical case when planners can change planning decisions in every time period and are not provided with any decision guidance or systemic restrictions (see first column on Figure 17). For simplicity, this situation is labelled as planning option *Short*.

In the Long time horizon, the framework allows to capture two possibilities. One of these possibilities is that planners are asked to make decisions every few periods and the business policy of making decisions every so often is binding. The system, therefore, restricts planners to stick to the initial schedule of relatively 'long-term' decision and prevents them from making short-term adjustments to the long-term plan. This situation captures a hypothetical case when a company introduces strict policy that does not allow planners to change existing long-term plan and bind them to stick to their long-term decisions (see third column on Figure 17). For simplicity, this situation is labelled as planning option *Long*.

Another possibility is more realistic and constitutes the main innovative component in the analytical framework. This possibility asks planners to make a long-term plan but has a non-binding policy that allows planners to alter the plan in each time period. In other words, planners have decision guidance not to alter the plan but they can do so if they wish. This situation captures a scenario that is close to the real world planning decision-making process within organizations when planners make a long-term plan but are then allowed to adjust this plan in response to various events (e.g., changes in demand, etc.). This situation is reflected in the second column on Figure 17. For simplicity, this situation is labelled as planning option *Hybrid*.

Considering all three planning options (Short, Long, and Hybrid), the framework predicts the following. In Short, where planners are unrestricted by policy and have no decision guidance, the planning performance will be primarily influences by the human nature (specifically, by MLA). Therefore, planners will evaluate their plans frequently and, due to a combination of loss aversion and mental accounting will make too many adjustments to their plans at a higher magnitude which will result in decline in profits. In other words, the expectation is that planners' performance in Short will be relatively poor (captured by the label '- -' on Figure 17). In Long, the System thinking will prevail and offset the negative effects of MLA due to policy and systemic restrictions. Therefore, in this treatment, planners will not 'overadjust' and reach a relatively good performance (captured by the label '+ +' on Figure 17). Finally, in Hybrid the framework predicts that the performance will depend on the individual propensity of planners' to follow a long-term plan. This performance will be relatively good if they stick to the decision guidance and will be relatively poor if they fail to follow the decision guidance in which case the MLA may prevail and negatively influence the performance.

The proposed framework is not only consistent with the previous decision-theoretic research, but also with literature on O&SCM. For example, Zhao and Lee (1993) argue that '…*less frequent re-planning improves system performance*…' (Zhao & Lee, 1993, p.185) which suggests that binding the plan to the previous choice should have a positive effect on performance. Similarly, some evidence from O&SCM implied that planning with longer time horizons should exhibit better performance than planning with short time horizons (Simpson, 1999). It has also been shown that DP can be manipulated through policy, i.e., that planning performance may be affected by policy parameters (de Kok & Inderfurth, 1997). Finally, policy can be implemented via decision support that can either offer guidance or restrict decisions (Silver 1991). Given the proposed framework, following hypotheses are formulated:

Hypothesis 1: *Planning performance in the planning option Long will be better than that in the planning option Short (i.e., planners will achieve higher profit in Long compared to Short).*

Hypothesis 2: *Planning performance in the planning option Hybrid will depend on the extent to which individual planners are able to follow the long-term plan: those planners who stick to the long-term plan (i.e., those who mimic behaviour in Long) will achieve higher profit than those who try to alter the plan frequently (i.e., those who mimic behaviour in Short).*

3.4 Human and system: individual differences

Since the framework predicts that in the planning option Hybrid, the profitability will essentially depend on the planners' individual planning behaviour. Insights from psychological research are brought in to formulate additional hypotheses about how (i) individual exposure to planning as well as (ii) other individual characteristics may help or hinder planning performance under conditions of this planning option. In what follows, relevant literature is discussed and hypotheses developed about the potential impact of factors (i) an (ii) on the planning performance for the Hybrid option.

3.4.1 Overview of individual differences

To understand the impact of individual traits and individual differences on planning, it is necessary to step back and first look at psychological research on intelligence. Intelligence is one of the most studied topics in psychology. One of the most replicated and consistent empirical finding was Spearman's (1904) result that individuals performing well on one mental task tend to perform well on most others regardless of variations in the task (Deary 2000). This set the direction of social psychology focused on the processes that can predict performance in specific situations. Alternatively, individual differences are used in studies that try to uncover regularities between individuals that can be generalised across different situations. The most studied behavioural differences are personality traits and cognitive abilities. However, to date, behavioural analysis has in great part ignored individual differences despite the claim that the focus is on individual behavioural instead of group averages (Williams et al. 2008). As a result, the general trend in psychology is that many studies are carried out at a group level and individual differences (individual heterogeneity) are treated as errors rather than a relevant phenomenon (e.g., Williams et al. 2008; Maltby et al. 2013).

Nonetheless, individual differences is one of the largest most interesting phenomena in social psychology focusing on how individuals differ from each other in terms of behaviour (e.g., Maltby et al. 2013). Each individual is considered unique. Consequently, differences can be observed at different levels. Individual differences can include many different measurable factors and dimensions offering a virtually infinite number of combinations. For example, individuals can differ in terms of memory, intelligence (e.g., I.Q.), knowledge, personality (e.g., Big Five), sex, height, age, education and experience, as well as many other characteristics used to explain (part of) the observed variance between individuals.

Following the idea that some cognitive types are *a priori* better decision-makers than others (Davis 1982), personality has been repeatedly demonstrated to significantly affect decision making (e.g., Dahlbäck 1990; Lauriola & Levin 2001; Soane & Chmiel 2005; Davis et al. 2007; Icellioglu & Ozden 2012; Filiz & Battaglio 2015; Mihaela 2015). Personality traits have been found to be significant to many different dimensions related to decision making, linked to job performance (Hurtz & Donovan 2000; Furnham 2008), academic achievement (Barakat & Othman 2005; Poropat 2009), risk attitude (Dahlbäck 1990; Lauriola & Levin 2001) and, problem solving (Weinman et al. 1985). Associated to high academic achievement and performance it has been found to be associated to Intellect/Openness (Heaven & Ciarrochi 2012) and conscientiousness (Conrad & Patry 2012).

One of cornerstones of individual differences is the study of personality and individual uniqueness. Personality is arguably one of the most complex and controversial areas of psychology research. The major complexity in the study of personality comes from the idea that each individual is unique and there are as many distinct personalities as conscious human beings ever lived on earth. However, the need for a simplified structure motivated researchers towards defining personality as a combination of traits. Larsen and Buss (2010, p.4) defined personality as '*…the set of psychological traits and mechanisms within the individual that are organized and relatively enduring and that influence his or her interactions with, and adaptations to, the intrapsychic, physical, and social environments.'* From this definition comes that first, personality is relatively stable and second that it will influence one's interaction with the environment, hence it will affect individual decision-making.

The conceptual foundations of trait theory date back to the works of Peterson (1968) and Mischel (1968). The basic structure of personality has converged to the Big Five through many trait theorists via independent studies (Digman 1990; Goldberg 1990; Pervin 1994; Digman 1997; John &

Srivastava 1999; John et al. 2008; Fleeson & Jayawickreme 2015). Formally, the five-factor model was suggested in the nineties by Digman (1990) based on a review of previous studies on personality inventories and its communalities. Over the last decades, the Big Five has been repeatedly validated and is considered a reliable instrument to represent the personality of an individual. The first order constructs of the Big Five are extroversion, conscientiousness, agreeableness, neuroticism and openness (Digman 1990; Goldberg 1990; Goldberg 1992; John & Srivastava 1999).

Many businesses believe that certain personality traits can boost performance. For example, a large number of companies around the globe use Myers-Briggs Type Indicator® (MBTI) personality inventory[7] to determine individual traits for individual development. Similarly, it is suggested that there is a 'right' mind-set (personality) for different planning (e.g., Lapide 2007). Yet, the tests of whether and how personality traits influence performance in DP have not been conducted before and it is difficult to derive such information from field (non-experimental) data as (i) managers differ in their abilities to perform their duties; and (ii) most importantly, there is too much complexity in the decision environment to distinguish between results which depend on personality and results that depend on external factors. Given both the evidence from the psychology literature as well as the consensus among practitioners that personality trait tests bring value to business, the expectation is that individual personality traits will have significant impact on performance. Since personality traits are likely to matter most in the Hybrid DP option, the following hypothesis is formulated.

Hypothesis 3: *Individual differences and individual personality traits are a significant predictor of demand planning performance. The effect of individual differences and individual personality traits should be particularly strong in the Hybrid planning option where individuals have a choice between following and not following the decision guidance to stick to the long-term plan.*

---

[7] See http://www.myersbriggs.org/

Individual differences offer a broad range of constructs. The following section discusses what individual differences have been associated to decision-making performance, with emphasis on DP. The review will concentrate on personality inventories and psychometric scales that measure individual differences relevant for the purpose of this thesis.

3.4.2 Individual exposure to demand planning processes: experience versus theoretical knowledge

The first and one of the most important individual characteristic considered is individual exposure to DP. The literature on leadership and business performance distinguishes between practical experience (henceforth, experience) and theoretical knowledge (henceforth, knowledge).[8] Experience and individual knowledge are known to significantly affect decision making (Ackerman 1996; Bolton & Katok 2004). However, there is some disagreement regarding the effect of experience on performance. Theoretically, it is expected that training and prior experience benefits performance through better decisions (Ackerman 1996). The evidence around this is mixed with some studies observing experience as beneficial (e.g., Goodall & Pogrebna 2015) while others, specifically comparing naïve and expert groups in experiments find the opposite (e.g, Haigh & List 2005; Brown & Tang 2006; Bolton et al. 2012). Considering that intuition is recognition which comes from experience (Simon 1969) other authors challenge the claim that experts poses high level of intuitive skills (Sjöberg 2003). However, intuitive forecasting by experienced managers seems to benefit accuracy when combined with statistical forecasting (Blattberg et al. 1990).

Both practice and decision-making experiments challenge the claim that experience and knowledge benefit performance. For example, from a study on 60,000 forecasts and their outcomes, managers decided to adjust the forecast in most forecasts which often reduced accuracy (Fildes et al. 2009).

---

[8] See, e.g., Goodall, A (2009) "Socrates in the Boardroom: Why Research Universities Should Be Led by Top Scholars", Princeton University Press, for an extensive discussion on this literature.

Moreover, a study by Rieg (2010) of forecasting in automotive industry over 15 years finds no improvement over time. In literature around expert judgement, this has been called the *'process-performance paradox in expert judgement'* (Camerer & Johnson 1991, p.8).

Experimental results suggest a similar relationship between experience and performance. Sub-optimal decision making can resist to professional experience or background training (Bolton & Katok 2008; Lurie & Swaminathan 2009; Bolton et al. 2012). This same idea links back to MLA hypothesis in professional traders where Haigh and List (2005) detected that professionals suffered from greater bias than naïve students. Moreover, in an inventory management experiment, both professional buyers and students exhibited the same behaviour (Brown & Tang, 2006). This aligns with findings by Bolton et al., (2012) who compared experienced procurement managers and naïve students to solve the newsvendor problem, finding that both groups exhibit the same kind of pull-to-centre bias. Bolton et al., (2012) detected that professional managers use information and task training no better than students do. Additional knowledge about the demand did not improve performance either in experimental settings (Schweitzer et al. 2000). Regarding the available information (knowledge), for example, knowing the demand distribution affects the behaviour but it does not lead to a better performance (Benzion et al. 2009).

Considering this mixed evidence from various studies, the hypothesis about the relative impact of planning exposure will be based on the following approach. It enables the distinction between theoretical or knowledge exposure to planning (labelled as Theory), practical exposure or experience in planning (labelled Practice) and no exposure to planning (labelled as Naïve). This approach is summarised in figure 16 in the Exposure-Performance Matrix (EPM).

The rationale behind the EPM is the following. Naïve planners (those who do not have any previous exposure to planning) will exhibit the worst performance as they have neither theoretical knowledge nor practical experience of the planning issues that may arise in the planning process.

Theory planners (those who have theoretical knowledge rather than practical experience) will exhibit the best performance as they are familiar with potential planning problems (such as, e.g., severe overadjustment of long-term plans) and, therefore, they are more likely to carry out the planning process 'by the book' making minimal number of overadjustment errors. EPM predicts that Practice planners (those who have practical experience in planning) will exhibit performance somewhere in-between Naïve and Theory planners. This is because practical knowledge of planning decisions within organisations will make those planners more likely to make additional adjustments to the long-term plan (this feature of the EPM comes from the empirical observations of excess adjustment observed in the real life decisions). In other words, Practice planner may perform as well as Theory planners but their practical experience will (generally) hinder their performance making them chase demand more than Theory planners would do. It is important to note that Naïve planners necessarily have neither theoretical nor practical exposure to planning. Theory planners necessarily do not have any practical experience while Practice planners may have some theoretical knowledge of the planning problems but they should necessarily have practical experience in dealing with planning problems in real-life situations.

Finally, policy is a way of overriding lack of exposure and improving performance. Policy can be enforced via restrictiveness or guidance, e.g., restricting managers from making bad decisions or guiding them by providing the necessary information to make a good decision.

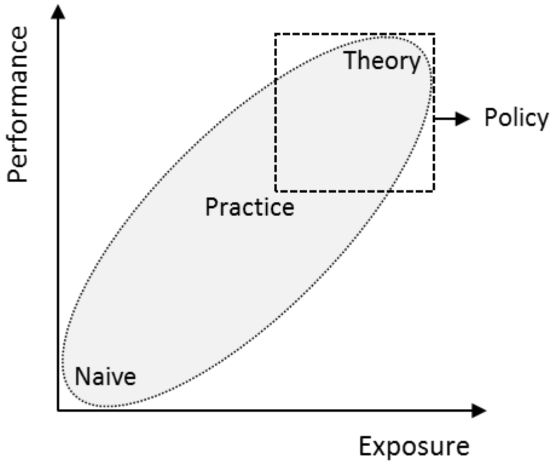


Figure 18 Exposure-Performance Matrix

Considering EPM, the following hypothesis is formulated.

Hypothesis 3-A: *Individual differences with regard to exposure to planning is an important determinant of planning performance with Theory planners performing better than Practice and Naïve planners*

3.4.3 Naïve interventionism hypothesis

DP performance is affected by naïve interventionism in the form of frequent adjustments to the plan. Worth noting that this considers conditions when no additional information is available to exclude the beneficial adjustments of additional knowledge about demand shaping decisions. A subject completely unaware, with no exposure to theory or experience, is considered perfectly naïve. Naïve interventionism is an individual characteristic that is related to individual's background, experience and practice. Hence, such attribute belonging to the domain of individual differences. Naïve interventionism encompasses plan instability or plan changes.

Hypothesis 3-B: *Greater level of naïve interventionism leads to worse demand planning performance*

The following section focuses specifically on individual differences and DP performance.

3.4.4 Personality inventory: Big Five

The field of personality traits is one of the largest in individual difference research. Based on the trait theory part of the researchers agree on the fundamental elements describing a personality (e.g., Digman 1990; Goldberg 1990; Pervin 1994; Digman 1997; John & Srivastava 1999; John et al. 2008; Fleeson & Jayawickreme 2015).

Each factor of the Big Five describes a distinct aspect of personality. Openness refers to curiosity, originality, and ingenuity. Alternatively, Openness can be referred to in different ways. It can be Culture because it includes intellectualism, polish, and independence of mind. In addition, Openness can be sometimes referred to as Intellect as it gives emphasis on intelligence, reflection, and sophistication. Conscientiousness includes dependability, responsibility, and orderliness. Sometimes referred to just as Dependability. Extraversion relates to energy, talkativeness, and assertiveness. Extraversion is sometimes called surgency. Agreeableness includes trust, good-naturedness, and cooperativeness – therefore some studies refer to it as a combination of compliance and friendliness. Finally, Neuroticism refers to how easy it is to upset the individual and stands opposite to emotional stability (which can be obtained by simply reverting the scores).

Strohhecker and Größler (2013) identify intelligence as the strongest predictor of performance but 'openness to new experiences' was detrimental to performance. Similarly, neuroticism has been attributed to impulsive behaviour leading to over-reactions, mistrust and second-guessing which is known to have a negative impact on the SC (Christopher & Lee 2001).

3.4.5 Specific personality construct measures

To measure specific personality constructs there is a wide range of alternatives. It was particularly interesting to include a measure of impulsiveness, since it has been associated to underperformance in previous OM literature (e.g., Martin & Potts 2009; Ockenfels & Selten 2015). Barratt Impulsiveness Scale (BIS), being one of the latest revisions referred to as BIS-11 proposed by Patton et al. (1995) was selected as it is one of the most used impulsiveness scales. The full list of questions for the BIS is included in the 6.7Appendix C.

DP requires making considerations about the future events. Previous OM literature suggests that behaviours such as risk or loss aversion (e.g., De Véricourt et al. 2013; Ma et al. 2015) affect performance significantly and some of the anecdotal explanations mentions that it has to do with how individuals evaluate the consequences of their actions. Elaboration on Potential Outcomes (EPO) is a scale proposed by Nenkov et al. (2008) designed to evaluate potential positive and negative consequences of individual behaviours. The complete list of questions for the EPO is available in 6.7Appendix B.

Finally, expert intuition has been referred to as one of the main traits involved in solving complex judgement problems. While expert intuition is hard to measure and no specific scale for DP has been proposed yet, overall intuition as part of decision-making style is included in the General Decision Making Style (GDMS) scale. Self-reported decision making style has been found to predict behavioural decision making (Franken & Muris 2005). The GDMS was designed to assess how individuals approach decision situations. One of the validations of the GDMS was done by Gambetti et al. (2008). The GDMS distinguishes between five decision styles (Scott & Bruce 1995). A rational style focuses on a careful search for and logical evaluation of alternatives. The avoidant style concerns avoiding or postponing decisions. A dependent style refers to searching for directions or advice from others. An intuitive style relies on hunches and feelings. Finally, a spontaneous style regards to immediacy and wish to finish the decision making process quickly. The full list of questions for the GDMS can be consulted in 6.7Appendix D.

Hypothesis 3-C: *Greater level of impulsiveness leads to worse demand planning performance*

3.4.6 Other individual differences

Individuals can also be distinguished by demographic characteristics such as sex and age. Most of the previous experimental research related to DP collected these demographics most of the time for descriptive purposes (e.g., Bolton et al. 2012; Moritz et al. 2014). De Véricourt et al. (2013) finds sex to be a significant predictor of DP performance. Therefore, the demographic sub-hypothesis 3-D is formulated as follows:

Hypothesis 3-D: *Male subjects outperform female subjects in the planning task*

3.4.7 Individual differences hypothesis overview

The conceptual framework around individual differences is illustrated in Figure 19. The human and system situation offers choice and this is where individual differences are expected to be predictors of DP performance. Decision support is offered as decisional guidance. Exposure, considered as experience or knowledge varies between naïve and theoretical with practice in the middle. Naïve interventionism (as plan instability) is expected to negatively affect performance. Psychometrics are measured using the BIG5, GDMS, EPO and BIS. Demographics measured are sex and age. It is expected that individual differences will explain at least part of the differences in performance.

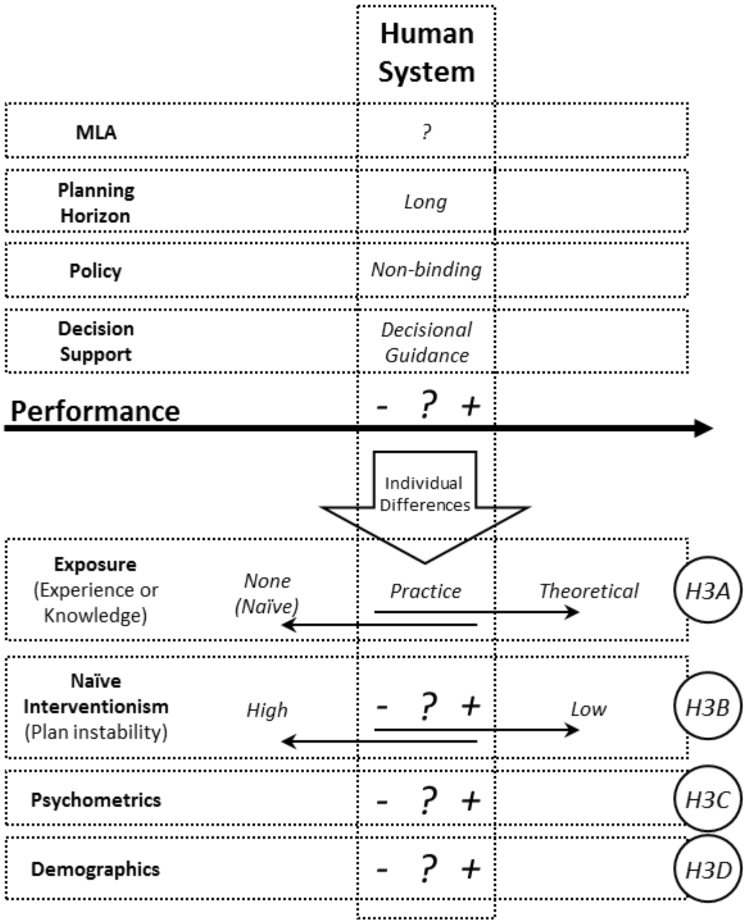


Figure 19 Conceptual framework focused on individual differences

3.5 Complete conceptual framework, testable hypotheses and predictions

The following conceptual framework is proposed for this thesis (see Figure 20). This framework is inspired by the previous theoretical and empirical results discussed above. The outcome of DP decisions (measured by planning performance) is influenced by systemic factors (organisational policies and processes) and human factors (individual exposure to planning decisions, individual propensity to follow a plan, and other individual traits and characteristics such a personality).

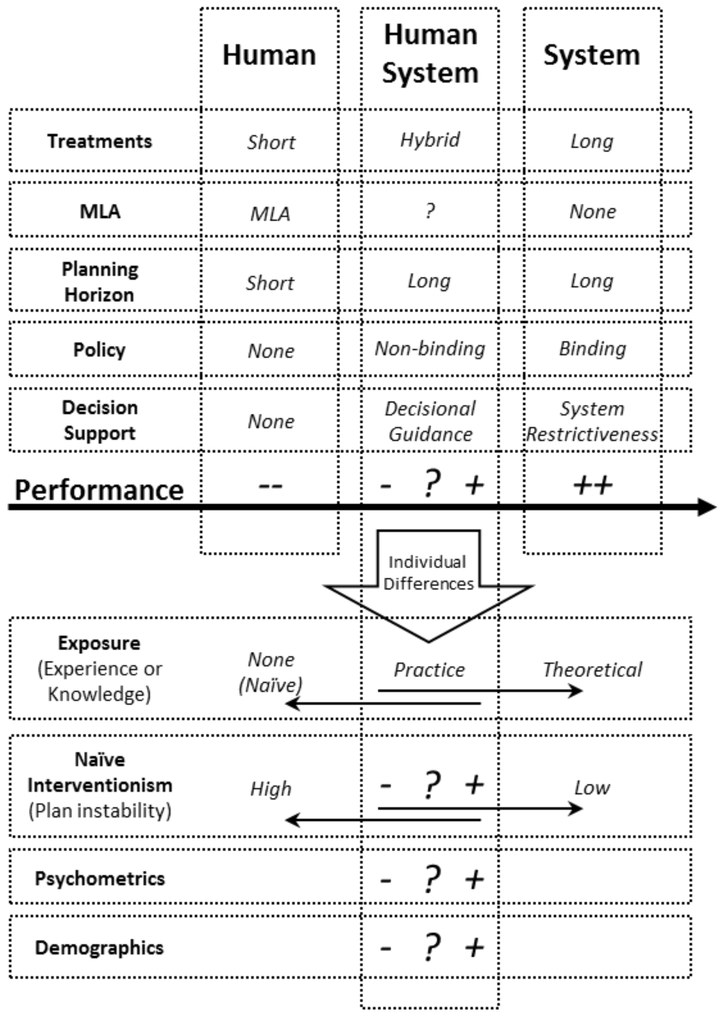


Figure 20 Complete Conceptual Framework

While there could be a large number of policies and processes, the focus is only on systemic factors affecting time horizons. Specifically, on policy commitments with regard to the planning time horizons.

In this regard, policy can ask planners to make planning decisions in each period (non-binding policy commitment). It can also restrict planners to only make decisions every several periods (Binding policy commitment). Finally, it can also ask planners to adopt an Optional (Hybrid) policy whereby planners are asked to make decisions every once in a while (once in several periods) but this policy in non-binding – i.e., while planners make a plan for several periods, then can change their decisions in every period. The empirical tests use non-binding (short), binding (long), and optional (hybrid) treatments to test whether policy affects planning outcomes. This policy is enforced via decision support as none, system restrictiveness and, decisional guidance respectively. It is expected that, due to MLA, varying policy will lead to the following planning outcomes. In the Long treatment, mostly systemic factors will be at play and therefore, planners will stick to the original plan, be less hectic and reach better profit levels. In the Short treatment, human factors will dominate systemic factors as loss averse planners due to mental accounting will notice losses more frequently than those in other treatments and will try to change the original plan making the outcome of planning worse. Finally, both systemic and human factors can be at play in hybrid treatment where people with different individual characteristics will be converging either towards behaviour mimicking Long or behaviour mimicking Short treatments.

With regard to individual characteristics, most interesting outcomes are expected from the Hybrid treatment. The assumption is that performance in the Long treatment will not be affected by individual differences. Similarly, in Short treatment loss aversion and mental accounting combined will outweigh any effects of individual differences.

With regard to the Optional (Hybrid) treatment, it is expected that people with no exposure to planning will drift towards behaviour similar to that in the Short treatment. People with theoretical exposure to planning will mimic Long treatment behaviour because they would have knowledge of potential damage that may occur due to over adjusting the plan. Finally, people with practical exposure may drift towards Short or Long, however, given the fact that over adjustment in planning is a big problem among practicing professionals, the expectation is that they are more likely to drift towards Short than Long.

3.6 Chapter summary

DP is subject to human judgement. The dominant model of human behaviour, CPT, is adopted which is a time-free deterministic theory. In CPT, all decisions are assumed to be made at a specific point in time and preferences towards time are not described. CPT with mental accounting bias can be combined in order to formulate meaningful predictions for planning decisions. In order to understand the contribution of individual differences, the focus is on individual's exposure (experience) and naïve interventionism (the plan instability). To observe individual differences, the focus is on personality, psychometrics and demographics. Personality is assumed to as combination of the five constructs of the big five. Psychometrics and demographics include decision-making style, elaboration and evaluation of planning and impulsiveness. Demographics include exposure (experience/education), sex and age. It is expected that some of the constructs will highlight the differences. The following chapter contains the research design adopted to test the framework and respective hypothesis.

4 Research design

4.1 Introduction

The theoretical framework presented in Figure 20 (p.77) is used to build a decision making experiment. The structure of Chapter 4 is outlined in Table 7. Following this introduction, the philosophical positioning of this research is presented. After a review of decision-making experiment methodology, this chapter focuses on the experimental task. The experiment is based on the newsvendor problem due to its previous successful applications in similar situations. A section on measuring individual differences provides an overview of how personality traits, psychometrics, and demographics can be captured. Finally, the experimental design is presented with the description of the treatments, sample and other methodological choices. The chapter closes with a summary.

Table 7 Chapter 5 structure

Section	Overview
4.2 Ontological and epistemological perspective	The philosophical positioning is discussed. This thesis adopts positivism, which is aligned with the philosophy adopted by most previous research informing this thesis.
4.3 Decision making experiments: methodological considerations	The methodology of decision-making experiments is reviewed along with its justification, task, validity, the requirements for compensation, sample size and ethical considerations.
4.4 Experimental task: newsvendor problem	The newsvendor problem is chosen as basis for the task in the decision making experiment. The justification is presented followed by a literature review around the newsvendor problem and its previous applications.
4.5 Measuring individual differences	In order to capture individual differences, a personality inventory, several psychometric scales and demographic questions are selected.
4.6 Experimental design	The experimental design section provides an overview, chosen strategy regarding incentives, target sample size and how performance and individual differences are going to be measured. The three experimental treatments are explained. As part of the requirements of decision-making experiment methodology and ethical considerations information sheets and declarations of informed consent are developed. The experimental procedure is summarised followed by eligibility criteria, risks and benefits for participants. The section closes with a summary of the experimental flow.

4.2 Ontological and epistemological perspective

The philosophical position is positivism due to author's personal preference influenced by background training and education in engineering. The present research builds on contributions from the fields of Economics, Psychology, OR and OM which are predominantly positivistic (e.g., Kagel et al. 1995; Starmer et al. 2009). These fields can be considered relatively mature. The exception is the field of OM that due to its broad nature includes other philosophical views such as critical realism, action research or interpretivism. The advantage of a mature field is that it allows theory testing. Edmondson and McManus (2007) recommends keeping the philosophical consistency with the core literature informing the research.

The aim of this research is to develop new knowledge and the assumption is that reality is objective. This means that reality is constructed of measurable and testable phenomena and objects that exist regardless of whether there is someone directly experiencing or observing it (O'Gorman & MacIntosh 2012). Findings from previous research are assumed independent from the researchers, replicable and generalizable as long as their results are statistically significant and valid. The same is expected of the results of this research. Assuming an objective reality opens the possibility to use a wide array of techniques, data and enables replication and comparison of the results with previous studies (O'Gorman & MacIntosh 2012).

Epistemologically, valid and reliable knowledge is developed following a positivist tradition that matches the objective ontology. Positivism comes from natural sciences and is one of the most common epistemologies in science. From a positivistic view follows that findings can be directly observed, verified and replicated regardless of the observers. The desire of this research is to formulate claims that can be generalizable rules.

The dominant methodologies in the positivistic paradigm are experimental research and survey research. Decision making experiments are widely used in experimental economics and other fields studying decision making (see Kagel et al. 1995). The most commonly used methods to analyse experimental results are quantitative methods, e.g., econometric analysis.

Any other philosophical positioning rather than positivism would invalidate most of the assumptions made. The closest alternative philosophy is pragmatism that allows the use of most methods as long as they are suited to the research problem and potentially avoiding philosophical debates. It recognises that each method has its limitations so the main characteristic is to use mixed methods and triangulation. This alternative is compatible with the present problem and the use of the same theories. However, due to personal preference, the drawbacks of positivist tradition are acceptable. This does not mean that introspective and intuitive knowledge are ignored, it is, however, used anecdotally to inspire objective observation.

Positivism is often criticised as reductionist. However, the reduction of complexity can be advantageous. People can be very complex. However, the assumption is that individual differences can be objectively measured. Whilst an individual is an extremely complex system, the belief is that it is still an objective system. It is possible to measure some of higher order traits using psychometric scales and personality inventories developed in psychology.

The following sections describe methodological choices and data-analysis techniques considering an objective reality and a positivist epistemology.

4.3 Decision making experiments: methodological considerations

The aim of this research is to test the hypotheses that behavioural biases and individual differences affect planners decision making following the methodology of decision making experiments (see Kagel et al. 1995). The aim is to set a task and measure performance that will depend on the subject's judgement. The task will have a theoretical optimum and subject's performance is expected to deviate to some degree from the optimal solution under the assumption of a normative view of rationality. Studying systematic errors and biases in judgement allows an insight on cognitive limitations and underlying works of statistical and logical intuition (Kahneman & Tversky 1982). Special attention is paid to methodological rigour as application of behavioural experimentation in the context of O&SCM *'provides much less evidence of an understanding of what "rigor" with such methods entails'*

(Bachrach & Bendoly 2011, p.5). There are two main alternatives to run decision making experiments, one is a laboratory experiment and the other is a framed field experiment where both can either be in form of an incentivised decision making experiment or a non-incentivised survey. Both methodologies are well established in economics and social psychology. For a detailed methodological overview see Plott and Smith (2008) regarding the methodology of laboratory experiments and Harrison and List (Harrison & List 2004) for the methodology of framed field experiments

4.3.1 Why experiment

The choice to approach this study using an experiment is because it is the most used methodology in behavioural economics and behavioural OR, it allows hypothesis testing and manipulation of particular factors. Running experiments is now an established method to explain and/or describe economic and business activity bringing these fields into alignment with many of the natural sciences that rely on experimental methods (e.g., physics and biology). This is backed up by publications, citations and even a Nobel prize (Smith 2002). Over the last 12 years 11% of the most-cited papers in economics are experimental which is roughly the same number as theoretical papers.

Experiments are not as common in OM and OR literature, although the methodology has been recently gaining increasing attention (Bachrach & Bendoly 2011). Beyond the context of economics, the methodology of decision-making experiments has been extensively used to test policy, designs or best practices. The advantage of using experiments is that it can test theories under precisely controlled and/or measured conditions that are typically unavailable in field data.

4.3.2 Experimental task

In order to design a decision making experiment it is necessary to create a situation where the subject is required to judge the provided information and make a decision (Kagel et al. 1995). The use of computer simulations to study judgement and decision making is widely adopted (e.g., Funke 2001). There are several considerations about internal and external validity to take into consideration in order to retain the methodological rigour.

This research focuses on DP performance, particularly how managers react to a demand signal in a situation of uncertainty. DP decisions are usually made repeatedly with a pre-determined frequency for an individual product or family of products (in case of goods). Similar happens for services. The problem targeted in this research is the one of mistrust, over-reactions and second-guessing. This irrational behaviour can have an insignificant effect as an isolated episode but repeated over many periods can have a significant toll on the overall business performance. Hence, the experiment has to have repeated decisions over many periods.

Considering the demand signal, in order to observe a 'clean' behaviour to the demand signal, the simulated demand must be pattern free so any judgement made relies solely on the decision-maker's interpretation. A pattern free demand signal can be obtained using a uniform distribution commonly used in decision making experiments on planning, ordering and forecasting in behavioural operations literature (Schweitzer et al. 2000; Bolton & Katok 2004; Benzion et al. 2008; Bostian et al. 2008; Croson, Ren, et al. 2013). By presenting managers with a pattern free demand signal, it is expected that any judgement bias will stand out clear.

4.3.3 External validity

To ensure external validity, the experimental task and context must be close to practice. DP activity is usually made from manager's own workstations (desktops or laptops). This allows to run the experiment outside laboratory environment, i.e., in a field setting, using a platform that can run from subject's own workstation and consequently improving the ecological validity (Berkowitz & Donnerstein 1982). To achieve this the framed field experiment should be web-based. There are many other practical advantages of conducting the experiment online. First, it allows reaching participants more easily. Second, it takes advantage of the current survey platforms such as

Qualtrics[9], which allows high levels of customisation through Java Scrip programming. Third, it can be disseminated via targeted mailing, advertised in speciality groups (e.g., linked in groups) going beyond geographical limitations, as subject's physical presence in the laboratory is not necessary. For more a detailed overview of framed field experiment see Harrison and List (Harrison & List 2004).

4.3.4 Compensation

Compensation is common in decision making experiments (Gneezy & Rustichini 2000). In the previous literature on decision making experiments in OM, the behaviour of participants in previous studies was measured sometimes in incentivized experiments (e.g., Schweitzer & Cachon 2000; Bolton & Katok 2008; De Véricourt et al. 2013; Ovchinnikov et al. 2015) and sometimes in non-incentivized surveys (e.g., Brown & Tang 2006; Rudi & Drake 2010). However, much research in economics and social psychology (see Gneezy & Rustichini 2000; Camerer et al. 2004; Gneezy et al. 2011) shows that in a field setting (when approaching professionals) much cleaner (less noisy) results were obtained when professionals were asked to participate in non-incentivized (non-paid) surveys rather than in incentivised (paid) experiments. *'Offering money did not always produce an improvement and subjects performed poorly'* (Gneezy & Rustichini 2000, p.791). Moreover, Remus et al. (1998) conducted experiments specifically to determine whether or not financial incentives affect judgmental forecasting experiments, finding that incentives had no significant impact on forecasting accuracy. Hence, the experimental design offers flexibility in terms of compensation and there are both advantages (commonly accepted practice) and disadvantages (introduction of noise). It is worth noting that incentives do not necessary need to be monetary (although it's the most common way). Subjects, particularly professionals, are harder to compensate in a meaningful way with money as a professional's time is usually much more expensive than the one of a student. However, incentives can take a different form and in return for

---

[9] Qualtrics (www.qualtrics.com) is a comprehensive web-based software to manage customer experience via survey

participation subjects can be offered performance feedback in form of individual profiles as well as access to reports on findings from the study (e.g., Moritz et al. 2013). Although this form is arguably not proportional to performance in the experimental task, it can be assumed that a subject that puts effort into the task will receive results that are more meaningful.

4.3.5 Sample size requirements

Considerations about the sample size are of critical importance with implications on the choice of data analysis methods (Kagel et al. 1995). Most commonly, experimental results are analysed using econometric methodology. List et al. (2011, p.168) suggests three principles to decide on the minimum sample size. First, *'[with] a continuous outcome measure one should only allocate subjects equally across treatment and control if the sample variances of the outcome means are expected to be equal in the treatment and control groups. i.e., if the treatment effect is homogenous.'* For example, assuming of homogenous treatment effects, it is necessary n = 16 (64) observations in each treatment cell to detect a one (one-half) standard deviation change in the outcome variable (following the standards in the literature of a significance level of 0.05, and setting power to 0.80). To detect a one-tenth standard deviation change, 1,568 subjects are needed in each treatment cell. *Second, 'in those cases where the sample variances are not equal, the ratio of the sample sizes should be set equal to the ratio of the standard deviations.'* Finally, third, 'i*f the cost of sampling subjects varies across experimental cells, then the ratio of the sample sizes is inversely proportional to the square root of the relative costs.'* For this thesis, the assumption is to work with approximately two standard deviation points. Therefore, a minimum of 30 participants per treatment will be required in order to ensure statistical significance.

4.3.6 Ethical considerations

Given that decision making experiments involve people, it is essential to have appropriate ethical considerations (Kagel et al. 1995). This must include an information sheet about the study and a consent form. The assumption is that such study will only include adult participants that are not considered vulnerable in any way.

Participants must be provided with an information sheet about the study that includes a clear identification of the researchers involved, overview of the purpose, information on research outlining what the participants are required to do. Special attention must be given to risks, and participants must be notified that there is a potential loss of confidentiality due to data storage breach in an unlikely event as well as information about information retraction. A highlight of potential benefits of this study should be included. Anonymity considerations must be made clear as results are not shared with third parties and identifying information is on not kept with the results that are codified and anonymised distinguished only by an individual ID number for each participant. It must be made clear that participation is voluntary and participants can stop the experiment at any time. Finally, any experiment of this nature must have an appropriate Biomedical & Scientific Research Ethics Committee (BSREC) ethical approval explicitly mentioned in the information sheet and participants must be provided with an address for complaints about the study in case they wish to do so for any reason.

After the information sheet, participants are expected to be able to provide formal consent. For this purpose, it is often used a consent form (hard copy or virtual) where participants are again provided with contacts of the researchers. The form lists a number of points to confirm that they are legal adults, have understood the information sheet for the project, agree to participate and follow instructions provided, and understand how the provided information will be used as well as their participation is voluntarily. The consent form must have some sort of validation either via signature or (in case of electronic format) via a button validation where they validate that they do understand all the points mentioned.

4.4 Experimental task: newsvendor problem

4.4.1 Why newsvendor problem

This study is set to test hypotheses that behavioural biases and individual differences affect planners' decision making via a decision making experiment. The crucial component of the experiment is the task that will be used to measure performance depending on the subject's judgement. The newsvendor problem is commonly used in OM and OR research. Carlson and O'Keefe (1969, p.483) reported one of the first uses of the newsvendor problem in decision making experiments suggesting that '*subjects can make reasonably good decisions on an ad hoc basis without having been taught a formal rule'* but finding '*almost every kind of mistake also being made*'.

The newsvendor problem allows mathematical formulation and does have an optimal solution that maximises profit. This fulfils the essential task requirements in order to measure relative deviation in judgement assuming a normative view of rationality. Due to this characteristics, the newsvendor problem has been widely used in laboratory and field experiments (Schweitzer et al. 2000; Bostian et al. 2008; Bolton & Katok 2008; Becker-Peth et al. 2013; Moritz et al. 2013). Moreover, the sub-optimal performance in the newsvendor task has been attributed to behavioural factors such as loss, risk and waste aversion and underestimation of opportunity cost (Fisher & Raman 1996; Schweitzer et al. 2000; Agrawal & Seshadri 2000; Benzion et al. 2008; Ma 2008; Wang & Webster 2009; Herweg 2013; Ma et al. 2015) Hence, the characteristics and application of the newsvendor problem make it an appropriate candidate for modification to be used in a DP scenario.

For clarity, the newsvendor problem is not central to this research and although literature offers a rich insight on its theoretical formulations and empirical application, it is only used instrumentally to serve as basis for a DP task. Therefore, the following sections provide a brief overview of literature about the definition of the problem and empirical findings of its application.

4.4.2 Newsvendor problem origin

The newsvendor problem deals with a single period inventory and is an iterative purchase decision of a perishable item when facing uncertain demand (Whitin 1955). The origin of the newsvendor problem tracks back to Edgeworth (1888) who presented a very similar mathematical solution to a problem concerning how much cash banks had to keep to meet unknown customer's demand against the amount of cash to lend at a certain profit.

It is important to mention for disambiguation that the newsvendor problem has been referred to with slight variations by different authors. The 'newsvendor' problem is also known as 'newsboy' problem (Tiwari et al. 2011) also often referred to as newsvendor game (Ockenfels & Selten 2015), hence it is common to refer to decision-makers as players and the task itself as a game.

The newsvendor problem is one of the two typical situations in SCs. One type is a monopolist facing a downward-sloping demand from the market (Tsay et al. 1999) and the other type is a newsvendor situation facing random demand from the market and exogenous retail prices (Lariviere 1999). This is one of the limitations of the newsvendor problem applied to O&SCM. It represents only certain types of products (perishables) in a real life situation.

4.4.3 Decision making research using newsvendor

The newsvendor problem has been repeatedly used in laboratory experiments (e.g., Schweitzer & Cachon, 2000; Bostian et al., 2008; Bolton & Katok, 2008; Becker-Peth et al., 2013) to demonstrate that the assumption that real decision makers act in a way that maximizes their expected utility does not hold – decision makers' orders volumes systematically deviate from the profit maximising optimum order quantity. Subjects normally exhibit learning and convergence, affected by the mean demand, the size of the optimal order quantity and the demand of the last round, the order sizes tend to be between the mean demand and the optimal order quantity (Benzion et al. 2008).

Feedback and practice the newsvendor problem

Extended task experience improves performance but more frequent feedback does very little (Bolton & Katok 2008). In one separate experiment, frequent feedback had modest improvement on the performance (Wachtel & Dexter 2010). However, more frequent feedback resulted in a worse performance because it leads to excessive attention on more recent data and failure to compare information across time (Lurie & Swaminathan 2009). The more feedback is provided, the greater is the recency effect which comes from greater attention to the immediate information (e.g., Lurie & Swaminathan, 2009; Gavirneni & Isen, 2009; Wachtel & Dexter, 2010). This interpretation of excessive attention to more recent data aligns with myopia in mental accounting (Thaler 1999). Some studies suggest that on an individual level such behaviour can resist to professional experience or background training (Bolton & Katok 2008; Lurie & Swaminathan 2009; Bolton et al. 2012).

Main biases in newsvendor

Typically, on average across decision-makers, the ordered quantity sits between the average demand and the optimum (Schweitzer et al. 2000). This has been called pull-to-centre effect by ordering few low high profit products and too many of low profit products (Schweitzer et al. 2000). *'Pull-to-centre'* effect is defined as the average order quantities are too low when it should be high and vice-versa. (Bostian et al. 2008, p.590). Different interpretations have been offered to such behaviour (Su 2008; Kremer et al. 2010). The most common interpretation is demand chasing and anchoring effects. Order chasing effect is the bias describing the adjustment of order quantities based on the most recent demand. Anchoring effect is the tendency to anchor on the mean demand and then adjust towards the optimum, placing orders typically between the mean demand and expected profit-maximizing quantity. Most of the subjects are able to identify the over/under costs but fail to transfer this into the optimal order quantity (Gavirneni & Isen 2009).

Newsvendor and external factors

The newsvendor problem has been also used to consider how context factors affect performance (Rudi & Drake 2010). External factors can take many forms, such as available information, single versus group decision making, region and culture, and finally contracts and policies.

Regarding the available information, for example, knowing the demand distribution affects the behaviour but it does not lead to a better performance (Benzion et al. 2009). The effect of forecast in the newsvendor problem has been also observed as significant (Gurnani & Tang 1999; Zheng et al. 2016).

Expanding beyond single-player newsvendor, the newsvendor performance has been compared between individual decision making and groups (Gavirneni & Isen 2009) with a special focus on group dynamics (Gavirneni & Xia 2009). One of the main findings was that subjects working in groups exhibit less propensity for errors (Gavirneni & Xia 2009).

In an exploratory study with a relatively small sample size, regional and cultural factors have been observed as potentially significant in the newsvendor setting (Cui et al. 2013). Although Cui et al. (2013) warns to a small subject pool, the main findings are interesting, suggesting that Chinese players compared to Americans required more information, proposed new numbers instead of repeating previous ones and were more aware of the salvaging costs.

An important external factor can be contracts and policies. Experimental results suggest that performance of the newsvendor problem can be improved through contracts designed considering a behavioural model over contracts designed using standard models (Becker-Peth et al. 2013).

Finally, it has been demonstrated over an inventory control task (arguably sharing some similarities to the newsvendor) in experimental conditions that feedback format plays an important role in how well subjects perform (Atkins et al. 2002). Although Atkins et al. (2002) did not look specifically at the newsvendor problem, the conditions and the type of decisions that subjects had to make are potentially compatible and findings are applicable.

Individual differences in newsvendor

Considering individual differences in judgement, some interesting contributions have been arising over the last years suggesting that some traits can be good predictors of performance in the conditions similar to the newsvendor game. Strohhecker & Größler (2013) tested personal traits in an inventory management simulation game (sharing some communality with the newsvendor problem) concluding that intelligence was the strongest predictor of performance. Considering other individual traits such as intuition, training or experience, it is expected that experienced players can arrive at the solution of the newsvendor problem based on intuition (Bolton et al. 2012). However, previous experiments show that both MBA students and professional buyers deviate from the optimal (Brown & Tang 2006). Counter-intuitively, experimental results suggest that years of experience for professionals had a negative correlation with performance while the managerial position had a positive correlation instead (Bolton et al. 2012). Looking at other behaviour that is trait compatible, such as impulsiveness, newsvendor performance seems to be affected by impulses (Ockenfels & Selten 2015) and this tendency varies significantly between individuals (Bolton & Katok 2008).

Focusing on differences in professional background for two distinct groups, the behaviour of operating room managers working in planning is not significantly different to the behaviour exhibited by students, potentially suggesting that organisational and professional background play no role in the newsvendor performance (Wachtel & Dexter 2010).

Considering attitude to risk, it has been observed how subjects cope with loss aversion in newsvendor problem (Sun & Xu 2015). Comparing the risk averse and risk neutral newsvendors in a condition where the demand is a function of the price determined by the newsvendor, Agrawal and Seshadri (2000) found that comparing to a risk neutral newsvendor, a risk averse one will charge higher prices and order less when the price affects the scale of the distribution. However, in a condition when the change in price will only affect the location of the distribution the risk averse will charge a lower price compared to a risk neutral.

Performance affected by impulses, feedback and decision frequency

Emerging from a number of distinct contributions using newsvendor, different authors suggest that decision and feedback frequency (Lurie & Swaminathan 2009), as well as impulsivity (Ockenfels & Selten 2015) are important for the task performance. One of the perspectives on the newsvendor problem was that of the fact that *the 'newsvendors are driven by impulses which occur whenever there is an ex post inventory error'* and that by constraining the standing order for a sequence of periods moves the average orders towards the optimum (Ockenfels & Selten 2015, p.1). This points to two interesting aspects, first that of impulsivity trait being potentially significant[10] and second, to decision-frequency (or myopia) aligning with other contributions suggesting that less frequent interventions and longer commitment benefits overall performance. An alternative interpretation to ex post inventory error is demand chasing which is observed at the individual level (Kremer et al. 2010). An alternative way to improve performance was to offer fewer choices and placing the optimal order quantity in the middle rather than an extreme value leads to better performance, complementing the flat-maximum hypothesis. (Feng et al. 2011). Finally, restricting making quick decisions based on insufficiently large samples had a positive effect on the newsvendor performance (Bolton & Katok 2008). Again, highlighting the importance of individual differences, the tendency to make quick decisions based on insufficient data seems to vary depending on the individual level (Bolton & Katok 2008). This suggests two possible interpretations, first restrictiveness had a dampening effect on impulsivity and second it reduced the frequency of decisions which aligns with previous research (Ockenfels & Selten 2015).

4.4.4 Newsvendor formulation

In the newsvendor problem, the decision maker decides the order quantity of goods for the next selling period. The decision is on a single-period inventory

[10] To our best knowledge no research has tested the impulsiveness trait as predictor of performance in the newsvendor game using a relevant scale (*e.g.*, Barratt Impulsiveness Scale (Patton et al., 1995))

(Whitin 1955). The cost minimisation problem is only equivalent to profit maximisation when other variables external to cost minimisation are assigned optimal values (Whitin 1955)

The historical demand is known and the actual demand is updated but the future demand is unknown. Demand is usually generated from a uniform distribution (Schweitzer et al. 2000; Benzion et al. 2008; Bolton & Katok 2008; Bostian et al. 2008; Gavirneni & Isen 2009). In fact, most studies use uniform distribution (e.g., Schweitzer & Cachon, 2000; Bolton & Katok, 2008; Feng et al., 2011) and very few studies use normal distribution (Benzion et al. 2008; Benzion et al. 2009; Moritz et al. 2013)

In a typical formulation, the surplus (unsold inventory) has no value after the sales period. Behind this lies one of the basic assumptions of the newsvendor problem that assumes single-period decisions not influencing following periods as unsold stock is lost or has a salvaging cost. In the original newsvendor problem, the decision-maker chooses the order quantity q at the beginning of each selling period. D is the stochastic demand of the product with mean μ that is unknown before the end of the selling period. F is the distribution function of demand and f the density function. The two standard assumptions are that: (a) F is continuous, differentiable and strictly increasing, and (b) the decision maker has an unbiased forecast of demand distribution. The cost of each unit is c and selling price per unit is p, where $p>c$. In case of overstock ($q>D$), the salvage value of each remaining product unsold is s, where $s<c$. If the order quantity q is ordered, then min(q,D) units are sold. The number of unsold product is max(q-D,0). The cost of ordering is qc. $Profit(q,D)$ is the realised daily profit (Eq. 1):

$$Profit(q, D) = p \min(q, D) + s \max(q - D, 0) - qc \qquad \textbf{Eq. 1}$$

Optimal solution derived by Gallego (1995). From the normative perspective, it is assumed that the decision-maker wants to maximise profit. Therefore, the optimal order quantity q^* per period is given by Eq. 2:

$$F(q^*) = (p - c) / p \qquad \textbf{Eq. 2}$$

4.5 Measuring individual differences

Measures for individual differences offer a vast array of possibilities. This section lists the selected question together with a brief rationale. In general, the selected questions and psychometric scales include particular traits that have been reported previously as significant. It is worth noting that due to plethora of possibilities the selection of scales does not aim at being exhaustive including every single possible scale. Instead, the rationale is to use questions that have been previously used in similar experiments of decision-making. Included psychometric scales must have been previously validated at least once and published in high quality academic journals. It is also essential for the study to keep the wording in the scales exactly as it is in the original publication to ensure that this research is comparable with previous studies and publishable.

The aim is to measure individual differences most commonly mentioned in previous OR and OM literature as significant for planning-related tasks (e.g., forecasting, inventory management, purchasing). Table 8 provides a summary of the selected questions. The first part includes basic demographics together with some speciality questions. Follows a personality inventory (BIG five) and specific psychometric scales such as Barratt Impulsiveness Scale (Patton et al., 1995), Elaboration on Potential Outcomes (Nenkov et al., 2008) and General Decision Making Style (Scott & Bruce, 1995). The following sections will provide further detail about the selected questions to measure individual differences.

Table 8 Individual Differences Questions

Question	Pro	Student	Observations	Also used in
Gender Age	Yes	Yes	Standard demographic questions	(common to most, gender reported as significant in De Véricourt et al. 2013)
Background education	No	Yes	Students could be enrolled in different degrees. This question was omitted from the professionals questionnaire as it is potentially sensitive for professionals who could have progressed in the industrial career without formal education	(Franken & Muris 2005; Bolton & Katok 2008; Bolton et al. 2012; Moritz et al. 2013)
Managerial Level	Yes	No	Managers could be junior, middle and upper	
Active student of L&SCM (or related)	No	Yes	Students enrolled in speciality degrees such as Logistics and Supply Chain Management (or related) are assumed to have theoretical knowledge of DP	
Experience (in Planning)	Yes	No	If the subject has experience in planning and if so how many years. Assumed not applicable to students.	
Sector	Yes	No	Which sector they work in, for sample descriptive purposes	(Moritz et al. 2013)
BIG Five (mIPIP)	Yes	Yes	See 6.7Appendix A	(Donnellan et al. 2006)
Elaboration on Potential outcomes (EPO)	Yes	Yes	See 6.7Appendix B	(Nenkov et al., 2008)
Barrat Impulsiveness Scale (BIS)	Yes	Yes	See 6.7Appendix C	(Patton et al., 1995)
Global Decision Making Style (GDMS)	Yes	Yes	See 6.7Appendix D	(Scott & Bruce, 1995)

There are many different versions of questionnaires to measure personality traits. The Big Five, being one of the most commonly used and is composed of 60 questions. In order to make the assessment quicker, several reduced versions have been suggested (e.g., Gosling et al. 2003; Donnellan et al. 2006) The Mini International Personality Item Pool (mIPIP) was designed and successfully validated both as a 10 item and a 20 item version to assess the original constellation of traits defined by the Five Factor Theory of Personality (Donnellan et al. 2006). The full list of questions can be consulted in 6.7Appendix A.

One of the most basic individual differences are demographic-type items (e.g., gender, age). In line with previous studies (e.g., Haigh & List 2005; Brown & Tang 2006; Bolton & Katok 2008; Kremer et al. 2011; Moritz et al. 2013), some basic demographic questions are included together with some context specific questions. Worth noting, that professionals and students had two different sets of demographic questions.

To capture exposure, the assumption is that L&SCM students have theoretical exposure of DP while professionals have practical exposure. There was no control over which modules students attended and their results. Nor a knowledge test was performed to verify their theoretical knowledge. This can potentially constitute a limitation. If the assumption about theoretical exposure is incorrect, there will be no difference between students from specialised degrees (e.g., L&SCM) and students from any other degree (e.g., Art History). This is intrinsically incorporated in the exposure hypothesis that will be tested further.

4.6 Experimental design

4.6.1 Experimental design overview

The experiment was made up of two parts, first an incentivised decision making experiment and second a questionnaire (with students) and a survey (with professionals). It was carried out under the ethical approval by the Biomedical & Scientific Research Ethics Committee (BSREC) with reference REGO-2016-1736.

The methodological approach used in the study consists of two separate stages that should not be viewed as a unified randomized control trial. The main objective of the study is to test the MLA hypothesis via treatments to the decision-frequency relying on system's restrictiveness or systems guidance against a control group.

Table 9 Population and its sub-groups

Population	Incentives
Students	Incentivised decision making experiment
	Questionnaire
Professionals	Survey

The study targets two populations (listed in Table 9), students and professionals. Stage 1 of the study, relies on the methodology of a *laboratory experiment* where participants (all students at the university of Warwick) either participate in the incentivised decision making experiment (with monetary compensation) or a non-incentivised survey (without monetary compensation). Stage 2 uses a methodology of a *framed field experiment* in a form of a non-incentivised survey conducted with professionals. Worth noting that although stage 2 is a non-incentivised survey, it is important to distinguish that there is no performance-based monetary compensation. To attract participants to the study a non-monetary incentive for all interested is offered regardless of their performance.

The experimental design follows the approach of a seminal paper on comparison between laboratory and field experiments proposed by Haigh and List (2005) which since have been used in numerous top-published papers in business, management, economics, and social psychology literature. One of the aims is to compare behaviour of professionals with that of students in a non-incentivized (survey) setting using the Newsvendor task.

4.6.2 Incentives

In the previous literature on Newsvendor task, the behaviour of participants was measured sometimes in incentivised experiments and sometimes in non-incentivized surveys. While student behaviour in incentivised experiments and non-incentivised surveys appears to be similar, the first contribution of this study to the existing literature (achieved in stage 1) is to provide a comparison between these two conditions. The hypothesis is that providing incentives will not make any difference: i.e., it is expected to obtain similar results in incentivised experiment and non-incentivised survey. Should results in incentivised experiment be different from results in a survey, it would be necessary to conduct an additional (incentivised study) with professionals. Conducting an incentivised experiment is not the main aim of the proposed research.

Another rationale for conducting non-incentivized study with professionals is that while the budget[11] allows to offer sufficiently large monetary incentive to students (i.e., on average £10 per 30 minutes of work), this rate will not be sufficient to incentivise professionals who have much higher hourly rates. Much research in economics and social psychology (see e.g., Gneezy et al., 2011; Camerer et al., 2011) shows that in a field setting (when approaching professionals) much cleaner (less noisy) results were obtained when professional were asked to participate in non-incentivised (non-paid) surveys rather than in incentivised (paid) experiments. In this case, this research follows a well-established approach replicated in numerous studies.

4.6.3 Target sample size

For sample size calculation, the assumption is to work with approximately two standard deviation points, i.e., a minimum of 30 participants per treatment each providing 152 replies/observations in order to ensure statistical significance. Therefore, in order to reach the necessary statistical significance, it is necessary to recruit a sample size of at least 30 people per treatment. In total, the study must include at least 180 students for three treatments (90 paid and 90 non-paid) and 60 professionals (considering two-treatments). This follows the main principles proposed by List et al. (2011) of economics/social psychology experiment design principles which is based on maximizing the propensity of obtaining statistical significance.

4.6.4 Experimental treatments

The methodology of behavioural experiment was used by the core contribution by Haigh and List (2005) on MLA comparing laboratory and field

[11] Financial support for the study was provided by the Research Councils UK/Engineering and Physical Sciences Research Council grant EPL023911/1

experiments. The treatments are set to detect the presence of in DP. These treatments allow testing both Hypothesis 1 and 2 via comparison of means using Mann-Whitney-Wilcoxon (MWW) and Kruskal-Wallis Tests (KW Test).

Variations in treatment represent different DP policies via either system restrictiveness or system guidance (refer to Silver 1991), the base line has neither guidance nor restrictiveness. The following Haigh and List (2005) variation between treatments regards to frequency of decisions (or commitment length). The relationship between commitment period and decision frequency is inverse. Key differences from Haigh and List (2005) treatments is that first, feedback frequency is not altered, i.e., participants receive feedback each period; and second, a hybrid situation where participants are offered a recommendation and can choose to keep the previous decision or adjust it.

Table 10 Summary of experimental treatments

Treatment	Policy	Decision Frequency	Commitment	Decision Support	Description
T1	None	High	Short	n.a.	Baseline when subjects can make decisions each period without restrictiveness or guidance
T2	Binding	Low	Long	System's Restrictiveness	Subjects can only make decisions each third period, forced to order the same volume three times
T3	Non-binding	Varied	Hybrid	Decisional Guidance	Subjects are recommended to make decisions make decisions each third period but have the possibility of adjusting their choice

Each participant is assigned to a different treatment randomly when starting the experiment. Demand planners in treatment one can make decisions each period (Short Commitment) which also corresponds to high frequency decisions. In treatment two, planners make decisions three in three periods forced to keep the previous decision for second and third periods (Long Commitment) corresponding to low decision frequency. Finally, the third treatment consists of a hybrid situation where planners are recommended to keep the original decision for second and third rounds (Hybrid commitment) so the decision frequency varies. Considering DP policy (decision support) terminology, the three treatments can be referred to as unrestricted policy, restricted policy and guiding policy. The three treatments are summarised in Table 10 and represented in relation to planning periods in Figure 21.

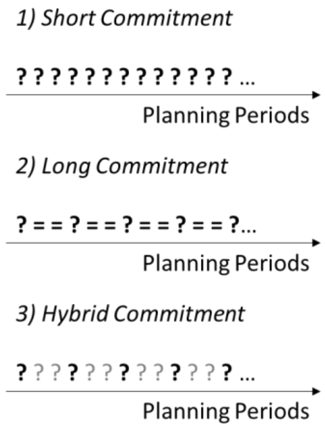


Figure 21 Representation of the experimental treatments

One particularity of the experiment in this study is that participants are primed with a forecast in all three treatments. They are instructed that the system's recommendation is a fixed amount per period. Unknown to participants is the fact that the recommended volume per period is the optimal solution to the problem. This is done to represent a common situation in practice when the system offers a recommendation and managers can either follow it or dismiss it. Since the problem focuses on mistrust, second-guessing and unnecessary reactions the provided recommendation must be the best possible to detect whether participants trust the recommendation considering all the information they have available.

4.6.5 Measuring performance and individual differences

The experimental task consists of a DP task (based on a modified newsvendor problem) where participants will be asked to plan consumer demand and make planning decisions. The survey is programmed in a web-based software Qualtrics. The experimental task was programmed from scratch along with the user interface (HTML), the Java Script code can be seen in 6.7Appendix F and the graphical interface together with the instructions in 6.7Appendix G. Considering user feedback and the discussion between the advantages of tabular versus graphical feedback, the decision was to include both in an interactive format. Participants had access to numerical feedback as well as a detailed graphical representation of the past demand and their decisions in a double-bar chart. The time-series graphic allowed zooming in and out, provided numerical feedback with mouse overlay. The participant is first presented with 9 practice rounds that do not count for the total profit, which is clearly identified in the interface in bold and red type font, followed by 30 decision rounds that count for the total profit. Feedback is provided after each round in a numeric and graphical format. However, the number of decisions they make during the 39 rounds will depend on which one of the three treatments they are assigned to at random when they launch the experiment. The number of rounds is decided based on multiples of three (following Haigh and List, 2005) with a relatively few practice periods, set to be less than one fourth of the total length of the experiment to mitigate learning biases and 'video-gaming' effect (Strohhecker & Größler 2013). The number of periods in previous applications of newsvendor problems varies significantly, ranging from 15 decision periods (e.g., Schweitzer & Cachon 2000) extending up to 100 periods (e.g., Bolton & Katok 2004). Greater number of periods allows greater experience which positively affects performance, however the improvement is on average very slow (Bolton & Katok 2004). Hence, the number of rounds was set to 30 for this experiment. Periods are abstract successive measures of time and are not timed to avoid inducing any time pressure to allow careful reasoning for as long as necessary (following suggestion in Größler 2004). Task information is kept standard, including standard provided in previous newsvendor settings (Benzion et al. 2008). In order to keep this as close to real-life situation as possible, characterisation of the demand distribution is excluded in order to address some of the main criticisms about unrealistic information in decision making experiments (e.g., Davern et al. 2008). Previous research shows that providing participants with the underlying demand distribution does not improve performance nor does it bring closer to the optimal solution (Benzion et al. 2009). In order to provide participants with some exposure to the underlying demand, practice rounds are provided.

After completing the DP task, participants are presented with some demographic questions and asked to answer several questions about their personality preferences using self-reporting scales[12]. For this, already validated and approved psychometric scales are used. They include the mIPIP (Donnellan et al. 2006) which can be consulted in 6.7Appendix A. Regarding specific personality constructs, it is included Barratt Impulsiveness Scale (BIS) (Patton et al. 1995) available in 6.7Appendix C, Elaboration on Potential Outcomes (EPO) (Nenkov et al. 2008) listed in 6.7Appendix B and finally, General Decision Making Style (GDMS) (Scott & Bruce 1995) as in 6.7Appendix D.

The above measures are used to test hypothesis about the contribution of individual differences to DP performance in the hybrid treatment (where participants have a choice). Specifically, to test hypothesis 3, 3A, 3B, 3C and 3D outlined in Chapter 3. This is achieved using regression analysis since all variables are numeric continuous. The main reason to choose regression analysis is that its outputs are easy to interpret. It is also one of the most commonly used analysis in the fields of psychology and economics.

4.6.6 Information sheets

In stage 1 of the study, students will be selected at random either to participate in an incentivised experiment or in non-incentivised survey through Decision Research at Warwick (DR@W) online recruitment system SONA[13] which to date has 1521 eligible participants who enrol in the study voluntarily. Much literature in economics and social psychology describes the dangers of changing the experimental information due to the so-called 'procedural invariance' effect (e.g., Loomes & Pogrebna, 2015). In previous studies with Newsvendor task, researchers either provided incentives or did not provide incentives and did not explain why incentives are provided or not. In fact, the mere mention of incentive provision in one group and no such

---

[12] Scales selected from the list of well-defined and accepted and previously validated psychological scales listed on http://www.sjdm.org/dmidi/ and http://ipip.ori.org/

[13] SONA https://warwick.sona-systems.com/default.aspx

provision in the other group can change participants' behaviour that (a) will make it impossible to compare the results of this study with existing results reported in the previous literature and, ultimately, (b) make this research unpublishable. For example, the results may be contaminated with envy considerations exhibited by participants in non-incentivized survey. The present study, therefore, relies on two separate Electronic Information Sheets (e.g., 6.7Appendix H) – one explaining the incentive provision (for paid experiment) and the other explaining the terms of participation in a non-incentivized survey. Participants are assigned randomly to either experiment or survey but then, in the invitation, are informed about all conditions of participation – i.e., all participants (in either experiment or survey) are fully informed about the conditions of participation and decide whether they want or do not want to take part. Therefore, participants in both experiment and survey have full information to provide informed consent. Participation in both experiment and survey is voluntary and participants can withdraw at any point in time.

4.6.7 Informed consent and anonymity

The consent for participation is gathered by requesting participants consent in the beginning of the survey (6.7Appendix I) where relevant information about the study is provided, allowing acceptance to continue or not. By pursuing with the experiment the participant is giving his/hers informed consent. This study uses an online consent form.

All data from the experiment is anonymised and answers are not associated with any personal information (name, address, etc.) from the participants. Participants will not receive any information about results and identities of other participants in this study. Participants' email addresses (should such addresses be provided for individual feedback) will be kept separate from the study data. All email addresses will be deleted after individual feedback is provided. The email address (if supplied) might reveal some personal information. The study data will include only an identification number for each participant. Once the questionnaire has been submitted, the data cannot be retrieved. However, if the participant opts to provide their email address, the record can be retrieved and deleted upon request.

4.6.8 Experimental procedure

The experiments are conducted online. The payments to the incentivised group of students (paid) is made to participants confidentially and in cash in person at the end of the experiment at agreed times between experimenters and participants. The earnings in the experiment are performance-based (the greater the total profit, the more they will earn). The estimate is that the mean earnings in the experiment to be 10 GBP that is equivalent to the standard experimental participation rate in the University of Warwick. The experimental task performance is measured in experimental tokens, which are converted into a monetary value (GBP) using a non-linear formula (Eq. 3).

$$\pounds\, Payment\ = \frac{\left(\frac{Total\ Profit}{10\ 000}\right)^3}{2} \qquad \textbf{Eq. 3}$$

Participants' personal data is confidentialised and unlinked, i.e., each participant is assigned a confidential ID number and participants' name and/or mail address is never associated with the data provided in experiments.

After results from the laboratory experiments are obtained and analysed, the proposed treatments will be repeated with actual planners (professionals) in the field experiment. Although professionals are expected to participate in a non-incentivised survey, there is still compensation offered. Practitioners are offered individual feedback and/or access to research report and findings, where participants can choose either or both. This optional compensation to professionals is offered just for participating regardless of their performance that should not be confused with performance-based monetary incentivisation used with students. Actual planners receive an invitation (6.7Appendix J) to take part in the online study along with the link to the experiment, attached to the mail invitation comes a sample individual report (6.7Appendix K). The recruitment of planners is done through the SCIP network[14], so participants will be directly invited to be part of the study. To broaden the reach of the call, invitation to participate in the experiment is disseminated across speciality groups around O&SCM particularly groups focused on planning. A full list of targeted groups can be consulted in 6.7Appendix E.

4.6.9 Eligibility criteria

Through all the treatments and groups it is applied the principle of non-discrimination: i.e., participants are not selected based on their gender, age, ethnicity or any other characteristic. It is required however, that all participants are adults of 18 years of age or older so that they could provide informed consent to study participation themselves. The online consent form (6.7Appendix I) is displayed in the beginning of the survey and the participant is informed explicitly that by continuing he or she agrees to its terms. The participants are expected to speak English sufficiently well to understand experimental instructions. These restrictions are clearly communicated to all potential participants prior to the study. Information about the study is provided in an Information Sheet (6.7Appendix H) that the participant can access at the beginning of the study.

4.6.10 Other ethical considerations: Risks and Benefits

This experiment is low-risk. It is non-invasive and only uses a survey based on a simple planning task and questions build from the list of well-defined and accepted psychological scales listed on http://www.sjdm.org/dmidi/ and http://ipip.ori.org/.

The data will be electronically stored on the University of Warwick secure server equipped with the latest safe authentication methodology and secure TLS tunnel for a period of ten years.

The major risk would be if Qualtrics servers are hacked and someone gains access to the data, similarly to if someone breaks into WMG offices at

[14] http://www2.warwick.ac.uk/fac/sci/wmg/research/scip/

University of Warwick and gets hold of the hard drive containing the raw data. Yet, these potential risks are minor since both Qualtrics and the University of Warwick have excellent cybersecurity rating and the precedence of confidentiality loss due to hacking have never happened before.

There are no immediate direct benefits to student participants in this research. However, this research will provide an opportunity to better understand how personality traits affect planning ability that will benefit the academic knowledge as well as industrial practice. Planners who take part in the field experiment may directly benefit from this research, as they will be able to understand how their personality affects their efficiency at work.

4.6.11 Experimental flow

Finally, the experiment follows the logic illustrated in Figure 22. At the very beginning after the information sheet and informed consent, the participant is assigned at random to one of the three experimental treatments. The participant is unaware of the alternative treatments. After the completion of the task, the participant responds to a questionnaire collecting and measuring data on individual differences. Two alternative sets of demographic questions are available, one for students and another for professionals. Follows four psychometric scales. The task is estimated to last from 15 to 30 minutes. At the end of the survey the participant (in the voluntary survey) can provide a mail contact and choose if he or she wishes to receive the individual report and/or access to a report on the main findings of the study.

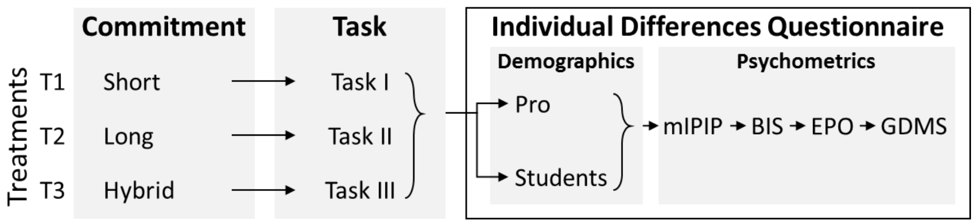


Figure 22 Experiment design

4.7 Chapter summary

The problem of sub-optimal DP decisions is partly attributed to behavioural issues considering both the individual as well as its relation with the decision support. To test the MLA and individual differences hypothesis an experiment is set following decision-making experiment methodology and relying on previously validated and widely accepted psychometric scales. The experiment targeted three distinct groups: professional and two groups of non-professionals (majority students) which parted in incentivised and non-incentivised groups. The experimental task is a modified newsvendor problem with unknown demand distribution but priming via a recommended order quantity (forecast) for each period with the optimum. The chosen uniform distribution is demand with equal probability of occurrence from 1 to 300 on a high-profit product (price of 12 and cost of 3) where the respective optimum planned volume is 225 per period. Participants go through 9 practice rounds followed by 30 rounds. The number of decisions made depend on the randomly assigned treatment at the beginning of the experiment. First treatment participants make decisions each period, second treatment participants are forced to keep the previous decision for second and third periods, and finally third treatment participants are recommended to keep the previous decision over to second and third periods. This corresponds to no decision support, system restrictiveness and decisional guidance respectively. Alternatively, considering commitment and decision frequency to high frequency short commitment, low frequency long commitment and finally varied frequency and optional commitment. The first and second treatments are designed to detect MLA while the third is expected to highlight potential personality traits affecting decision due to either offering a choice to follow or not the recommendation.

After the experimental task participants answer two distinct demographic question sets (for professionals and non-professionals) followed by common psychometric scales and personality inventory. Personality is measured using Mini-IPIP - Mini International Personality Item Pool (Donnellan et al., 2006) composed of 20 questions measuring measures 5 constructs (4 questions each). The measured constructs are: (a) Extraversion; (b) Agreeableness; (c)

Conscientiousness; (d) Emotional Stability and; (e) Imagination. Impulsiveness is measured using BIS - Barratt Impulsiveness Scale (Patton et al., 1995) over 30 questions. The BIS questions measure 3 first order constructs with 6 second order constructs (2 each). Follows EPO - Elaboration on Potential Outcomes (Nenkov et al., 2008) with 13 questions to measure 3 constructs (6, 3 and 4 questions per construct). Finally, GDMS - General Decision Making Style (Scott & Bruce, 1995) uses 25 questions to measure 5 constructs (5 questions each construct). The questionnaire finishes with the suggestion to leave a contact address so individual feedback and early access to results can be sent back if the participant wishes so. On overall, the experiment is estimated to last 15 minutes on average.

This thesis contributes methodologically to the newsvendor problem literature with an innovative design. The newsvendor problem allows variation of the planning horizon/ decision frequency. Treatments are enforced via different DP policy. Finally, the experimental design incorporates personality inventories and psychometric scales which has not been tested before with the newsvendor problem.

This research is carried out under the ethics approval from Biomedical & Scientific Research Ethics Committee (BSREC) reference REGO-2016-1736.

The following chapter reports on the results and analysis.

5 Results and analysis

5.1 Introduction

Following the previously defined research design, the following section focuses on the results and analysis. The structure of Chapter 5 is outlined in Table 11. After this introduction, the resulting sample is broken into groups based on professional background, incentives and education. The performance of incentivised and non-incentivised groups is compared followed by the analysis of performance across treatments to validate the MLA hypothesis. Finally, the analysis focuses on the hybrid treatment to observe what are the traits predicting performance. This is done for sophisticated students (L&SCM students), professionals and overall.

Table 11 Chapter 6 structure

Section	Overview
5.2 Overview of the framework and experimental treatments	The previously defined framework is reviewed because it provides the structure for the analysis
5.3 Resulting experimental	Provides a summary of the participants and the sub-groups. A total of 339 participants after clearing incomplete responses with 222 students and 117 non students of which 84 are professional planners.
5.4 Incentives and performance	As the choice of incentives is deeply rooted into decision-making, one of the first analysis concerns the effect of incentives on performance.
5.5 Performance between treatments	The performance between treatments is analysed across groups and treatments
5.6 Individual differences as predictors of performance in hybrid treatment	Focusing on the human and system interactions the analysis of the hybrid treatment (optional policy) participants allows the understanding of what attributes are predictors of performance. The analysis is first done over the sophisticated students, followed by professional planners and finishes with an overview of individual differences and performance

5.2 Overview of the framework and experimental treatments

The conceptual framework with the hypothesis is illustrated in Figure 23. Following the descriptive analysis of the resulting samples, performance is compared between treatments and between groups to test Hypothesis 1, Hypothesis 2 and partially Hypothesis 3 (H3A). Regression analysis is then used to understand differences in performance in the hybrid treatment. This allows testing the remaining of Hypothesis 3 (including H3B, H3C and H3D) focusing on individual differences and personality traits.

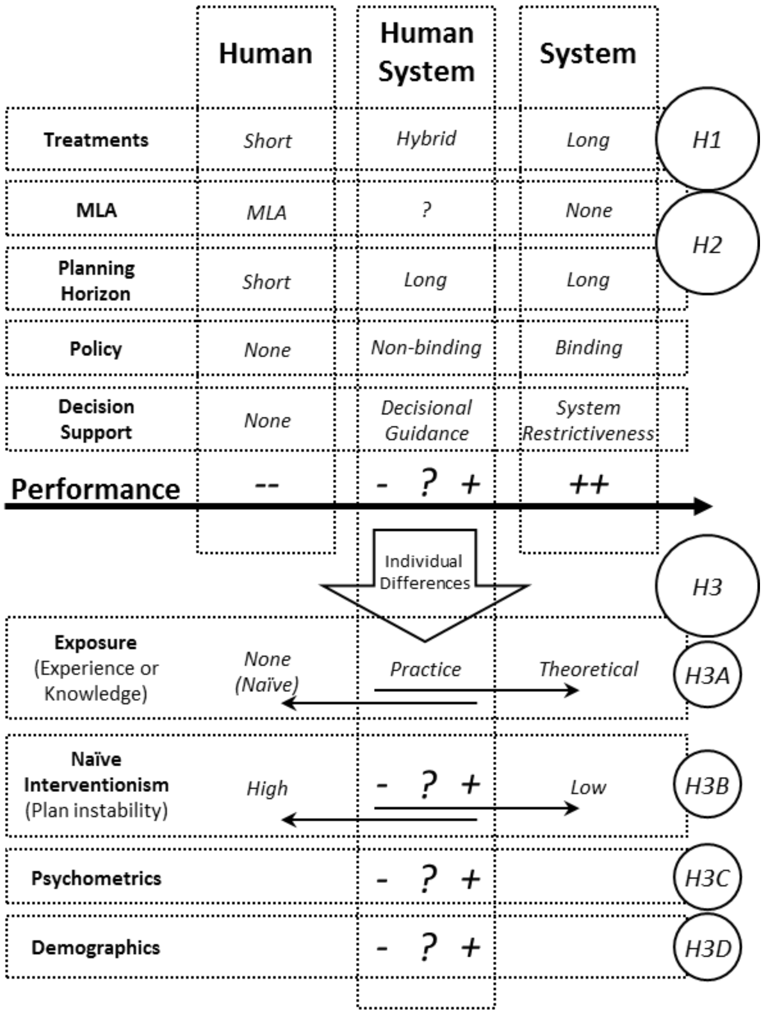


Figure 23 Conceptual Framework, Experimental treatments and hypothesis

5.3 Resulting experimental sample

Overall, 339 participants took part in the study (see Figure 24). The ultimate goal was to understand whether and to what extent professional planners were prone to MLA and whether MLA could explain planners' decisions. However, it was also important to conduct a baseline study with a sample of 'naïve' people. Naïve people did not have previous exposure to planning tasks. The aim was to later compare naïve with professionals. The average number of years of experience for professional planners was approximately 10 years and 4 months. For the purposes of this study, the recruited 'naïve' people were students who did not have planning experience in the industry.

Therefore, of the 339 participants, 222 were students at the University of Warwick while 117 were non-students (professional planners, consultants, non-academic researchers, and other people who did not belong to the student population).

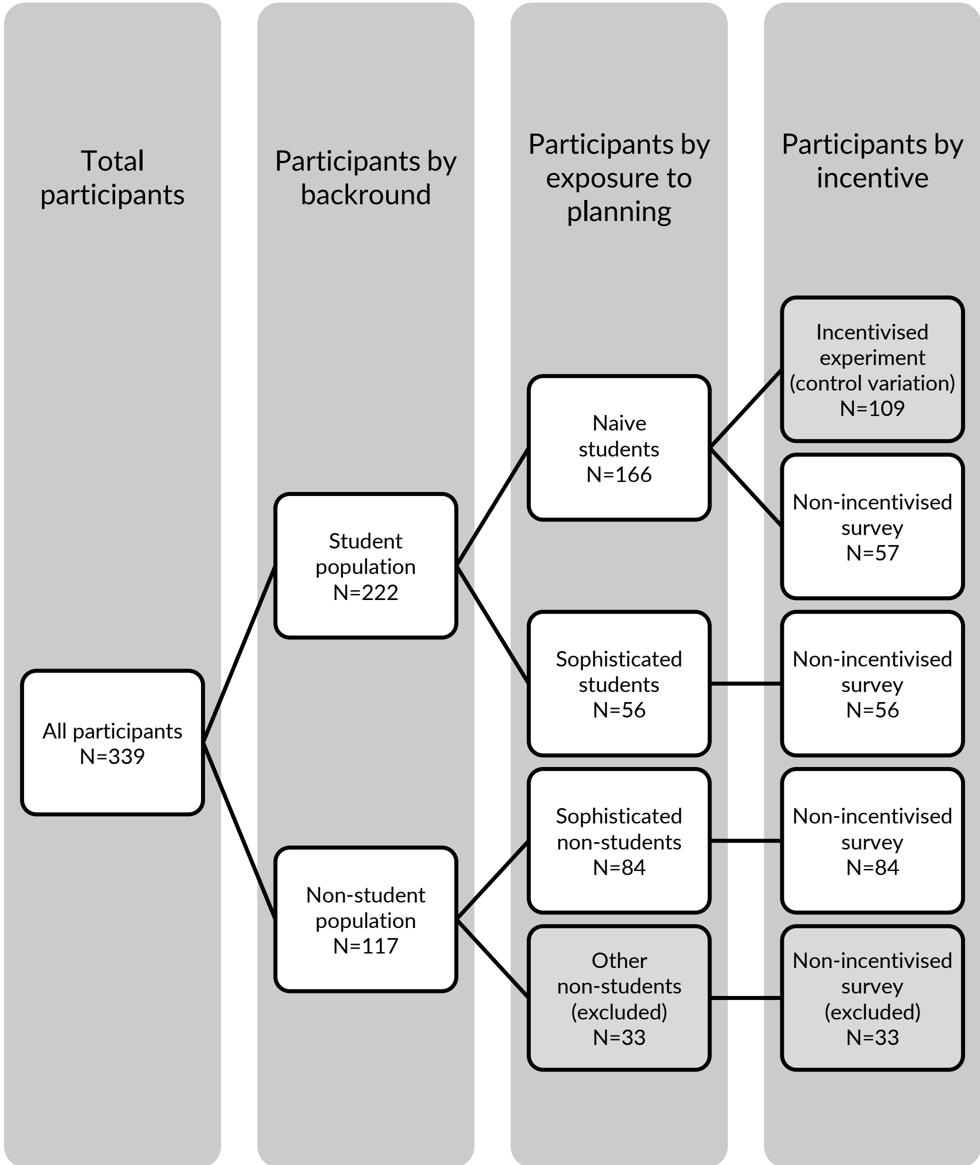


Figure 24 Study Overview: Sample and Incentives

To introduce several layers of comparison with professionals, two types of individuals are distinguished within student sample: '*naïve*' and '*sophisticated*'. Naïve students did not have any exposure to planning tasks before while sophisticated students were majoring in Logistics and Supply Chain Management (L&SCM) and, therefore, could have been familiarised with planning contexts within the frame of their educational programme. Since both types of students did not have practical experience in planning, having both naïve and sophisticated group of unexperienced subjects allowed not only to compare experienced (professional planners) and unexperienced (students) samples, but also to understand which of the two factors - theoretical or practical knowledge of planning - affects planning performance. By theoretical knowledge, it is meant education in Logistics or SCM which was the main differentiating characteristic of sophisticated students in the study. By practical knowledge, it is meant first-hand industrial experience in planning which was the main distinguishing characteristic of professional planners in the study.

Naïve and sophisticated students were compared with the sample of sophisticated non-students - professional planners (84 participants). It was also collected data from 33 non-student participants. However, their background was both too diverse to be combined in one sample. This subsample of participants included non-planners from industry, SCM researchers, and SCM consultants. It was also too vague to understand whether their exposure to planning allowed classifying them either as naïve or as sophisticated non-students. It was often unclear whether researchers and consultants had practical planning experience in industry. Therefore, observations from these 33 individuals were excluded from the analysis. The resulting subject pool consisted of 306 participants. Amongst these 306, three samples are distinguished: naïve students (166); sophisticated students (56), and professional planners (84).

A non-incentivised survey with professional planners was used because it would be difficult to design monetary incentives that would be sufficient to properly incentivise professional planners in the experimental planning task (newsvendor problem). The goal was to compare the performance and behaviour of professional planners with naïve and sophisticated students in non-incentivised study. However, since the unexperienced sample was student sample and the overwhelming majority of studies with students used incentive-compatible mechanisms, it was important to understand whether behaviour in incentivised setting was different from non-incentivised setting.

Hence, it is introduced incentivised and non-incentivised variations of planning task for naïve students' sample in order to (i) have a good comparison of the results with previous literature and (ii) understand whether monetary incentives made a difference for performance in planning task. In other words, it is not the purpose of this study to compare experienced and unexperienced participants under incentivised and non-incentivised conditions. Rather, the aim is to test the theory of MLA in planning decisions under different time horizons with experienced agents (professionals), controlling (a) for the role of experience (comparison with naïve students) and education (comparison with sophisticated students) as well as (b) for the role of incentives (comparison between incentivised and non-incentivised implementation of planning task).

In the following analysis, the measure of total profit (payoff) obtained by participants in the planning task (either real monetary profit or hypothetical profit) is used as a measure of performance in the planning task. This total profit, therefore, is the dependent variable in all estimations reported below (unless specified otherwise).

5.4 Incentives and performance

The analysis starts with the population widely studied in the previous literature: a sample of students without previous planning experience that is called 'naïve' students. A total of 109 received an incentivised planning task and 57 received the same planning task as a hypothetical (non-incentivised) problem.

Table 12 Naïve Students: Comparison of Incentivised and Non-Incentivised Experiment

Treatment	Incentivised (N participants, mean profit and standard deviation in points)		Non-incentivised (N participants, mean profit and standard deviation in points)		Mann-Whitney-Wilcoxon test results
Short	32	24,683.91 (3,639.01)	21	25,108.00 (2,523.54)	z = 0.364 p = 0.7161
Long	41	26,660.63 (3,301.98)	12	27,104.75 (1,944.97)	z = -0.234 p = 0.8152
Hybrid	36	24,040.33 (4,250.46)	24	23,627.00 (2,748.06)	z = -0.573 p = 0.5664
Total	109	25,214.89 3,876.98	57	24,904.79 (2,805.61)	z = -0.889 p = 0.3738

Table 12 shows that in the sample of naïve students provision of monetary incentives does not influence performance in the planning task. This is true both for the overall comparison of profits of participants who completed the planning task with (109) and without (57) incentives (Mann-Whitney-Wilcoxon, or MWW, test p=0.3738) and for each of the three treatments: Short (MWW test p=0.7161), Long (MWW test p=0.8152), and Hybrid (MWW test p=0.5664). Since there are no statistically significant differences between performance in the planning task with and without incentives, the data for all naïve students can be pooled. Both incentivised and non-incentivised data can be used for comparison with other experimental samples (sophisticated students and professional planners).

Since the difference between incentivised and non-incentivised planning tasks is not statistically significant in the most neutral sample in the experiment, no statistically significant differences in other samples are expected. In order to confirm the conjecture, a series of comparisons are conducted between performances of the incentivised planning task versus non-incentivised task among all samples. In other words, performance of 109 naïve students who received the incentivised planning task was compared with all other samples - naïve students (57), sophisticated students (56) and professional planners (84) - who played without incentives (197 participants in total). No statistically significant differences are found (a) between incentivised and all non-incentivised samples taken together (109 vs 197 participants, MWW test p = 0.1315). Similarly, no statistically significant differences are found (b) by treatment (all MWW test probabilities are greater than 0.06). This analysis allows to pool data obtained from incentivised and non-incentivised planning tasks together to conduct further analysis.

5.5 Performance between treatments

The following analysis focuses on experimental treatments and testing of the hypotheses. As explained above, the following three following samples are considered: naïve students (pooled together across incentivised and non-incentivised variations of the planning task); sophisticated students (students majoring in SCM and/or logistics who completed non-incentivised planning task), and sophisticated non-students (professional planners who also took part in non-incentivised planning task). A series of non-parametric tests are conducted to (i) test whether and to what extent different evaluation periods (restricted by policy) influence performance and (ii) how different samples in the experiment react to changes in policy. Results of the analysis are reported in Table 13.

Table 13 Comparison of Performance in Planning Task by Treatment and Sample

Treatment	Sample			Total	MWW & Kruskal-Wallis Test results
	Naïve students	Sophisticated students	Sophisticated non-students		
Short	24,851.94 (3,222.99) *24,669.00* N=53	25,921.58 (3,073.68) *25,896.00* N=19	24,372.00 (3,181.95) *24,195.00* N=20	24,968.51 (3,192.85) *24,864.00* N=92	**Naïve students vs Sophisticated students:** z= -1.207 p= 0.2273 **Sophisticated students vs sophisticated non-students:** z= 1.447 p= 0.1478 **Naïve students vs sophisticated non-students:** z= 0.495 p= 0.6208 **Kruskal-Wallis test:** chi-squared = 2.239 p= 0.3264
Long	26761.19 (3,036.84) *27,552.00* N=53	28,373.31 (2,303.58) *28,695.00* N=13	27,138.50 (2,979.78) *27,975.00* N=36	27,099.82 (2,953.81) *27,892.50* N=102	**Naïve students vs Sophisticated students:** z= -2.056 p= 0.0398 **Sophisticated students vs sophisticated non-students:** z= 1.585 p= 0.1129 **Naïve students vs sophisticated non-students:** z = -0.660 p = 0.5090 **Kruskal-Wallis test:** chi-squared = 4.349 p= 0.1137
Hybrid	23875.00 (3,701.76) *23,860.50* N=60	27,574.13 (2,118.48) *27,922.50* N=24	25,095.75 (3,941.80) *24,498.00* N=28	24,972.86 (3,756.51) *25,371.00* N=112	**Naïve students vs Sophisticated students:** z= -4.624 p= 0.0000 **Sophisticated students vs sophisticated non-students:** z= 2.331 p= 0.0197 **Naïve students vs sophisticated non-students:** z= -1.407 p= 0.1596 **Kruskal-Wallis test:** chi-squared = 20.123 p = 0.0001
Total	25,108.41 (3,540.03) *25,350.00* N=166	27,198.96 (2,660.36) *27,757.50* N=56	25,798.89 (3,542.90) *26,131.50* N=84	25,680.54 (3,475.19) *26,053.50* N=306	**Naïve students vs Sophisticated students:** z= -4.043 p= 0.0001 **Sophisticated students vs sophisticated non-students:** z= 2.233 p= 0.0255 **Naïve students vs sophisticated non-students:** z= -1.697 p= 0.0897 **Kruskal-Wallis test:** chi-squared = 16.519 p = 0.0003
MWW test results	**Short vs Long:** z= -3.153 p= 0.0016 **Long vs Hybrid:** z= 4.355 p= 0.0000 **Short vs Hybrid:** z= 1.602 p= 0.1091	**Short vs Long:** z = -2.436 p = 0.0148 **Long vs Hybrid:** z= 1.575 p= 0.1153 **Short vs Hybrid:** z= -1.883 p= 0.0597	**Short vs Long:** z= -3.300 p= 0.0010 **Long vs Hybrid:** z= 2.098 p= 0.0359 **Short vs Hybrid:** z= -0.544 p= 0.5866	**Short vs Long:** z= -4.930 p= 0.0000 **Long vs Hybrid:** z= 4.493 p= 0.0000 **Short vs Hybrid:** z= -0.116 p= 0.9080	

* Each cell of the table shows mean profit (standard deviation), *median profit*, and N number of participants.

First, the difference between different treatments is considered. It is obvious from Table 13 that performance of study participants in the long treatment is better than that in other treatments. Indeed, the overall results (across all 3 samples) show that participants earn more points in the Long treatment (27,099.82) than in the Short treatment (24,968.51). This difference is statistically significant (MWW p=0.0000). This result confirms Hypothesis 1 that says that: Planning performance in the planning *option Long will be better than that in the planning option Short (i.e., planners will achieve higher profit in Long compared to Short).*

Performance in the Long treatment (across all samples) is also better than that in the Hybrid treatment (24,972.86). This difference is also statistically significant (MWW p=0.0000). However, there is no statistically significant differences between the Short treatment and the Hybrid treatment (MWW p=0.9080). These results partially confirm Hypothesis 2 that says that factors other than treatment variation is likely to influence performance in the Hybrid treatment. Yet, this result also suggests that, across all samples, participants react to policy: restricting the planning period to Long evaluation (binding policy) helps to improve planning performance and earn higher profit/minimise losses.

The comparison of different samples shows results that are more interesting. In the Short treatment (KWallis test p=0.3264) and the Long treatment (KWallis test p=0.1137) performance is similar among all three samples in this study (naïve students, sophisticated students, and professional planners). This suggests that irrespective of theoretical and practical exposure to planning, policy equally influences performance of sophisticated and naïve people. Restricting evaluation period increases performance in all samples, while unrestricting leads to overadjustment in all samples. However, in the Hybrid treatment samples are dissimilar (KWallis test p=0.0001): specifically, sophisticated students seem to perform better than both naïve students (MWW p=0.0000) and professional planners (MWW p=0.0197). At the same time, the difference between performance of professional planners and naïve students in the Hybrid treatment is not statistically significant (MWW p=0.1596). This pattern from the Hybrid treatment also drives test results across all treatments where professionals and naïve students also appear similar and sophisticated students show better performance. Looking at between-treatment comparisons within each sample helps to explain this result (see last row in Table 13).

Results conducted for each sample separately show that naïve students and professionals behave as if the Hybrid treatment is similar to the Short treatment and dissimilar to the Long treatment, while sophisticated students behave as if Hybrid treatment is similar to the Long treatment and dissimilar to the Short treatment. Since the Long treatment improves performance, sophisticated students outperform other samples in the Hybrid treatment because they are mimicking the Long treatment in the Hybrid treatment while other samples are mimicking the Short treatment in the Hybrid treatment (see Table 13 for test results).

This partially confirms Hypothesis 3 and fully confirms Hypothesis 3-A. Overall, the analysis shows that all participants are influenced by policy: the Long evaluation period improves performance which is consistent with the MLA hypothesis. At the same time, sophisticated students (Theory planners) perform better than other samples (Naïve planners and Practice planners), yet, this performance difference is primarily due to their behaviour in the Hybrid treatment. Sophisticated students (Theory Planners) seem to be better than other samples because in the Hybrid treatment they behave similarly to the Long treatment which is consistent with the Exposure-Performance Matrix and Hypothesis 3-A. In the next sections, it is explored whether other individual characteristics influence planning performance and test hypotheses 3-B, 3-C, and 3-D.

5.6 Individual differences as predictors of performance in hybrid treatment

The aim of following analysis is to understand what individual differences can explain variation in performance in the Hybrid treatment since variation in the Short and the Long treatments can be fully captures by MLA and policy (as proved by the non-parametric analysis reported above). The same three samples are considered: naïve students; sophisticated students (L&SCM students), and sophisticated non-students (professional planners). It was previously observed that sophisticated students significantly outperform both the naïve subjects and the professional planners, while naïve subjects and professional planners perform similarly.

Each participant provided 30 decisions; the quality of these decisions determined the total profit. Naïve interventionism is calculated as the mean absolute deviations in the planned demand volumes. The assumption is that the rational approach is to maintain the optimal and stick to the suggested forecast without deviations. The task was followed by a questionnaire developed to capture individual differences. The Big 5, EPO, BIS and GDMS all provide numeric scores derived from Likert scales for its constructs and sub-constructs. The selected psychometric scales have been successfully validated in previous literature and the stability analysis is listed in 6.7Appendix M. Additional individual differences are either categorical (e.g., sex), ordinal (e.g., managerial level in the organisation) or continuous numeric (e.g., age and years of experience). Therefore, simple OLS regressions are conducted followed by clustered multinomial logit regression to (i) test whether individual differences influence performance and (ii) how sophisticated students differ from naïve students and from sophisticated non-students (professionals). In order to obtain a more detailed insight into individual heterogeneity, clustered OLS regressions are used to expand on (i) whether individual differences influence performance. Finally, to test the differences between genders, MWW tests are used to compare performance between male and female subjects both across group and across treatments.

First, the overview of summary statistics of the individual differences (6.7Appendix L) suggest that groups differ (at least numerically) from each other. The variation between groups seems to suggest that some individual characteristics may potentially predict performance considering that previous analysis has shown that sophisticated students significantly outperformed the rest. For example, Naïve students exhibit on average higher level of naïve interventionism compared to both sophisticated students and non-students. Specifically, across all treatments, naïve students tend to make average jumps in their adjustments of predicted demand in the newsvendor problem equal to 33%, while sophisticated students make only 22% jumps and professional planners – 24% jumps. Professional planners exhibit greater levels of extroversion, conscientiousness, rational decision-making style than other samples. Yet, professional planners also exhibit lower levels of negative outcome focus, impulsiveness and avoidant decision-making style than other samples.

At this point, it is difficult to tell how any of these characteristics affect performance. To do so, a series of regressions considering different hypotheses are conducted. Since the previous analysis shows that treatment variation and exposure to planning are important determinants of performance in the Short and the Long treatment (as proved by the results of non-parametric tests), the analysis will concentrate on the Hybrid treatment. Regression results are presented for each sample of subjects separately (Table 14; Table 15 and; Table 16) because non-parametric tests in the previous sections show significant differences between samples.

Table 14 Regression equations estimated by an OLS for Hybrid treatment (dependent variable – payoff in Newsvendor Game) – A: Naïve students

Explanatory Variable	**Model 1 (NI)**	**Model 2 (BIG5)**	**Model 3 (EPO)**	**Model 4 (BIS)**	**Model 5 (GDMS)**	**Model 6 (ALL)**
Naïve interventionism (NI)	-5706.31*** (974.89)	-	-	-	-	-5963.36*** (1092.28)
Big 5 Extroversion	-	-2506.60 (3276.93)	-	-	-	-3025.609 (2892.909)
Big 5 Agreeableness	-	216.6763 (3543.534)	-	-	-	3177.809 (3711.617)
Big 5 Conscientiousness	-	3006.368 (3367.001)	-	-	-	-1618.553 (3438.331)
Big 5 Neuroticism	-	-869.1244 (3341.483)	-	-	-	171.6963 (3938.321)
Big 5 Intellect/Imagination	-	2968.267 (3455.163)	-	-	-	-1822.957 (3065.983)
EPO Generation & Evaluation	-	-	7943.469* (3531.4)	-	-	7546.506 (4544.1)
EPO Positive Outcome Focus	-	-	-4514.605 (2921.826)	-	-	-2955.272 (2921.234)
EPO Negative Outcome Focus	-	-	-4761.766 (3352.041)	-	-	-2351.115 (3616.497)
BIS (overall)	-	-	-	-11793.83 (5958.922)	-	-3431.408 (8160.028)
GDMS Rational	-	-	-	-	1105.01 (5021.967)	-4973.72 (5757.708
GDMS Intuitive	-	-	-	-	5956.937 (4200.287)	6264.592 (3958.502)
GDMS Dependent	-	-	-	-	1764.789 (3706.795)	-4653.381 (4276.654)
GDMS Avoidant	-	-	-	-	-4246.261 (3389.063)	1444.019 (3576.637)
GDMS Spontaneous	-	-	-	-	-7571.484 (4310.351)	-7183.17 (4561.824)
Constant	26263.09*** (559.0226)	21551.15*** (6030.715)	24202.16*** (3398.448)	30211.51*** (3227.557)	24549.71*** (5144.698)	33970.37*** (6860.123)
R-squared	0.3713	0.0439	0.0492	0.0643	0.1522	0.5276
N (observations)	60	60	60	60	60	60

*** Significant at 0.001 level; ** Significant at 0.01 level; * Significant at 0.05 level

Table 15 Regression equations estimated by an OLS for Hybrid treatment (dependent variable – payoff in Newsvendor Game) – B: Sophisticated Students

Explanatory Variable	Model 1 (NI)	Model 2 (BIG5)	Model 3 (EPO)	Model 4 (BIS)	Model 5 (GDMS)	Model 6 (ALL)
Naïve interventionism (NI)	-11161.4*** (2118.75)	-	-	-	-	-5542.33* (1858.579)
Big 5 Extroversion		-756.344 (3528.379)	-	-	-	3246.804 (1539.705)
Big 5 Agreeableness	-	-969.5214 (5176.804)	-	-	-	17012.12*** (3214.753)
Big 5 Conscientiousness	-	-3510.974 (4615.684)	-	-	-	-7930.929** (2122.356)
Big 5 Neuroticism	-	313.4589 (4711.668)	-	-	-	-12634.22** (2927.043)
Big 5 Intellect/Imagination	-	1549.179 (4141.662)	-	-	-	-8597.581** (1957.442)
EPO Generation & Evaluation	-	-	-6030.569* (2535.889)	-	-	-6883.546** (1862.641)
EPO Positive Outcome Focus	-	-	-3302.782 (3306.717)	-	-	-8491.019** (2140.269)
EPO Negative Outcome Focus	-	-	-1932.393 (3874.16)	-	-	-2777.076 (2609.604)
BIS (overall)	-	-	-	1951.031 (7137.208)	-	-11899.88 (6218.276)
GDMS Rational	-	-	-	-	-8190.567 (5418.701)	-13411.05* (4276.653)
GDMS Intuitive	-	-	-	-	2131.515 (4478.365)	13752.97** (3132.54)
GDMS Dependent	-	-	-	-	-1923.659 (4920.961)	486.3296 (3754.405)
GDMS Avoidant	-	-	-	-	-7021.832 (4110.47)	-12185.3** (2655.465)
GDMS Spontaneous	-	-	-	-	-1611.847 (4694.692)	-5415.894 (2510.59)
Constant	30038.93*** (552.6039)	29887.52*** (4835.042)	35148.96*** (4074.904)	26472.61*** (4053.653)	38498.71*** (5311.65)	62176.99*** (6087.501)
R-squared	0.5578	0.0778	0.2989	0.0034	0.3100	0.9446
N (observations)	24	24	24	24	24	24

*** Significant at 0.001 level; ** Significant at 0.01 level; * Significant at 0.05 level

Note: note that in Model 6 the majority of variables become significant. This is most probably because there are 24 independent observations and estimate a large number of variables (14 variables + constant). The fact that the majority of these variables are not significant in Models 1-5 shows that personal characteristics generally do not tend to play an important role in the performance of sophisticated students.

Table 16 Regression equations estimated by an OLS for Hybrid treatment (dependent variable – payoff in Newsvendor Game) – C: Professional Planners

Explanatory Variable	Model 1 (NI)	Model 2 (BIG5)	Model 3 (EPO)	Model 4 (BIS)	Model 5 (GDMS)	Model 6 (ALL)
Naïve interventionism (NI)	-7574.20*** (1269.01)	-	-	-	-	-4561.267 (2251.649)
Big 5 Extroversion		9412.602 (5841.343)	-	-	-	4485.43 (3922.912)
Big 5 Agreeableness	-	-8177.142 (8447.866)	-	-	-	1299.646 (6675.517)
Big 5 Conscientiousness	-	-2349.962 (6849.004)	-	-	-	-2784.686 (5071.057)
Big 5 Neuroticism	-	5150.837 (4828.676)	-	-	-	5185.63 (5334.191)
Big 5 Intellect/Imagination	-	-1822.596 (7073.359)	-	-	-	3357.99 (6359.43)
EPO Generation & Evaluation	-	-	-8769.407 (5871.844)	-	-	3133.497 (5604.055)
EPO Positive Outcome Focus	-	-	4324.502 (5650.728)	-	-	9643.395 (4866.71)
EPO Negative Outcome Focus	-	-	3678.106 (5065.21)	-	-	1977.419 (5756.445)
BIS (overall)	-	-	-	17383.89 (9841.517)	-	14979.29 (14387.6)
GDMS Rational	-	-	-	-	-14420.73* (6840.711)	-8977.799 (10414.98)
GDMS Intuitive	-	-	-	-	-9757.474 (5976.658)	-8880.252 (5480.624)
GDMS Dependent	-	-	-	-	13910.63* (5484.582)	-216.3101 (6084.171)
GDMS Avoidant	-	-	-	-	-6015.357 (4697.421)	-8870.179 (5870.254)
GDMS Spontaneous	-	-	-	-	1122.701 (5169.517)	8438.736 (6868.705)
Constant	27665.57*** (654.6099)	25848.38*** (9094.01)	26922.86*** (7620.96)	15907.12** (5251.166)	35504.71*** (8769.427)	13644.45 (13227.43)
Adjusted R-squared	0.5781	0.1373	0.1125	0.1071	0.4771	0.8440
N (observations)	28	28	28	28	28	28
*** Significant at 0.001 level; ** Significant at 0.01 level; * Significant at 0.05 level						

The regression analysis reveals several interesting results (refer to Table 14; Table 15 and; Table 16). First, naïve interventionism (NI) is a significant determinant of planning performance (payoff in Newsvendor game) for all samples. This supports hypothesis 3-B. However, this variable cannot explain why sophisticated students do better than other samples in the experiment because for all samples the effect of NI goes in the same direction. Specifically, the higher the NI, the lower the payoff. In other words, the more planners adjust their plan, the worse planning performance they achieve (across all samples). Moreover, NI is not always robustly significant in all regressions. For example, considering the sample of professional planners, NI is significant in Model 1 but not in Model 6. In order to understand better the contribution of NI, further analysis is required.

From the previous analysis with non-parametric tests follows that sophisticated students (Theory planners) perform better than other samples in the Hybrid treatment. One personal characteristic which may explain the differences between samples is EPO generation/evaluation variable which is one of the three components of the Elaboration on Potential Outcomes (EPO) scale proposed by Nenkov et al. (2008). EPO generation and Evaluation construct evaluates the extent to which the individual considers potential consequences of his/her actions. For naïve students (Naïve planners) this variable is positive and significant, for sophisticated students (Theory planners) it is negative and significant and for professional planners (Practice planners) it is negative but not significant.

In order to explore the differences between samples, a series of multinomial logit regressions are conducted. Individual characteristics are used as explanatory variables and sample (Naïve or Practice planners) as dependent variables. Since the interest is how Naïve planners (naïve students) and Practice planners (professional planners) differ from Theory planners (sophisticated students), Theory planners is the base category in the multinomial logit regressions (Table 17 and Table 18).

Table 17 Results of the Multinomial Logit Regressions for Hybrid Treatment (dependent variable – sample; base category – Theory planners/sophisticated students) – A: Naïve Students (Naïve planners) vs Sophisticated Students (Theory planners)

Explanatory Variable	Model 1 (NI)	Model 2 (BIG5)	Model 3 (EPO)	Model 4 (BIS)	Model 5 (GDMS)	Model 6 (ALL)
Naïve interventionism (NI)	1.086173 (0.5841851)	-	-	-	-	1.120055 (0.6221477)
Big 5 Extroversion		.0838016 (0.9567409)	-	-	-	0.7111813 (1.070276)
Big 5 Agreeableness	-	.5343226 (1.131673)	-	-	-	-0.4681442 (1.324114)
Big 5 Conscientiousness	-	.2806386 (1.117035)	-	-	-	-0.8968197 (1.349083)
Big 5 Neuroticism	-	-.8743524 (0.9337394)	-	-	-	-2.843971 (1.164247)
Big 5 Intellect/Imagination	-	0.3619613 (1.066678)	-	-	-	1.619874 (1.238055)
EPO Generation & Evaluation	-	-	1.272142 (1.061026)	-	-	-0.0765192 (1.361714)
EPO Positive Outcome Focus	-	-	0.3743729 (1.035722)	-	-	-0.5287629 (1.17086)
EPO Negative Outcome Focus	-	-	2.587809* (1.136466)	-	-	2.888265* (1.341246)
BIS (overall)	-	-	-	-2.558179 (1.970952)	-	-3.818222 (2.885316)
GDMS Rational	-	-	-	-	2.119413 (1.522602)	0.0548745 (1.911569)
GDMS Intuitive	-	-	-	-	0.9416385 (1.469622)	1.329873 (1.559568)
GDMS Dependent	-	-	-	-	2.842393* (1.291512)	2.695333 (1.441875)
GDMS Avoidant	-	-	-	-	1.457123 (1.137685)	1.580458 (1.369016)
GDMS Spontaneous	-	-	-	-	-1.563602 (1.335612)	-1.250468 (1.538751)
Constant	0.7941177*** (0.2092666)	0.6975441 (1.421964)	-1.708326 (1.205288)	2.478716* (1.105184)	-3.083344 (1.654658)	-0.5244026 (2.820571)
Pseudo R-squared	0.0104	0.0444	0.0890	0.0251	0.0855	0.1689
N (observations)	306	306	306	306	306	306

*** Significant at 0.001 level; ** Significant at 0.01 level; * Significant at 0.05 level

Table 18 Results of the Multinomial Logit Regressions for Hybrid Treatment (dependent variable – sample; base category – Theory planners/sophisticated students) – A: Professional planners (practice planners) vs Sophisticated Students (Theory planners)

Explanatory Variable	Model 1 (NI)	Model 2 (BIG5)	Model 3 (EPO)	Model 4 (BIS)	Model 5 (GDMS)	Model 6 (ALL)
Naïve interventionism (NI)	0.2668453 (0.6671818)	-	-	-	-	0.442446 (0.732484)
Big 5 Extroversion		1.205688 (1.085533)	-	-	-	1.523608 (1.167813)
Big 5 Agreeableness	-	1.88103 (1.317286)	-	-	-	0.7056266 (1.478365)
Big 5 Conscientiousness	-	3.295388 (1.319714)	-	-	-	1.145372 (1.577957)
Big 5 Neuroticism	-	-2.803944** (1.075182)	-	-	-	-2.489834 (1.310105)
Big 5 Intellect/Imagination	-	-.9428772 (1.235076)	-	-	-	-0.5560833 (1.389789)
EPO Generation & Evaluation	-	-	4.484905*** (1.206153)	-	-	2.200433 (1.540314)
EPO Positive Outcome Focus	-	-	-2.875194* (1.22596)	-	-	-3.408303* (1.349688)
EPO Negative Outcome Focus	-	-	-3.697672** (1.291566)	-	-	-3.006533 (1.555804)
BIS (overall)	-	-	-	-8.00411*** (2.283878)	-	-3.197738 (3.294688)
GDMS Rational	-	-	-	-	4.538446** (1.761031)	3.33564 (2.307201)
GDMS Intuitive	-	-	-	-	-1.672094 (1.614827)	-1.812319 (1.74239)
GDMS Dependent	-	-	-	-	2.713656 (1.435263)	2.445807 (1.612073)
GDMS Avoidant	-	-	-	-	-2.865059* (1.249215)	-0.6358813 (1.493963)
GDMS Spontaneous	-	-	-	-	-0.6491749 (1.5312)	0.7159121 (1.762815)
Constant	0.3432037 (0.2314037)	-1.857212 (1.659884)	1.1224 (1.4084)	4.683012*** (1.243537)	-1.786024 (1.907126)	0.8802183 (3.343933)
Pseudo R-squared	0.0104	0.0444	0.0890	0.0251	0.0855	0.1689
N (observations)	306	306	306	306	306	306
*** Significant at 0.001 level; ** Significant at 0.01 level; * Significant at 0.05 level						

Note: Since both effects for Naïve and Practice planners are evaluated jointly, Table 17 and Table 18 share the same pseudo R^2 and N (observations).

The multinomial logit analysis (refer to Table 17 and Table 18) reveals that Naïve planners are different from Theory planners in their EPO negative outcome focus, while Practice planners are different from Theory planners in their EPO positive outcome focus. However, since neither EPO negative outcome focus nor EPO positive outcome focus have a significant effect on performance (see the OLS regression analysis), these differences do not shed light on why Theory planners perform better in Hybrid treatment than other samples. Therefore, the only individual characteristic that seems to matter for performance is exposure to planning as well as the level of naïve interventionism. Therefore, at this stage Hypothesis 3 is only partially confirmed, with Hypothesis 3-A being fully confirmed.

Finally, hypothesis 3-B and 3-C are tested using a clustered regression (Table 19) on each decision made through the task. It is clustered by participant and each participant made 30 decisions. A clustered regression allows a more detailed insight into individual heterogeneity. The analysis for the most interesting two groups is listed in Table 19. For both groups, naïve interventionism, as previously observed, contributes negatively to DP performance. The negative contribution of naïve interventionism is highly significant (at 0.001 level). The same was observed in the previous simple OLS regression. Therefore, Hypothesis 3-B is supported but 3-C is rejected due to inconsistent significance.

For sophisticated students, all five constructs of the big five contribute significantly to explain variation in performance. Extroversion and agreeableness contribute positively while conscientiousness, neuroticism and imagination contribute negatively. In terms of elaboration on potential outcomes, generation and evaluation as well as focus on positive outcomes both contribute negatively to performance. Impulsiveness (BIS) also contributes negatively to performance that is aligned with prior expectations. Surprisingly, results differ for sophisticated students and sophisticated non-students in terms of GDMS. Preference for rational, avoidant and spontaneous decision-making styles all contribute negatively to performance while the preference towards intuition seems to benefit planning performance. This is different for professionals, with preference for intuitive decision-making style contributing negatively to performance. Although professionals do not differ in terms of how spontaneous decision-making style contributes to performance, its magnitude is significantly different.

Finally, for professional planners, feedback about volume and profit both contribute significantly to predict performance. Feedback on volume contributes negatively, which can be interpreted as positive variations in demand volume lead to worse performance that is aligned with the assumption of loss aversion. In contrast, feedback on profit contributes positively to performance, suggesting that professionals make use of this information.

Table 19 Clustered OLS regression for sophisticated students and sophisticated non-students

Explanatory Variable	Sophisticated Students Model Coefficient (standard error)	Sophisticated non-students Model Coefficient (standard error)
Naïve interventionism	-5643.8820*** (1076.0940)	-4031.9210*** (1102.6180)
Big 5 Extroversion	3430.9450*** (800.7175)	2843.81 (4462.412)
Big 5 Agreeableness	16856.6900*** (2167.5470)	-456.5991 (7213.092)
Big 5 Conscientiousness	-7725.0220*** (1328.6680)	-3850.434 (4994,299)
Big 5 Neuroticism	-12149.7300*** (1805.2410)	6728. 987 (4234.167)
Big 5 Intellect/Imagination	-8782.9610*** (1380.6520)	3445.436 (4496.55)
EPO Generation & Evaluation	-6070.5770*** (1032.9670)	3063.979 (3291.402)
EPO Positive Outcome Focus	-8123.0830*** (1052.8700)	11738.7900** (3926.3610)
EPO Negative Outcome Focus	-2146.018 (1301.048)	70.2044 (5147.532)
BIS (overall)	-15590.2200*** (3637.5710)	21018 (11544.23)
GDMS Rational	-16027.9800*** (2777.3440)	-2515.68 (10861.3)
GDMS Intuitive	14148.4600*** (1910.8740)	-11408.7700* (4251.0100)
GDMS Dependent	932.5941 (2607.379)	10027.71 (6732.635)
GDMS Avoidant	-12261.0700*** (1559.7840)	1954.158 (6656.998)
GDMS Spontaneous	-4115.9740** (1498.9990)	-10700.4000* (5139.9440)
Feedback on volume	-0.1657 (0.106386)	-2.3191*** (0.6529)
Feedback on profit	0.0169 (0.0133)	0.3321*** (0.0896)
Absolute plan jump	-0.3787 (0.3240)	0.0787 (0.5931)
Age	-79.8843 (195.5656)	-306.2873 (631.7543)
Sex	476.1240 (302.2109)	-615.8712 (1449.99)
Level in organisation	n.a.	683.3845 (728.2778)
Years of experience	n.a.	118.1885 (123.2514)
Prob > F	0.0000	0.0000
R-squared	0.9520	0.8625
N (observations)	696	812
*** Significant at 0.001 level; ** Significant at 0.01 level; * Significant at 0.05 level		

Regarding the last Hypothesis 3-D, the performance between male and female groups is compared across sample (Table 20). None of differences in performance between men and women is significant.

Table 20 Gender results by sample (profit and N subjects in brackets)

Sex at birth	Naïve students	Sophisticated students	Professional planners
Male	25458.13 (72)	27359.08 (26)	25829.78 (55)
Female	24774.92 (91)	27203.9 (29)	25794.89 (27)
Prefer not to state	26831 (3)	22893 (1)	25003.5 (2)
MWW test male vs female	z = 1.637 p = 0.1015	z = 0.152 p = 0.8794	z = 0.059 p = 0.9528

The same analysis is performed across treatments (Table 21). None of differences in performance between men and women is significant.

Table 21 Gender results by treatment (profit and N subjects in brackets)

Sex at birth	Short	Long	Hybrid
Male	25143.26 (46)	27543.56 (48)	25191.15 (59)
Female	24836 (45)	26671.74 (50)	24782.42 (52)
Prefer not to state	22893 (1)	27126 (4)	21996 (1)
MWW test male vs female	z = 0.091 p = 0.9273	z = 1.396 p = 0.1626	z = 1.087 p = 0.2769

Given the non-significant results between men and women both across groups (Table 20) and across treatments (Table 21) Hypothesis 3-D is rejected.

5.7 Chapter summary

The analysis focuses on three main groups, naïve students, sophisticated students (majoring in L&SCM) and sophisticated non-students (professional planners). Variations in performance between long and short commitment treatments confirms Hypothesis 1: Planning performance in the planning *option Long will be better than that in the planning option Short (i.e., planners will achieve higher profit in Long compared to Short).*

Considering experience, professional planners outperform naïve students numerically, but this difference is not statistically significant. In contrast, sophisticated students significantly outperform the remaining two groups. One possible interpretation is that L&SCM students had exposure to theory and understand the correct strategy for the newsvendor type of game. The results partially confirm Hypothesis 2 that says that factors other than treatment variation is likely to influence performance in the Hybrid treatment. Yet, this result also suggests that, across all samples, participants react to policy: restricting the planning period to Long evaluation (binding policy) helps to improve planning performance and earn higher profit/minimise losses.

To understand the differences in performance, the focus is on the hybrid treatment. Considering the contribution of individual differences to explain differences in performance, results only support part of Hypothesis 3: *Individual differences and individual personality traits are a significant predictor of demand planning performance. The effect of individual differences and individual personality traits should be particularly strong in the Hybrid planning option where individuals have a choice between following and not following the decision guidance to stick to the long-term plan*

Results fully confirm Hypothesis 3-A: *Individual differences with regard to exposure to planning is an important determinant of planning performance with Theory planners performing better than Practice and Naïve planners.*

Moreover, results also confirm Hypothesis 3-B: *Greater level of naïve interventionism leads to worse demand planning performance* – Worth noting that this was accepted because NI appears as significant in all variants of the analysis except in one sub-group regression including all items. The exception is for professional planners in an all-inclusive model. For the rest, NI appears as significant or highly significant always with the same negative contribution to performance. The remaining hypotheses 3-C and 3-D are rejected. No consistently significant contribution has been found from psychometric and demographic variables.

The following chapter will focus on conclusions, discussing the results, findings, limitations and their implications to theory and practice.

6 Conclusions

6.1 Introduction

This is the final chapter of the thesis. Following this introduction, a summary of the research is provided. Findings are presented focusing on how the identified research gap is being addressed, followed by the contribution to theory and to practice. Limitations and further research are outlined. This thesis ends with a final personal reflection. The structure of Chapter 6 is summarised in Table 22.

Table 22 Chapter 7 structure

Section	Overview
6.2 Summary of the research	A summary of the research is provided, revising the initial purpose, problem and research question addressed.
6.3 Findings	The section discusses the addressed research gap along with the findings.
6.4 Contribution	Findings are discussed in terms of contribution to theory and to practice.
6.5 Limitations	The section discusses the shortcomings of the chosen approach emphasising compromises and limitations.
6.6 Further research	Further research links back to some of the limitations discussed in the previous section. Recommendations for further work are provided. Most of further effort is an expansion and suggestions for alternative treatments in the experiment.
6.7 Final reflection	The final section is a personal view on the research, specifically the motivation and the individual learning.

6.2 Summary of the research

Balancing supply and demand is critical for increasingly complex business systems of exchange of goods and services (e.g., Deming 1986; Christopher & Lee 2001). These systems are managed by people. Managers, just as any human beings, are complex systems themselves. It is generally assumed that everyone involved in management wants to perform well and make good decisions. However, practice shows that often best efforts can be destructive (e.g., Fildes et al. 2009). One of the most common form of destructive efforts are unnecessary interventions. Deming (1986) proposes the idea that businesses cannot succeed without the following four core concepts:

knowledge (epistemology), understanding variation, appreciation of systems and, people (psychology).

Deming's (1986) idea can be simplified down to the duality of humans and systems as well as how business success depends on the consideration of both. DP is one of the most important processes in business. It balances supply and demand to deal with uncertainty. DP is one of the managerial processes that greatly depends on both humans and systems. DSS can offer system restrictiveness or decisional guidance. The predominant trend in DSS has been towards deliberate decisional guidance. Managers are often presented with options and guiding principles suggesting the best action. The general assumption behind decisional guidance is that managers are able to use the available information in the most rational way.

One of the main issues in DP are unnecessary interventions and constant overriding of the system. Managers are often over-confident about their ability to improve statistical forecasts and make adjustments. The separation between what is best achieved via statistics and what is best suited for human judgement is often confused in practice (Fildes et al. 2006). Therefore, DP performance greatly depends on both system factors and human factors. How well managers perform will not only depend on the task, but also on their individual differences, e.g., personality (e.g., Lapide 2007). For example, one of the main approaches to improve manager's performance is via training.

When it comes to judgement, previous research sometimes fails to detect benefits of experience (e.g., improvement over time) or evidence of skill (e.g., consistently good performance). For example, companies do not become better at forecasting over time (e.g., Rieg 2010). Similarly, in experimental conditions, professionals perform as well as naïve subjects (e.g., Haigh & List 2005; Bolton et al. 2012). However, this should not be surprising since many of the decision making biases are known to be resistant to practice or training (e.g., Tversky & Kahneman 1974). Similarly, some environments simply do not allow good judgement (Todd & Gigerenzer 2007).

This thesis had the following research question: *What is the contribution of individual differences and planning policy parameters to demand planning performance?*

This research is one of the few studies focusing on DP because of system factors and individual differences. The problem has been partially observed by either DP system (e.g., OR literature), or individual differences (e.g., cognitive psychology literature). The aim was to develop a theoretical model drawing on theory from behavioural economics and psychology to identify planning policy parameters and individual traits that can be used to predict DP performance.

Relying on CPT and the mental accounting bias, MLA is used to build a prediction around how the length of commitment will influence DP performance. To capture individual differences trait theory is used together with previously developed and validated psychometric scales. As a result, this thesis proposes a theoretical framework. The methodology of decision making experiments and econometric analysis methods are used to test three main hypotheses. The obtained sample represented both naïve subjects, sophisticated students and professional planners.

6.3 Findings

Following the results and analysis, Figure 25 summarises the hypothesis and respective results. The application of CPT to planning behaviour provides results consistent with the theoretical expectation. MLA is successfully detected in planning (H1 and H2). This can have significant implications for both theory and practice. Comparing treatments, MLA significantly affects performance. Longer commitment led to better performance while more frequent decisions deteriorated it. To understand the contribution of individual differences a hybrid treatment is used. This treatment relies on a non-binding policy that was enforced via decisional guidance.

Considering the differences between treatments, an important finding concerns DSS, specifically, if planning decisions should be guided or restricted (or none). Informative guidance as a form of deliberate decisional guidance has shown to have different results depending on the task. While it has been generally ineffective with tasks related to forecasting, it has shown to improve other tasks (Montazemi et al. 1996). Deliberate decisional guidance has been particularly effective when used as memory support (Singh 1998). The comparison between the treatment without decision support and the other two treatments shows that having decision support significantly improves task performance. Results suggest that when planners have no exposure to theory, system restrictiveness will lead to better results. Alternatively, when participants have good understanding of theory, decisional guidance can also lead to good results. In general, system restrictiveness leads to better results when compared to decisional guidance because it dampens variation.

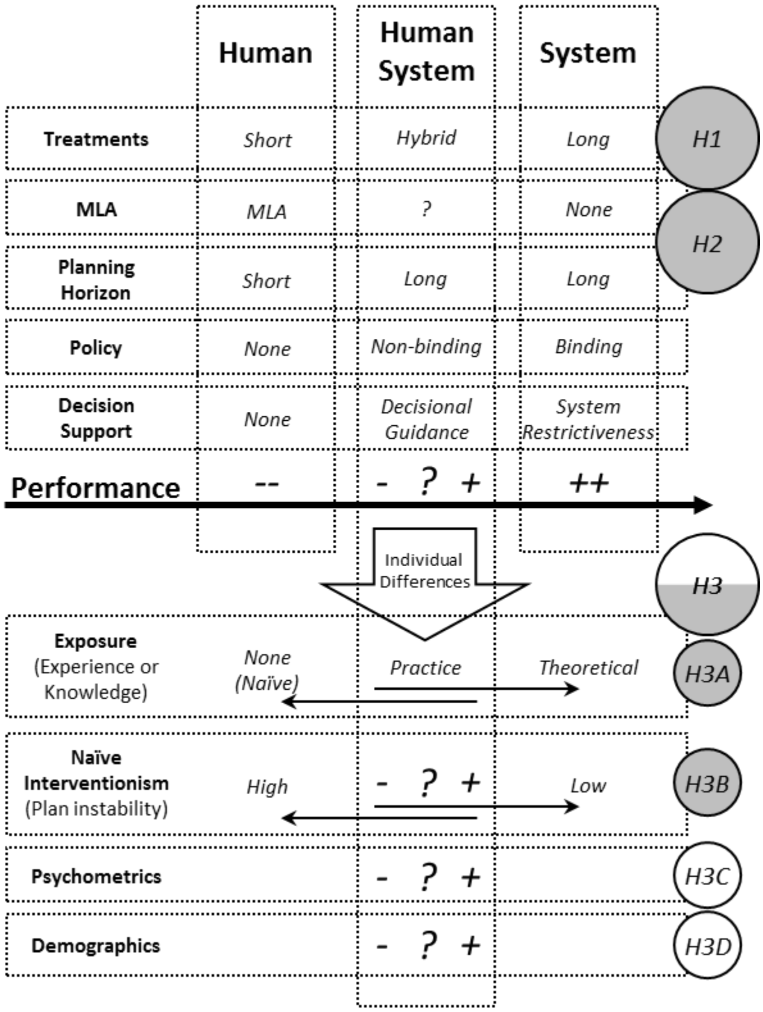


Figure 25 Conceptual framework, hypothesis and results (grey = confirmed)

Although individual traits and personality were expected to play a significant role in DP performance, hypothesis 3 is only supported partially.

Exposure (H3A) significantly affects performance. A new model describing how exposure relates to DP performance is designed, the EPM. Theoretical knowledge allows the optimal solution while naive knowledge results in the worse performance. Exposure to practice is in between. Results successfully support a prediction that is aligned with previous research. Professional planners did not perform significantly better than naïve subjects (only numerically better). However, postgraduate L&SCM students with no practical planning experience significantly outperformed the remaining groups. L&SCM students are exposed to theory, potentially remembering what the best strategy for the newsvendor problem is.

Naïve interventionism contributes negatively to DP performance as expected (H3B). Higher NI leads to lower profit. This means that frequent adjustments to the over the previously planned volume contribute negatively to performance (profit).

Finally, considering personality traits (Big 5), psychometric scales (BIS, EPO, and GDMS) and demographics (sex, age, years of experience, and level in the organisation). All show inconsistent significance. Demographics such as gender, age or years of experience are not significantly associated with DP performance. This contradicts some of the previous research. Specifically, claims that there are differences between genders (e.g., De Véricourt et al. 2013) are not supported. Similarly, claims that impulsiveness is associated with underperformance (e.g., Martin & Potts 2009; Ockenfels & Selten 2015) are also not supported. Overall, this thesis does not support previous research which claims that there is a 'right mind-set' for planning (e.g., Lapide 2007). Therefore, this thesis challenges the relevance of psychometric tests often used by Human Resources departments (e.g., MBTI® test) as individual development tool. Instead, DP policy can lead most individuals to perform better regardless of his/her individual characteristics. In this regard, this thesis supports E. Williams Deming's (1986) idea that a bad system will beat a good person every time.

6.4 Contribution

This is a cross-disciplinary thesis that observes a problem from O&SCM. It borrows theory from the fields of economics and psychology. This thesis contributes to the growing body of literature of the newly formed field of behavioural operations research (BOR). However, due to the immature state of BOR, further discussion is made by each field separately. Hence, this thesis contributes on different levels to fields of engineering, management, psychology and economics.

6.4.1 Contribution to theory

The main contributions of this thesis are to Engineering and Management while psychology and economics are secondary. This research seeks an explanation to a phenomena observed in DP using theory from psychology and economics. This research contributes to the theoretical understanding of what are the contributions of individual differences and planning policy parameters to DP performance. At the same time, fields of psychology and economics benefit from an innovative experimental design and application of its theory in a new context. Figure 26 provides an overview of the main contributions of this thesis.

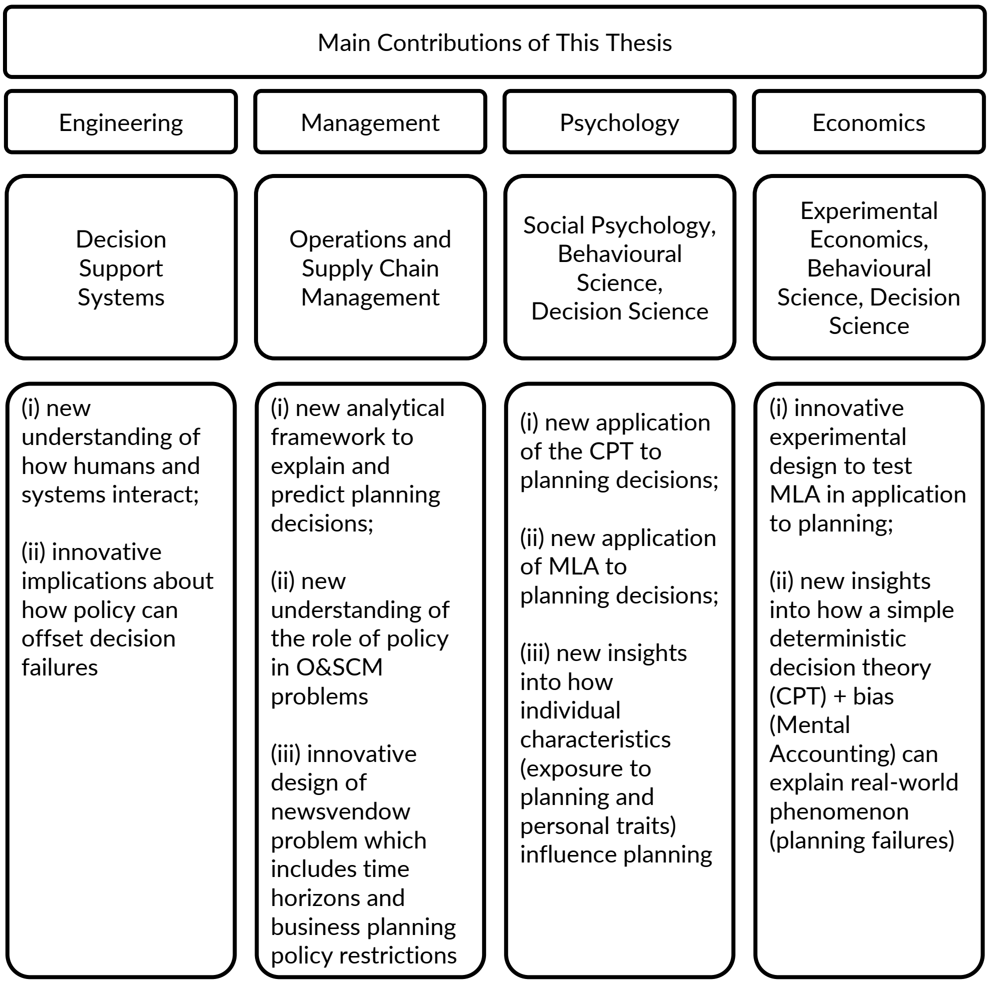


Figure 26 Summary of the main contributions of this thesis to theory

This research contributes to engineering and DSS with a new understanding of how humans and systems interact using decisional guidance or system restrictiveness. This can be attributed to theory in behavioural OR in three ways. This thesis provides an explanation of one of the tasks of DP to balance supply and demand. This is one of the most common types of decisions in OM. This is achieved by borrowing from the field of Behavioural Economics the CPT and using MLA to explain one of the persistent behavioural issues observed in practice. So far this has never been done in OM or OR. This constitutes the main contribution of this research. It is not new to detect that frequency of interventions degrade planning performance. Similarly, unnecessary interventions and adjustments to statistical forecasts are known to be negative. Whilst most of the existing research focuses on how to improve on the problem of integrating judgement with statistics, explanations to why managers behave in such way are limited. This thesis provides an explanation of the human behaviour in DP through CPT using MLA.

Contributing to O&SCM, a new analytical framework is proposed explaining and predicting planning decisions. It allows new understanding of the role of policy in O&SCM. Additionally, an innovative design of the newsvendor problem is proposed. It includes manipulation of the time horizon and planning policy that has also never been done so far.

Previous studies of this nature have observed variation in performance between individuals but no explanation of potential factors have been provided. Despite being common good practice in experimental research to collect measures on individual attributes (e.g., demographics) results are usually reported as average and little to no explanation is provided on the heterogeneity of the results. Drawing on the Trait Theory from the field of psychology previous studies provide evidence that individual differences, e.g., experience, personality, can significantly impact decision making performance (e.g., Weinman et al. 1985; Stanovich & West 2000). This study tested the trait theory hypothesis through multiple previously validated psychometric scales, failing to detect any stable profile of what characteristics make up for a good planner.

One concerning fact with this study and its results comparing experts/professionals with naïve subjects, similar to previous research (e.g., Arnold et al. 2000; Bolton & Katok 2004; Haigh & List 2005; Bolton et al. 2012), is that it also fails to successfully detect skill. Once again it is successfully demonstrated how practical experience is unrelated to performance – Expert's (professional planners) judgement was not significantly better than the judgement of naïve planners. This idea is not new and has been found in many studies before this one. Camerer and Johnson (1991, p.203) goes as far as claiming that *'[the] depressing conclusion from these studies is that expert judgements (…) are no more accurate than those of lightly trained novices (…) expert judgement have been worse than those of simplest statistical models in virtually all domains that have been studied'*. This thesis supports this view. The intuitive expectation would be that experienced individuals would outperform on average naïve subjects. However, skill is not detected at an overall level, neither at treatment level. Whilst the experimental task is arguably not fully representative of the tasks professionals do in real life, the newsvendor situation represent one of the most basic decision making processes. The demand planner must make a decision reacting to a demand signal under uncertainty and risk.

As secondary contributions, this thesis contributes to social psychology, behavioural economics and decision science with a new application of CPT to planning decisions. MLA is successfully applied to planning decisions. This is the first application of MLA to planning. New insights are provided on how individual differences influence planning decisions. This thesis contributes to methodology by its use of validated psychometric scales together with a multi-treatment modified newsvendor problem that is designed to represent a real situation from management. It is designed to detect both individual differences as well as overall decision-making bias. The experimental task can be considered representative of a real-life situation as the participant is presented with both a system's recommendation, unknown incoming demand and graphical and tabular feedback. The performance is measured in terms of overall profit that is different from measuring accuracy in experiments on judgemental forecasting.

Priming participants with the optimal solution has not been done before. This priming was intended to represent the statistical forecast offered by DSS systems to DP planners. All treatments included priming so no conclusions are made on whether it improves performance or not. However, even provided with the optimal solution, the majority of participants with the exception of very few chose to deviate from the recommendation and react to variations in the demand.

Finally, contributing to the field of experimental economics, behavioural science and decision science, this thesis proposes an innovative experimental design to test MLA in application to planning. It allows drawing insights on how a simple deterministic theory (CPT) with the mental accounting bias (as part of the MLA) can explain a real-world phenomenon of sub-optimal DP in O&SCM.

6.4.2 Contribution to practice

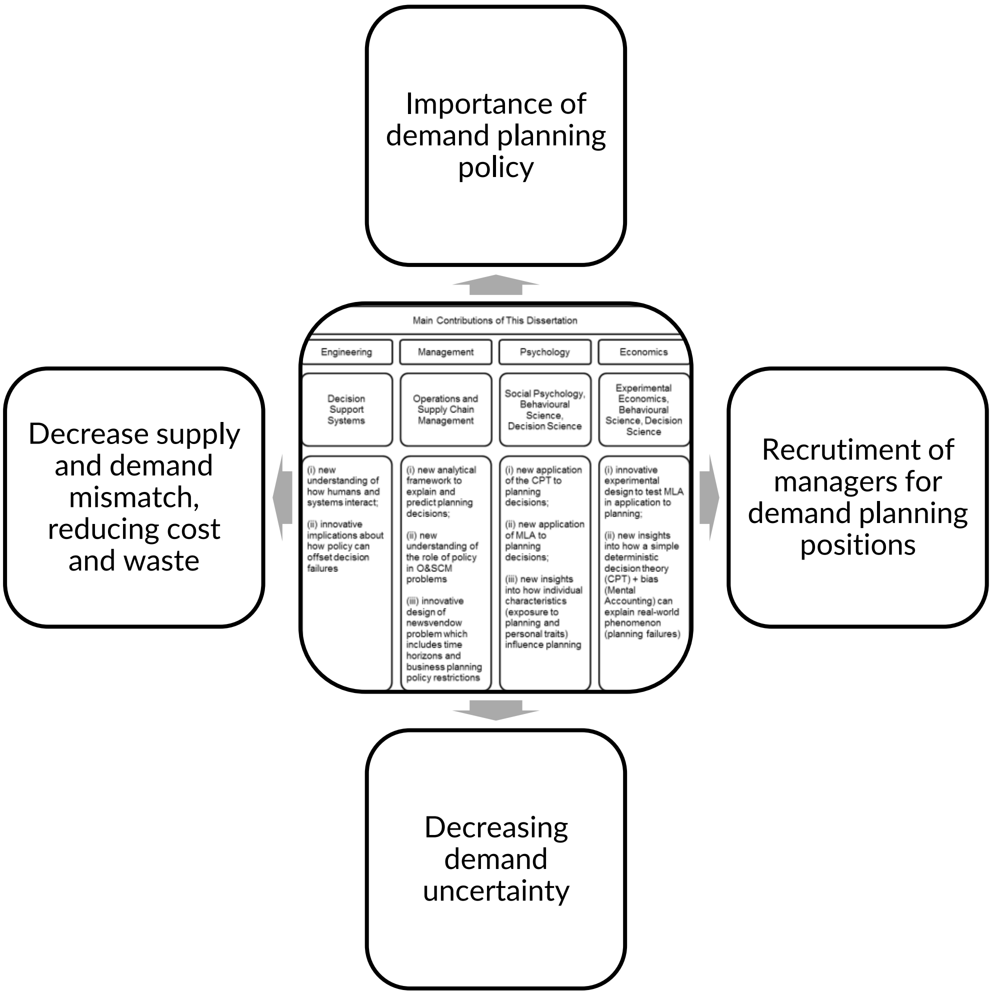


Figure 27 Summary of the main contributions of this thesis to practice

Findings from this research have implications for DP policy design, recruitment of manager for DP positions, decreasing supply and demand mismatch and reducing induced demand uncertainty.

DP policy is shown to be important to manipulate DP performance. Considering the problem of operationalisation, DSS is suggested as means to enforce different DP policies. One major implication for practice comes from the discussion between the use of decision support and elaborating on the idea of whether the decision-maker should be guided or restricted. Systems should be designed in such a way that people are not exposed to situations where judgement failure is likely. In situations of repeatable decision-making, judgement should be automated as much as possible. If automation is not possible, the recommendation is to resort to system restrictiveness to avoid unnecessary or damaging interventions. However, it is important to consider the warning by Fildes et al., (2006, p.355) that *'absolute restrictiveness can be dangerous if it is wrongly applied'*. Therefore, it is essential to take into consideration when judgement should be allowed and where it should be restricted. The suggestion following this study is to moderate the intervention frequency.

Another important implication for practice is that frequent interventions or adjustments are indeed harmful in most cases. The results suggest that planners, like most human beings, suffer from mental accounting and are prone to MLA. Whilst this is a theoretical contribution, in practice this means that the problem of DP should be addressed via reducing the frequency of interventions and increasing time horizons over which judgements are made. Too frequent analysis and interventions can lead to overall underperformance. Even if it can seem like a local improvement, considering MLA, it will be most likely worse overall. Informed by MLA and the negative effect of NI, it is possible to reduce demand amplification and information distortion using DP policy. Hence, enforcing MLA-informed policies will reduce supply and demand mismatch as well as reduce demand uncertainty.

Considering individual differences, this thesis challenges the commonly accepted idea that there is a 'right' personality to be a good planner, as defended by Lapide (2007). Consequently, this affects the recruitment and training for the role of DP. There is a general (intuitive) belief that individual differences and training are very important for the role. However, this thesis claims that good policy is more important than personality or individual differences. Demand planners should not be selected based on psychometric tests.

This thesis was originally set to understand what the contribution of individual differences to DP performance is. As a result, it would provide recommendations on what aspects of one's personality make up for a good demand planner. However, no significant contribution of personality traits nor psychometrics to DP performance was found. This allows both an optimistic and a pessimistic interpretation. An optimistic interpretation suggests that planning performance does not depend on individual's personality. This does not mean, however, that in a situation of additional information (i.e., being aware of a promotion that the marketing planned in advance), a manager will not outperform an 'unaware' one. In situations of equal information, different individuals might perform as well if a system is designed considering the shortcomings of the human mind. A pessimistic interpretation of the little to no importance of individual characteristics is that people are simply unfit for the task of planning. This experiment fails to detect evidence of skill or experience as significant predictors in the DP task. If this is the case, then management should focus efforts on both deploying good planning systems and allowing greater automation by decision support. Planning policies must prevent people from intervening with the system and making unnecessary changes. This thesis advocates for the adoption of greater system restrictiveness over investing in decisional guidance. This is reinforced by the fact that participants, even provided with the optimal solution to the problem, quickly ignored it trying to 'game' the system.

Finally, findings about naïve interventionism have implications for practice. Some individuals exhibited greater levels than others did. This variation had significant implications for performance. For example, students with theoretical knowledge performed less frequent but relatively large changes comparatively to professionals and naïve students. This can have implications on what kind of restrictions are imposed on planners, e.g., similarly to results by Fildes et al. (2009), by restricting small adjustments and allowing only relatively large ones.

6.5 Limitations

Individual traits are complex to measure and the selected psychometric scales as well as personality inventory are not free of criticism. Most of the criticism is partly linked to the positivist paradigm that is often criticised as reductionist. This research inherits some of the common criticisms in the field of psychology on the study of individual differences and personality. Relying on simplified measures to describe one's personality provides a relatively limited insight. However, it is unpractical to try to measure exhaustively all traits that define one unique self. Hence, regardless of the care to select relevant psychometric scales, only a small fraction about individual differences is actually measured and analysed.

The use of newsvendor as decision-making task is not universal and does not represent all DP decisions in practice. One of the main assumptions of the newsvendor is that each decision is made one-shot for a single period. It does not influence the follow up periods. Excess is savaged and shortage is lost. In real-life, DP decisions are much more complex. However, the purpose of this experiment was to detect a fundamental decision making bias. Therefore, the decision making process had to be greatly reduced to its most basic formulation.

Time is not considered. The timing of the decisions is ignored and the focus is only on its sequence. There is no control for the time taken to make the decisions. Moritz et al. (2014) observes the decision speed in a judgemental forecasting task, finding that too quick or too long decisions tend to increase error. How quick and how long is vaguely defined. For the purpose of this thesis, participants are allowed as long as necessary to make the decisions without any time pressure assuming that each would choose the optimum decision time.

The experimental task can also be criticised because it uses only one demand series of 30 responses (observations). In order to minimise the possible effects of single series, a randomly generated sequence drawn from uniform distribution was used for the study. Since the purpose of this experiment was to study behaviour of different types of individuals across different treatments, it was difficult to create multiple demand scenarios because the experiment either would then require a much larger subject pool or would take significantly longer time. Another reason for limiting the task to only one sequence was the fact that professional respondents (planners) could only be engaged in the study if a short task could be designed. A different random order was not generated for each participant because this would create a difficulty in analysing the data: this would not allow for direct comparison of decisions, as they would be likely to be influenced by order effects. A further exploration of other sequences within newsvendor problem is necessary to confirm the robustness of results reported in the earlier sections.

Another limitation associated with the experimental task is the use of a single high profit product. In practice, managers often manage several products or portfolios with distinct profit margins. Observing a single high margin product allows isolating the parameters but is arguably only partially representative of reality.

One of the limitations of the research design can be attributed to the effort that the subject puts in the task. To address this limitation, an incentivised experiment is run with part of the students. Comparing the performance between incentivised and non-incentivised students, no significant difference is found between providing incentives or not. Considering the common criticism of students not being able to make as good use of the provided information as a professional would, experimental results in this thesis show the opposite. Instead, naïve students and professionals do not differ in performance and sophisticated students significantly outperform the previous two. These results challenge the criticism of effort and expertise. However, in an ideal setting it would be better to provide meaningful monetary incentives to all groups.

Another criticism could be raised with regard to the study's subject pool. Specifically, the proportion of undergraduate students and postgraduate students was unequal in different experimental groups. Specifically, Naïve planners were primarily undergraduates and Theory planners were primarily postgraduates. This was mainly because it was impossible to find an undergraduate program at the University of Warwick where students would be sufficiently knowledgeable about the planning theory. This criticism could be easily addressed because (i) the overwhelming majority of undergraduates (96%) were in their last year of study and all postgraduates (100%) were in the first (and only) year of study. Therefore, the two groups were quite close in terms of their level of education with the only difference that postgraduates had exposure to Supply Chain Management theory in their program.

The response rates in the study were 75% for students and 35% for professionals. It could be argued that because the response rates were not 100% a selection bias could have affected experimental results. However, this is highly unlikely for the following reasons. Students were drawn from a homogeneous population of the subject pool of the Decision Research at Warwick group. This homogeneity was insured by the administrators of the recruitment system. Furthermore, 25% dropout among students could be explained by the fact that they were given several days to complete the study online. Yet, they were also informed that once the study reaches maximum number of participants (110), the study will close automatically (places in the study were available on first-come-first-serve basis). The number 110 was determined based on the available budget. Therefore, 25% of students who started and did not finish the study were (most likely) those who underestimated the competition from other potential participants or those who were not very interested in the study. This suggests that 75% of students who completed the study were representative of the student population at the University of Warwick.

The response rate of 35% among professionals is unprecedentedly high since the majority of marketing studies commissioned commercially only reach out to 15% of professionals. In fact, this study recruited the largest sample of planning professionals studied in the literature to date. Even if one assumes that only those professionals who were particularly interested in planning completed the study, this makes the study results even stronger. The study finds that professionals in many treatments perform worse than students do and if only ‘sophisticated’ planners were attracted to the study, they should be expected to do better instead of worse.

One may also question the study results because it could be argued that the respondents intervened in the experiment (they were being asked to play an active role) rather than generate reported effects due to loss aversion. Indeed, in the Short treatment the respondents are asked to act in every time period, while in the Long treatment the respondents are asked to defer actions. Yet, in the Hybrid treatment, participants have a choice between action and not acting. In the Hybrid treatment participants are asked not to act too often (they receive Decision Support/Policy), but, as results suggest, many of them do act more often than advised. Therefore, it is unlikely that reported results are generated by something other than a personal trait and the experimental design of this study is careful with separating the effect of personality characteristics from the impact of loss aversion that appears to explain the observed behaviour.

This thesis did not detect significant contribution of personality over DP performance. This can be due to the adaptive nature of individual differences (Buss & Greiling 1999; Buss & Hawley 2010). It claims that 'humans possess a complex array of evolved psychological mechanisms, only a subset of which is activated at any particular time' (Buss 1999, p.259). This means that some personality traits essential to DP might only activate in special circumstances that have not been present in the simulation. Alternatively, any suggestion about what traits make up for a good planner are subject to this same idea. Up until a particular moment (i.e., need to make DP decisions) the relevant individual differences might not be active. Hence, attempts at measuring particular traits beforehand can be inconclusive or even misleading. To counter this limitation, subject's personality traits are measured following the simulation task representative of the DP task.

Finally, a criticism might be raised with regard to the regression models used in the study. Specifically, it could be argued that some of the personality scales' variables do not come out significant because they could be correlated with Naïve Interventionism variable (i.e., the clustered regressions suffer from overfitting bias). Yet, it is highly unlikely that the presented analysis generated biased results. According to the correlation analysis (reported in Appendix N), Naïve Interventionism is significantly correlated with only three other personality measures: Extraversion (negative correlation), Neuroticism (positive correlation) and Negative Outcome Focus (positive correlation).

Furthermore, several robustness checks were reported in the regression summaries where Naïve Interventionism is taken out of the regression while other (correlated) characteristics were present as independent variables (see tables 14-18). The fact that Naïve Interventionism in many tables remains significant across models while other variables are not significant even when taken in isolation indicates that analysis is unlikely to produce untrustworthy results.

Table 16 reports results for N=28 professionals and shows that none of the variables are significant when taken jointly. One may argue that this is due to the small sample size. However, the same table also reports that Naïve Interventionism is significant when taken in isolation as well as two GDMS traits are significant when taken separately. These GDMS traits are not correlated with Naïve Interventionism according to the correlation analysis (shown in Appendix N). It is, of course, possible that estimating a model with 16 variables in the extended model does not return significant results because the sample is rather small. Yet, since variables that produce correlations in the narrowed-down models are not correlated with each other, statistically, we should expect at least one of these variables to come out significant in the extended model.

6.6 Further research

Much of the further research comes from the limitations of this study. Due to time limitations, alternative experimental treatments are left untested. The present research focused on a particular situation in DP, it was a repeated individual task based on a modified newsvendor problem set with a single high-margin product.

Further research should focus on a replication of the experiment using a small margin product. It can potentially lead to very different results (behaviour differed between high and low margin in De Véricourt et al., 2013a). Individual decision making and group decision-making is different and whilst much of behavioural evidence suggests that the bias persists across groups, both personality as well as group dynamics can potentially be predictors of performance. Mental accounting bias suggests that individuals will react differently not only variations in frequency, but also to portfolios as opposed to single elements. Managing multiple products with different levels of aggregation is also expected to be prone to MLA. Literature discusses the advantage of planning at a product family level for long horizon and at stock keeping unit at a short horizon. However, this discussion is mostly explanatory and provides the recommendation as a rule of thumb. Applying CPT and MLA to product aggregation in the context of DP could potentially provide theoretical understanding of why this happens.

Further research should also consider testing variations using a dynamically complex decision making task. A dynamically complex task according to Brehmer (1992, p.212) requires that the decisions must be made as a series rather than single so many decisions are required to reach a given goal. Moreover, the decisions must be constrained by previous decisions so they are not independent. Additionally, the states of the system change both with the context as well as with the decisions made and finally, the decisions must be made in real-time. A dynamically complex decision making task and individual differences has already been tested (Strohhecker & Größler 2013), however, the objective was solely focused on explaining individual differences.

The newsvendor problem assumes that each decision is made one-shot for a single period not influencing the follow up periods where excess is savaged and shortage is lost. A dynamically complex decision making task comes closer to real-life decisions in terms of being compatible with situations of carrying over the excess stock for following periods as well as delivering back orders later. The application of the Beer Game (Forrester 1958; Niranjan et al. 2009; Yang et al. 2011; Ancarani et al. 2013) instead of the newsvendor is interesting as it would address both the shortcoming of the newsvendor regarding the dynamically complex decision making task as well as individual versus group decision making. Variations regarding system restrictiveness and decisional guidance relying on principles derived from MLA can lead to distinct results when experimental treatments vary the frequency of decisions made by the group or part of the group. Moreover, the application of a task where participants must consider flows and stock would potentially provide an alternative explanation to why individuals are not good at understanding the effects of flows on a stock (e.g., Sweeney & Sterman 2000; Strohhecker & Größler 2013).

Regarding individual differences, further research should include alternative psychometric scales since previous research suggests that individual differences significantly affect decision-making performance. Although the selected scales fail to support the individual differences hypothesis completely, other scales can potentially highlight missing personality aspects that can potentially predict planning performance.

6.7 Final reflection

The interest in the human element in OM and OR came from the frustration of watching people making mistakes over and over, without losing enthusiasm. Over the last years collaborating with industry, I was lucky to observe anecdotal evidence of managers worried to gather evidence to support their points of views rather than to develop a point of view. Any analysis, in practice seems to be in good part intuition driven, mostly used to validate rather than to discover. Coming down to small, day-to-day tasks, people are not afraid of making small changes without realising that if everyone does the same it can be a problem. Small changes would not be a problem if they were perfectly random – this way it would simply average itself and cause no harm. However, small changes are governed by some common decision patterns, heuristics, and prone to similar biases. This way, much like in a boat if all passengers make a small step to the same side, the whole boat will capsize.

The interesting fact about biases is its similarity with illusions. The illusion does not disappear even if people are aware of it. During my years of formal arts education during the life-drawing sessions, most of the training was focused on working around the illusion. One particularity of the life-drawing practice was that nobody was interested in making the illusion of multidimensional space disappear, it was considered a pointless effort, and the common direction was to learn how to live with the illusions and the shortcomings of the mind. Instead of repeatedly showing individuals what they did wrong, the training provided people with tools and warnings about when the mind will be playing a trick on them. People were provided with a mechanical aid to live with the illusion, the visual bias. This aid, as a framework was given in form of physical cues such as a straight pencil at a fixed distance, or a checked grid placed between the eye and the object. Great masters such as Leonardo Da Vinci during the Renascence, Johannes Vermeer during the Dutch Golden Age or Caravaggio during the Baroque movements used mechanical aids such as *camera obscura*, *camera lucida* and fixed concave mirrors with directional light to paint in a realistic style. Regardless of the criticism of ‘cheating’ on the art of painting, some of the greatest masters of all times relied on systems to paint. And what else is a painting than a virtually infinite number of decisions made by the artist. Although knowing about the illusion did not make it disappear, a solid mechanical aid (a framework) provided means to draw realistically.

REFERENCES

Abdellaoui, M., Vossmann, F. & Weber, M., 2005. Choice-Based Elicitation and Decomposition of Decision Weights for Gains and Losses Under Uncertainty. *Management Science*, 51(9), pp.1384–1399.

Ackerman, P.L., 1996. A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, 22(2), pp.227–257.

Agrawal, V. & Seshadri, S., 2000. Impact of Uncertainty and Risk Aversion on Price and Order Quantity in the Newsvendor Problem. *Manufacturing & Service Operations Management*, 2(January 2015), pp.410–423.

Allais, M., 1953. Le Comportement de l â€™ Homme Rationnel devant le Risque : Critique des Postulats et Axiomes de l â€™ Ecole Americaine. *Econometrica: Journal of the Econometric Society*, 21(4), pp.503–546.

Ancarani, a., Di Mauro, C. & D'Urso, D., 2013. A human experiment on inventory decisions under supply uncertainty. *International Journal of Production Economics*, 142(1), pp.61–73.

Anderson, J., 1996. ACT: A simple theory of complex cognition. *American Psychologist*, 51(4), pp.355–365.

Apgar, V., 1953. A proposal for a new method of evaluation of the newborn infant. *Current researches in anesthesia & analgesia*, 32(4), pp.260–7.

Armstrong, J.S., 2001. *Principles of forecasting: a handbook for researchers and practitioners*, Springer Science & Business Media.

Arnold, V., Collier, P. a., Leech, S. a. & Sutton, S.G., 2000. The effect of experience and complexity on order and recency bias in decision making by professional accountants. *Accounting and Finance*, 40(2), pp.109–134.

Ashenfelter, O., 2008. Predicting the quality and prices of bordeaux wine*. *The Economic Journal*, 118(529), pp.174–184.

Ashton, R.H. & Kennedy, J., 2002. Eliminating recency with self-review: the case of auditors' "going concern" judgments. *Journal of Behavioral Decision Making*, 15(3), pp.221–231.

Atkins, P.W.., Wood, R.E. & Rutgers, P.J., 2002. The effects of feedback format on dynamic decision making. *Organizational Behavior and Human Decision Processes*, 88(2), pp.587–604.

Azar, O.H., 2014. The default heuristic in strategic decision making: When is it optimal to choose the default without investing in information search? *Journal of Business Research*, 67(8), pp.1744–1748.

Bachrach, D. & 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.

Barakat, A. & Othman, A., 2005. The Five-Factor Model of Personality and its Relationship to Cognitive Style (Rush and Prudence) and Academic Achievement among a Sample of Students. , 1735, pp.156–165.

Bazerman, M.H., 2005. *Judgement in Managerial Decision Making* 6th ed., New York: John Wiley & Sons.

Becker-Peth, M., Katok, E. & Thonemann, U.W., 2013. Designing contracts for irrational but predictable newsvendors. *Management Science2*, 59(8), pp.1800–1816.

Benartzi, S. & Thaler, R., 1995. Myopic loss aversion and the equity premium puzzle. *The quarterly journal of Economics*, 110(1), pp.73–92.

Bendoly, E., 2006. Incorporating behavioral theory in OM empirical models. *Journal of Operations Management*, 24(6), pp.735–736.

Bendoly, E., Croson, R., Goncalves, P. & Schultz, K., 2009. Bodies of Knowledge for Research in Behavioral Operations. *Production and Operations Management*, 19(4), pp.434–452.

Benzion, U., Cohen, Y., Peled, R. & Shavit, T., 2008. Decision-making and the newsvendor problem: an experimental study. *Journal of the Operational Research Society*, 59(9), pp.1281–1287.

Benzion, U., Cohen, Y. & Shavit, T., 2009. The newsvendor problem with unknown distribution. *Journal of the Operational Research Society*, 61(6), pp.1022–1031.

Berkowitz, L. & Donnerstein, E., 1982. External validity is more than skin deep: Some answers to criticisms of laboratory experiments. *American Psychologist*, 37(3), pp.245–257.

Binmore, K., 2009. *Rational decisions*, Princeton & Oxford: Princeton University Press.

Blackburn, J.D., Kropp, D.H. & Millen, R.A., 1986. A Comparison of Strategies to Dampen Nervousness in MRP Systems. *INFORMS - Management Science*, 32(4), pp.413–429.

Blackburn, J.D., Kropp, D.H. & Millen, R.A., 1985. MRP system nervousness: Causes and cures. *Engineering Costs and Production Economics*, 9(1–3), pp.141–146.

Blattberg, R.C., Hoch, S.J. & Science, M., 1990. Database Models and Managerial Intuition: 50% Model + 50% Manager. *Management Science*, 36(8), pp.887–899.

Blavatskyy, P.R., 2005. Back to the St. Petersburg Paradox? *Management Science*, 51(4), pp.677–678.

Blavatskyy, P.R. & Pogrebna, G., 2010. Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics*, 25(6), pp.963–986.

Bolton, G.E. & Katok, E., 2004. Learning-by-Doing in the Newsvendor Problem A Laboratory Investigation of the Role of Experience and Feedback.

Bolton, G.E. & Katok, E., 2008. Learning by Doing in the Newsvendor Problem: A Laboratory Investigation of the Role of Experience and Feedback. *Manufacturing & Service Operations Management*, 10(3), pp.519–538.

Bolton, G.E., Ockenfels, a. & Thonemann, U.W., 2012. Managers and Students as Newsvendors. *Management Science*.

Bone, J., Hey, J.D. & Suckling, J., 2009. Do people plan? *Experimental Economics*, 12(1), pp.12–25.

Bostian, A.A., Holt, C.A. & Smith, A.M., 2008. Newsvendor “Pull-to-Center” Effect: Adaptive Learning in a Laboratory Experiment. *Manufacturing & Service Operations Management*, 10(4), pp.590–608.

Bowers, J.S. & Davis, C.J., 2012. Bayesian just-so stories in psychology and neuroscience. *Psychological bulletin*, 138(3), pp.389–414.

Brehmer, B., 1992. Dynamic decision making: Human control of complex systems. *Acta Psychologica*, 81(3), pp.211–241.

Brenner, L.L.A., Koehler, D.D.J., Liberman, V. & Tversky, A., 1996. Overconfidence in probability and frequency judgments: A critical examination. *Organizational Behavior and Human Decision Processes*, 65(3), pp.212–219.

Brown, A.O. & Tang, C.S., 2006. The impact of alternative performance measures on single-period inventory policy. *Journal of Industrial and Management Optimization*, 2(3), pp.297–318.

Burstein, F. & Holsapple, C., 2008. *Handbook on Decision Support Systems 1 : Basic Themes*,

Buss, D.M., 1999. Adaptive individual differences revisited. *Journal of Personality*, 67(2), pp.259–264.

Buss, D.M. & Greiling, H., 1999. Adaptive individual differences. *Journal of Personality*, 67(2), pp.209–243.

Buss, D.M. & Hawley, P.H., 2010. *The Evolution of Personality and Individual Differences*, Oxford University Press.

Camerer, C. & Johnson, E., 1991. The process-performance paradox in expert judgment: How can experts know so much and predict so badly? *Toward a General Theory of Expertise: Prospects and Limits*, pp.195–217.

Camerer, C.F., Loewenstein, G. & Rabin, M., 2004. *Advances in behavioral economics* C. F. CAMERE, G. LOEWENSTEIN, & M. RABIN, eds., Oxford: Princeton University Press.

Campbell, K.L., 1971. Scheduling is not the problem. *Production and Inventory Management*, 12(2), pp.53–60.

Carey, J.M., 1991. The Effects of Graphical Versus Numerical Response on the Accuracy of Graph-Based Forecasts. *Journal of Management*, 17(1), pp.77–96.

Carlson, J.A. & Keefe, T.B.O., 1969. Buffer Stocks and Reaction Coefficients: An Experiment with Decision Making Under Risk. *The Review of Economic Studies*, 36(4), pp.467–484.

Carter, C., 2011. A call for theory: the maturation of the supply chain management discipline. *Journal of Supply Chain Management*, 42(2), pp.3–7.

Carter, C.R., Kaufmann, L. & 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.

Charness, G., Karni, E. & Levin, D., 2007. Individual and group decision making under risk: An experimental study of Bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and Uncertainty*, 35(2), pp.129–148.

Chater, N., Tenenbaum, J.B. & Yuille, A., 2006. Probabilistic models of cognition: conceptual foundations. *Trends in cognitive sciences*, 10(7), pp.287–91.

Chen, A., Hsu, C.-H. & Blue, J., 2007. Demand planning approaches to aggregating and forecasting interrelated demands for safety stock and backup capacity planning. *International Journal of Production Research*, 45(10), pp.2269–2294.

Childerhouse, P., Hermiz, R., Mason-Jones, R., Popp, A. & Towill, D.R., 2003a. Information flow in automotive supply chains – identifying and learning to overcome barriers to change. *Industrial Management & Data Systems*, 103(3), pp.137–149.

Childerhouse, P., Hermiz, R., Mason-Jones, R., Popp, A. & Towill, D.R., 2003b. Information flow in automotive supply chains – present industrial practice. *Industrial Management & Data Systems*, 103(3), pp.137–149.

Christopher, M., 2011. *Logistics & Supply Chain Management* 4th ed., FT Press.

Christopher, M. & Lee, H.L., 2001. Supply chain confidence. *Dulles: Vastera Inc*, (November).

Clemen, R.T., 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4), pp.559–583.

Conrad, N. & Patry, M., 2012. Conscientiousness and academic performance: a mediational analysis. *International Journal for the Scholarship of Teaching and Learning*, 6(1), pp.1–14.

Craighead, C.W., Blackhurst, J., Rungtusanatham, M.J. & Handfield, R.B., 2007. The Severity of Supply Chain Disruptions: Design Characteristics and Mitigation Capabilities. *Decision Sciences*, 38(1), pp.131–156.

Crandall, C.S., 1984. The Overcitation of Examples of Poor Performance: Fad, Fashion, or Fun? In Defense of the Citation Bias in the Judgment Literature. , (December), pp.1499–1500.

Croson, R. & Donohue, K., 2006. Behavioral Causes of the Bullwhip Effect and the Observed Value of Inventory Information. *Management Science*, 52(3), pp.323–336.

Croson, R. & Donohue, K., 2002. Experimental Economics and Supply-Chain Management. *Interfaces*, 32(5), pp.74–82.

Croson, R., Ren, Y. & Croson, R., 2013. Overconfidence in Newsvendor Orders: An Experimental Study. *Management Science*, 59(11), pp.2502–2517.

Croson, R., Schultz, K., Siemsen, E. & Yeo, M.L., 2013. Behavioral operations: The state of the field. *Journal of Operations Management*, 31(1–2), pp.1–5.

Cui, Y., Chen, L.G., Chen, J., Gavirneni, S. & Wang, Q., 2013. Chinese perspective on newsvendor bias: An exploratory note. *Journal of Operations Management*, 31(1–2), pp.93–97.

Dahlbäck, O., 1990. Personality and risk-taking. *Personality and Individual Differences*, 11(12), pp.1235–1242.

Dane, E., Rockmann, K.W. & Pratt, M.G., 2012. When should I trust my gut? Linking domain expertise to intuitive decision-making effectiveness. *Organizational Behavior and Human Decision Processes*, 119(2), p.187.

Daston, L., 1980. Probabilistic expectation and rationality in classical probability theory. *Historia Mathematica*, 7, pp.234–260.

Davern, M.J., Mantena, R. & Stohr, E.A., 2008. Diagnosing decision quality. *Decision Support Systems*, 45(1), pp.123–139.

Davis, C., Patte, K., Tweed, S. & Curtis, C., 2007. Personality traits associated with decision-making deficits. *Personality and Individual Differences*, 42(2), pp.279–290.

Davis, D.L., 1982. Are Some Cognitive Types Better Decision Makers Than Others? An Empirical Investigation. *Human Systems Management*, 3(3), p.165.

Deary, I.J., 2000. *Looking down on human intelligence : information processing and cognitive ability differences*, Oxford: Oxford University Press.

Deming, W.E., 1986. *Out of the crisis*, MIT Press.

Dieckmann, A. & Rieskamp, J., 2007. The influence of information redundancy on probabilistic inferences. *Memory & cognition*, 35(7), pp.1801–13.

Digman, J.M., 1997. Higher-order factors of the Big Five. *Journal of personality and social psychology*, 73(6), pp.1246–1256.

Digman, J.M., 1990. Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology*, 41(1), pp.417–440.

Doerr, K.H., Freed, T., Mitchell, T.R., Schriesheim, C. a & Zhou, X.T., 2004. Work flow policy and within-worker and between-workers variability in performance. *The Journal of applied psychology*, 89(5), pp.911–921.

Donihue, M.R., 1993. Evaluating the Role Judgment Plays in Forecast Accuracy. *Journal of Forecasting*, 12(July 1991), pp.81–92.

Donnellan, M.B., Oswald, F.L., Baird, B.M. & Lucas, R.E., 2006. The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment*, 18(2), pp.192–203.

Van Donselaar, K., Van Den Nieuwenhof, J. & Visschers, J., 2000. Impact of material coordination concepts on planning stability in supply chains. *International Journal of Production Economics*, 68(2), pp.169–176.

Dreyfus, H., 1992. *What computers still can't do: a critique of artificial reason*, Cambridge: The MIT Press.

Edgeworth, F.Y. & Society, R.S., 1888. The Mathematical Theory of Banking. *Journal of the Royal Statistical Society*, 51(1), pp.113–127.

Edmondson, A. & McManus, S., 2007. Methodological fit in management field research. *Academy of management review*, 32(4), pp.1155–1179.

Edmundson, R.H., 1990. Decomposition; a strategy for judgemental forecasting. *Journal of Forecasting*, 9(4), pp.305–314.

Edwards, W. & Fasolo, B., 2001. Decision technology. *Annual review of psychology*, 52(1), pp.581–606.

Einhorn, H. & Hogarth, R., 1981. Behavioral decision theory: Processes of judgment and choice. *Journal of Accounting Research*, 32, pp.53–88.

Ellsberg, D., 1961. Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics*, 75, pp.643–669.

Evans, J., 1984. In defense of the citation bias in the judgment literature. , (December), pp.1500–1501.

Feng, T., Keller, L.R. & Zheng, X., 2011. Decision making in the newsvendor problem: A cross-national laboratory study. *Omega*, 39(1), pp.41–50.

Fildes, R. & Beard, C., 1992. Forecasting Systems for Production and Inventory Control. *International Journal of Operations & Production Management*, 12(5), pp.4–27.

Fildes, R. & Goodwin, P., 2007. Against Your Better Judgement? How Organizations Can Improve Their Use of Mangement Judgement in Forecasting. *Interfaces*, 37(6), pp.570–576.

Fildes, R., Goodwin, P. & Lawrence, M., 2006. The design features of forecasting support systems and their effectiveness. *Decision Support Systems*, 42(1), pp.351–361.

Fildes, R., Goodwin, P., Lawrence, M. & Nikolopoulos, K., 2009. Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25(1), pp.3–23.

Fildes, R. & Hastings, R., 1994. The Organization and Improvement of Market Forecasting. *The Journal of the Operational Research Society*, 45(1), pp.1–16.

Fildes, R. & Stekler, H., 2002. The state of macroeconomic forecasting. *Journal of Macroeconomics*, 24(4), pp.435–468.

Filiz, E. & Battaglio, R.P., 2015. Personality and decision-making in public administration: the five-factor model in cultural perspective. *International Review of Administrative Sciences*.

Fisher, M. & Raman, a., 1996. Reducing the Cost of Demand Uncertainty Through Accurate Response to Early Sales. *Operations Research*, 44(1), pp.87–99.

Fleeson, W. & Jayawickreme, E., 2015. Whole Trait Theory. *Journal of Research in Personality*, 56(2015), pp.82–92.

Forrester, J.W., 1958. Industrial dynamics: a major breakthrough for decision makers. *Harvard business review*, 36(4), pp.37–66.

Franken, I.H.A. & Muris, P., 2005. Individual differences in decision-making. *Personality and Individual Differences*, 39(5), pp.991–998.

Franses, P.H. & Legerstee, R., 2013. Do statistical forecasting models for SKU-level data benefit from including past expert knowledge? *International Journal of Forecasting*, 29(1), pp.80–87.

Frederick, S., Kahneman, D. & Mochon, D., 2010. Elaborating a simpler theory of anchoring. *Journal of Consumer Psychology*, 20(1), pp.17–19.

Funder, D.C., 1987. Errors and mistakes: evaluating the accuracy of social judgment. *Psychological bulletin*, 101(1), pp.75–90.

Funke, J., 2001. Dynamic systems as tools for analysing human judgement. *Thinking and Reasoning*, 7(1), pp.69–89.

Furnham, A., 2008. *Personality and intelligence at work: Exploring and explaining individual differences at work*, Routledge.

Gallego, G., 1995. IEOR E4000 Production Management. *IEOR 4000: Production Management*, (April), pp.1–11.

Gambetti, E., Fabbri, M., Bensi, L. & Tonetti, L., 2008. A contribution to the Italian validation of the General Decision-making Style Inventory. *Personality and Individual Differences*, 44(4), pp.842–852.

Gans, N. & Croson, R., 2008. Introduction to the Special Issue on Behavioral Operations. *Manufacturing & Service Operations Management*, 10(4), pp.563–565.

Gavirneni, S. & Isen, A.M., 2009. Anatomy of a Newsvendor Decision: Observations from a Verbal Protocol Analysis. *Production and Operations Management*, 19(4), pp.453–462.

Gavirneni, S. & Xia, Y., 2009. Anchor Selection and Group Dynamics in Newsvendor Decisions--A Note. *Decision Analysis*, 6(2), pp.87–97.

Geary, S., Disney, S.M. & 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.

Gigerenzer, G., 2001. The adaptive toolbox: Toward a Darwinian rationality. *Nebraska symposium on motivation*.

Gigerenzer, G., 2008. Why Heuristics Work. *Perspectives on Psychological Science*, 3(1), pp.20–29.

Gigerenzer, G. & Brighton, H.J., 2009. Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science*, 1, pp.107–143.

Gigerenzer, G. & Gaissmaier, W., 2011. Heuristic decision making. *Annual review of psychology*, 62, pp.451–482.

Gigerenzer, G. & Goldstein, D.G., 1996. Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, 103(4), pp.650–69.

Gneezy, U., Kapteyn, A. & Potters, J., 2003. Evaluation Periods and Asset Prices in a Market Experiment. *The Journal of Finance*, 58(2), pp.821–837.

Gneezy, U., Meier, S. & Rey-Biel, P., 2011. When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, 25(4), pp.1–21.

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

Gneezy, U. & Rustichini, A., 2000. Pay Enough or Don't Pay at All*. *Quarterly Journal of Economics*, 115(3), pp.791–810.

Goldberg, L.R., 1990. An alternative "description of personality": the Big-Five factor structure. *Journal of Personality and Social Psychology*, 59(6), pp.1216–1229.

Goldberg, L.R., 1992. The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1), pp.26–42.

Goldstein, D.G. & Gigerenzer, G., 2002. Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109(1), pp.75–90.

Goodall, A.H. & Pogrebna, G., 2015. Expert leaders in a fast-moving environment. *The Leadership Quarterly*, 26(2), pp.123–142.

Goodwin, P., 2000. Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting*, 16(1), pp.85–99.

Goodwin, P., 2002. Integrating management judgment and statistical methods to improve short-term forecasts. *Omega*, 30(2), pp.127–135.

Goodwin, P. & Fildes, R., 1999. Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy? *Journal of Behavioral Decision Making*, 12(1), pp.37–53.

Goodwin, P., Fildes, R., Lawrence, M. & Nikolopoulos, K., 2007. The process of using a forecasting support system. *International Journal of Forecasting*, 23(3), pp.391–404.

Goodwin, P., Fildes, R., Lawrence, M. & Stephens, G., 2011. Restrictiveness and guidance in support systems. *Omega*, 39(3), pp.242–253.

Goodwin, P., Onkal-Atay, D., Thomson, M.E., Pollock, A.C. & Macaulay, A., 2004. Feedback-labelling synergies in judgmental stock price forecasting. *Decision Support Systems*, 37(1), pp.175–186.

Goodwin, P., Sinan Gönül, M. & Önkal, D., 2012. Antecedents and effects of trust in forecasting advice. *International Journal of Forecasting*, 29(2), pp.354–366.

Goodwin, P. & Wright, G., 1993. Improving judgmental time series forecasting: A review of the guidance provided by research. *International Journal of Forecasting*, 9(2), pp.147–161.

Gosling, S.D., Rentfrow, P.J. & Swann, W.B., 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), pp.504–528.

Größler, A., 2004. Don't let history repeat itself-methodological issues concerning the use of simulators in teaching and experimentation. *System Dynamics Review*, 20(3), pp.263–274.

Gurnani, H. & Tang, C.S., 1999. Note: Optimal Ordering Decisions with Uncertain Cost and Demand Forecast Updating. *Management Science*, 45(10), pp.1456–1462.

Haigh, M. & List, J., 2005. Do professional traders exhibit myopic loss aversion? An experimental analysis. *The Journal of Finance*, LX(1), pp.523–534.

Harrison, G.W. & List, J.A., 2004. Field experiments. *Journal of Economic literature*, 42(4), pp.1009–1055.

Harvey, N., 1995. Why Are Judgments Less Consistent in Less Predictable Task Situations? *Organizational Behavior and Human Decision Processes*, 63(3), pp.247–263.

Harvey, N. & Bolger, F., 1996. Graphs versus tables: Effects of data presentation format on judgemental forecasting. *International Journal of Forecasting*, 12(1), pp.119–137.

Hayes, R.H. & Clark, K.B., 1985. Explaining observed productivity differentials between plants: Implications for operations research. *Interfaces*, 15(6), pp.3–14.

Healy, a, 2000. Comparing Serial Position Effects in Semantic and Episodic Memory Using Reconstruction of Order Tasks,. *Journal of Memory and Language*, 42(2), pp.147–167.

Heaven, P.C.L. & Ciarrochi, J., 2012. When IQ is not everything: Intelligence, personality and academic performance at school. *Personality and Individual Differences*, 53(4), pp.518–522.

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

Hey, J.D., 2005. Why we should not be silent about noise. *Experimental Economics*, 8(4 SPEC. ISS.), pp.325–345.

Hey, J.D. & Knoll, J.A., 2007. How far ahead do people plan? *Economics Letters*, 96(1), pp.8–13.

Hilbert, M., 2012. Toward a synthesis of cognitive biases: how noisy information processing can bias human decision making. *Psychological bulletin*, 138(2), pp.211–37.

Ho, C.-J., 1989. Evaluating the impact of operating environments on MRP system nervousness. *International Journal of Production Research*, 27(7), p.1115.

Hogarth, R.M. & Karelaia, N., 2007. Heuristic and linear models of judgment: matching rules and environments. *Psychological review*, 114(3), pp.733–58.

Hurtz, G.M. & Donovan, J.J., 2000. Personality and job performance: the Big Five revisited. *The Journal of applied psychology*, 85(6), pp.869–879.

Hwang, M.I., 1995. The effectiveness of graphic and tabular presentation under time pressure and task complexity. *Information Resources Management Journal*, 8(3), p.25.

Icellioglu, S. & Ozden, M.S., 2012. Personality Effect on Decision-Making Among Normal University Students and Comparing Gambling Task Performances of Normal University Students with a Group of Normal Adults. *Procedia - Social and Behavioral Sciences*, 46, pp.1010–1014.

Jiang, J.J. & Klein, G., 2000. Side effects of decision guidance in decision support systems. *Interacting with Computers*, 12(5), pp.469–481.

John, O.P., Naumann, L.P. & Soto, C.J., 2008. Paradigm shift to the integrative Big Five Trait taxonomy. *Handbook of personality: Theory and research*, pp.114–158.

John, O.P. & Srivastava, S., 1999. *The Big Five trait taxonomy: History, measurement, and theoretical perspectives* 2nd ed., NY: The Guilford Press.

Jung, C., 1951. *Psychological Types*,

Kadipasaoglu, S.N. & Sridharan, V., 1995. Alternative approaches for reducing schedule instability in multistage manufacturing under demand uncertainty. *Journal of Operations Management*, 13(3), pp.193–211.

Kagel, J.H., Roth, A.E. & Hey, J.D., 1995. *The handbook of experimental economics*, Princeton university press Princeton, NJ.

Kahneman, D., 2003. Maps of Bounded Rationality: Psychology for Behavioral. *The American Economic Review*, 93(5), pp.1449–1475.

Kahneman, D., 2012. *Thinking, fast and slow*, New York: Penguin.

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

Kahneman, D. & Tversky, A., 1979. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 47(2), pp.263–292.

Kahneman, D. & Tversky, A., 1972. Subjective probability: A judgment of representativeness. *Cognitive psychology*, 3(3), pp.430–454.

Katsikopoulos, K. V, 2010. The less-is-more effect : Predictions and tests. *Judgment and Decision Making*, 5(4), pp.244–257.

Katsikopoulos, K. V. & Gigerenzer, G., 2013. Behavioral Operations Management: A Blind Spot and a Research Program. *Journal of Supply Chain Management*, 49(1), pp.3–7.

Kaufmann, L., Carter, C.R. & Buhrmann, C., 2010. Debiasing the supplier selection decision: a taxonomy and conceptualization. *International Journal of Physical Distribution & Logistics Management*, 40(10), pp.792–821.

Kerkkänen, A., Korpela, J. & Huiskonen, J., 2009. Demand forecasting errors in industrial context:

Measurement and impacts. *International Journal of Production Economics*, 118(1), pp.43–48.

Kleinmuntz, B., 1990. Why we still use our heads instead of formulas: toward an integrative approach. *Psychological bulletin*, 107(3), pp.296–310.

de Kok, T. & Inderfurth, K., 1997. Nervousness in inventory management: Comparison of basic control rules. *European Journal of Operational Research*, 103(1), pp.55–82.

Kremer, M., Minner, S. & Van Wassenhove, L.N., 2010. Do Random Errors Explain Newsvendor Behavior? *Manufacturing & Service Operations Management*, 12(4), pp.673–681.

Kremer, M., Moritz, B. & Siemsen, E., 2011. Demand Forecasting Behavior: System Neglect and Change Detection. *Management Science*, 57(10), pp.1827–1843.

Krueger, J.I. & Funder, D.C., 2004. Towards a balanced social psychology: causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *The Behavioral and brain sciences*, 27(3), pp.313-27-76.

Kruglanski, A.W. & Gigerenzer, G., 2011. Intuitive and deliberate judgments are based on common principles. *Psychological review*, 118(1), pp.97–109.

Langer, T. & Weber, M., 2008. Does commitment or feedback influence myopic loss aversion? *Journal of Economic Behavior & Organization*, 67(3–4), pp.810–819.

Langer, T. & Weber, M., 2005. Myopic prospect theory vs. myopic loss aversion: how general is the phenomenon? *Journal of Economic Behavior & Organization*, 56(1), pp.25–38.

Lapide, B.L., 2013. Supply ' s Demand- Shaping Roles.

Lapide, L., 2007. SALES AND OPERATIONS PLANNING (S&OP) MINDSETS. *The Journal of Business Forecasting*, 26(1), p.21–22,31.

Laplace, P. marquis de, 1902. *A philosophical essay on probabilities* First Edit., London: John Wiley & Sons, Ltd.

Lariviere, M., 1999. Supply Chain Contracting and Coordination with Stochastic Demand. *Quantitative models for supply chain management*, pp.233–268.

Larrick, R.P., Koehler, D.J. & Harvey, N., 2004. Debiasing. In D. J. Koehler & N. Harvey, eds. *Blackwell Handbook of Judgment and Decision Making*. Blackwell Publishing Ltd, pp. 316–337.

Larsen, R.J. & Buss, D.M., 2010. *Personality Psychology* 4th ed., McGraw-Hill.

Lauriola, M. & Levin, I.P., 2001. Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Personality and Individual Differences*, 31(2), pp.215–226.

Lawrence, M., Goodwin, P. & Fildes, R., 2002. Influence of user participation on DSS use and decision accuracy. *Omega*, 30(5), pp.381–392.

Lawrence, M. & Sim, W., 1999. Prototyping a financial DSS. *Omega*, 27(4), pp.445–450.

Lawrence, M.J., Edmundson, R.H. & O'Connor, M.J., 1985. An examination of the accuracy of judgmental extrapolation of time series. *International Journal of Forecasting*, 1(1), pp.25–35.

Lawrence, M.J., Edmundson, R.H. & O'Connor, M.J., 1986. The accuracy of combining judgemental and statistical forecasts. *Management Science*, 32(12), pp.1521–1532.

Lee, H.L., 2004. The triple-A supply chain. *Harvard business review*, 82(10), pp.102–12, 157.

Lee, H.L., Padmanabhan, V. & Whang, S., 1997. Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science*, 43(4), pp.546–558.

Leighton, R.T., 1981. Decision Support Systems. *Journal of Systems Management*, 32(2), p.40.

Lichtenstein, S. & Slovic, P., 1971. Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89(1), pp.46–55.

Liersch, M.J. & McKenzie, C.R.M., 2009. Duration neglect by numbers—And its elimination by graphs. *Organizational Behavior and Human Decision Processes*, 108(2), pp.303–314.

Lim, J.S. & O'Connor, M., 1996. Judgmental forecasting with interactive forecasting support systems.

Decision Support Systems, 16(4), pp.339–357.

Lim, J.S. & O 'connor, M., 1995. Judgemental Adjustment of Initial Forecasts: Its Effectiveness and Biases. *Journal of Behavioral Decision Making*, 8(January), pp.149–168.

Lin, V.S., Goodwin, P. & Song, H., 2014. Accuracy and bias of experts' adjusted forecasts. *Annals of Tourism Research*, 48, pp.156–174.

List, J.A., Sadoff, S. & Wagner, M., 2011. So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Experimental Economics*, 14(4), pp.439–457.

Loomes, G., Moffatt, P.G. & Sugden, R., 2002. A microeconometric test of alternative stochastic theories of risky choice. *Journal of Risk and Uncertainty*, 24(2), pp.103–130.

Loomes, G., Pinto-Prades, J.L., Abellan-Perpinan, J.M. & Rodriguez-Miguez, E., 2008. Modelling Noise and Imprecision in Individual Decisions. , (October 2008), pp.1–32.

Loomes, G. & Pogrebna, G., 2015. Do preference reversals disappear when we allow for probabilistic choice?

Loomes, G. & Sugden, R., 1982. Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal*, 92(368), pp.805–824.

Lopes, L.L., 1991. The Rhetoric of Irrationality. *Theory & Psychology*, 1(1), pp.65–82.

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

Ma, L., 2008. Loss-averse newsvendor problem with general profit target. *International Journal of Information and Decision Sciences*, 1(2), p.145.

Ma, L., Xue, W., Zhao, Y. & Zeng, Q., 2015. Loss-averse newsvendor problem with supply risk. *Journal of the Operational Research Society*, pp.1–15.

Maltby, J., Day, L. & Macaskill, A., 2013. Personality Individual Differences and Intelligence.

Martin, L.E. & Potts, G.F., 2009. Impulsivity in decision-making: An event-related potential investigation. *Personality and Individual Differences*, 46(3), pp.303–308.

Masatlioglu, Y. & Ok, E. a., 2005. Rational choice with status quo bias. *Journal of Economic Theory*, 121(1), pp.1–29.

Mason-Jones, R. & Towill, D.R., 2000. Coping with Uncertainty : Reducing " Bullwhip " Behaviour in Global Supply Chains.

Mason-Jones, R. & Towill, D.R., 1998. Shrinking the supply chain uncertainty circle. *IOM control*, 24(7), pp.17–22.

Mathews, B.P. & Diamantopoulos, A., 1990. Judgemental Revision of Sales Forecasts : Effectiveness of Forecast Selection. *Journal of Forecasting*, 9(April 1989), pp.407–415.

Mayhew, B.W. & Vitalis, A., 2014. Myopic loss aversion and market experience. *Journal of Economic Behavior and Organization*, 97(2014), pp.113–125.

Meehl, P.E., 1954. *Clinical versus Statistical Prediction*, Minneapolis: University of Minnesota.

Mihaela, P.L., 2015. Personality Variables in Decision - Making. *Procedia - Social and Behavioral Sciences*, 187, pp.658–662.

von Mises, R., 1957. *Probability, Statistics and Truth*, New York: The Macmillan Company.

Montazemi, A.R., Wang, F., Khalid Nainar, S.M. & Bart, C.K., 1996. On the effectiveness of decisional guidance. *Decision Support Systems*, 18(2), pp.181–198.

Moore, D. a. & Cain, D.M., 2007. Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes*, 103(2), pp.197–213.

Moritz, B., Siemsen, E. & Kremer, M., 2014. Judgmental forecasting: Cognitive reflection and decision speed. *Production and Operations Management*, 23(7), pp.1146–1160.

Moritz, B.B., Hill, A. V. & Donohue, K., 2009. Cognition and Individual Differences in the Newsvendor Problem: Behavior Under Dual Process Theory. *Working Paper*, 104(25), pp.10335–9.

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

Nenkov, G.Y., Inman, J.J. & Hulland, J., 2008. Considering the Future: The Conceptualization and Measurement of Elaboration on Potential Outcomes. *Journal of Consumer Research*, 35(1), pp.126–141.

Neumann, J. Von & Morgenstern, O., 1947. *The theory of games and economic behavior* 2nd ed., New Jersey: Princeton University Press.

Niranjan, T.T., Metri, B.A. & Aggarwal, V., 2009. Behavioral Causes of Bullwhip Effect: Breaking the Mould. *International Journal of Services and Operations Management*, 5(3), pp.350–374.

O'Connor, M., Remus, W. & Griggs, K., 1993. Judgemental forecasting in times of change. *International Journal of Forecasting*, 9(2), pp.163–172.

O'Gorman, K. & MacIntosh, R., 2012. Research Methods for Business Students. *the Global Management Series*, (SEPTEMBER), pp.1–696.

Ockenfels, A. & Selten, R., 2015. Impulse Balance and Multiple-Period Feedback in the Newsvendor Game. *Production and Operations Management*, 24(2014), p.n/a-n/a.

Oreg, S. & Bayazit, M., 2009. Prone to bias: Development of a bias taxonomy from an individual differences perspective. *Review of General Psychology*, 13(3), pp.175–193.

Ouwersloot, H., Nijkamp, P. & Rietveld, P., 1998. Errors in probability updating behaviour: Measurement and impact analysis. *Journal of Economic Psychology*, 19(5), pp.535–563.

Ovchinnikov, A., Moritz, B. & Quiroga, B.F., 2015. How to Compete against a Behavioral Newsvendor. *Production and Operations Management*, 24(11), pp.1783–1793.

Parackal, M., Goodwin, P. & O'Connor, M., 2007. Judgement in forecasting. *International Journal of Forecasting*, 23(3), pp.343–345.

Pareto, V. & Page, A.N., 1971. Translation of Manuale di economia politica ("Manual of political economy"). *AM Kelley*.

Parikh, M., Fazlollahi, B. & Verma, S., 2001. The effectiveness of decisional guidance: An empirical evaluation. *Decision Sciences*, 32(2), pp.303–332.

Patton, J.H., Stanford, M.S. & Barratt, E.S., 1995. Factor Structure of the Barratt Impulsiveness Scale. *Journal of Clinical Psychology*, 51(6), pp.768–774.

Payne, J., Bettman, J. & Johnson, E., 1993. *The adaptive decision maker*, New York: Cambridge University Press.

Pervin, L.A., 1994. A Critical Analysis of Current Trait Theory. *Psychological Inquiry*, 5(2), pp.103–113.

Plott, C.R. & Smith, V.L., 2008. *Handbook of experimental economics results*, Elsevier.

Poropat, A.E., 2009. A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135(2), pp.322–338.

Power, D.J., Sharda, R. & Burstein, F., 2015. Decision Support Systems. In *Wiley Encyclopedia of Management*. John Wiley & Sons, Ltd, pp. 629–633.

Remus, W., O'Connor, M. & Griggs, K., 1998. The impact of incentives on the accuracy of subjects in judgmental forecasting experiments. *International Journal of Forecasting*, 14(4), pp.515–522.

Rieg, R., 2010. Do forecasts improve over time? A case study of the accuracy of sales forecasting at a German car manufacturer. *International Journal of Accounting & Information Management*, 18(3), pp.220–236.

Rudi, N. & Drake, D., 2010. Observation Bias: The Impact of Demand Censoring on Newsvendor Level and Adjustment Behavior. *SSRN Electronic Journal*, (October 2015), pp.1–24.

Samuelson, P., 1938. A Note on the Pure Theory of Consumer's Behaviour. *Economica*, 5(17), pp.61–71.

Sanders, N.R. & Manrodt, K.B., 2003. Forecasting software in practice: Use, satisfaction, and performance. *Interfaces (Providence)*, 33(5), p.90.

Sanders, N.R. & Ritzman, L.P., 2001. Judgmental Adjustment of Statistical Forecasts. *Principles of forecasting*, pp.405–416.

Savage, L., 1954. *The Foundations of Statistics*, Wiley.

Savage, L., 1961. *The foundations of statistics reconsidered*, Berkeley: University of California Press.

Schkade, D.A. & Kleinmuntz, D.N., 1994. Information Displays and Choice Processes: Differential Effects of Organization, Form, and Sequence. *Organizational Behavior and Human Decision Processes*, 57(3), pp.319–337.

Schweitzer, M.E., Cachon, P. & Cachon, G.P., 2000. Decision Bias in the Newsvendor Problem with a Known Demand Distribution: Experimental Evidence. *Management Science*, 46(November 2015), pp.404–420.

Schwenk, C., 1988. The cognitive perspective on strategic decision making. *Journal of Management Studies*, 25(1), pp.41–55.

Scott, S.G. & Bruce, R.A., 1995. Decision-making style: The development and assessment of a new measure. *Educational and Psychological Measurements*, 55(5), pp.818–831.

Sedlmeier, P. & Gigerenzer, G., 2000. Was Bernoulli wrong? On intuitions about sample size. *Journal of Behavioral Decision Making*, 13(1), pp.133–139.

Shah, A.K. & Oppenheimer, D.M., 2008. Heuristics made easy: an effort-reduction framework. *Psychological bulletin*, 134(2), pp.207–22.

Shiloh, S., Salton, E. & Sharabi, D., 2002. Individual differences in rational and intuitive thinking styles as predictors of heuristic responses and framing effects. *Personality and Individual Differences*, 32(3), pp.415–429.

Silver, M.S., 1991. Decisional Guidance for Computer-Based Decision Support. *MIS Quarterly*, 15(1), pp.105–122.

Simon, H., 1955. A behavioral model of rational choice. *The quarterly journal of economics*, 69(1), pp.99–118.

Simon, H., 1990. Invariants of human behavior. *Annual review of psychology*, 41, pp.1–19.

Simon, H., 1969. *The sciences of the artificial* 3rd ed., London: The MIT Press.

Simon, H., 1992. What is an "explanation" of behavior? *Psychological Science*, 3(3), pp.150–161.

Simon, H.A., 1987. Bounded Rationality. In M. M. & P. N. J. Eatwell, ed. *The new Palgrave: A Dictionary of Economics*. London and Basingstoke: Macmillan.

Simon, H.A., 1956. Rational choice and the structure of the environment. *Psychological review*, 63(2), pp.129–38.

Simon, H.A., 1979. Rational Decision Making in Business Organizations. *The American Economic Review*, 69(4), p.493.

Simpson, N.., 1999. Multiple level production planning in rolling horizon assembly environments. *European Journal of Operational Research*, 114(1), pp.15–28.

Singh, D.T., 1998. Incorporating cognitive aids into decision support systems: The case of the strategy execution process. *Decision Support Systems*, 24(2), pp.145–163.

Sjöberg, L., 2003. Intuitive vs. analytical decision making: Which is preferred? *Scandinavian Journal of Management*, 19(1), pp.17–29.

Smith, V.L., 2002. Constructivist and ecological rationality in economics. *American Economic Review*, 93(3), pp.502–561.

Soane, E. & Chmiel, N., 2005. Are risk preferences consistent? The influence of decision domain and personality. *Personality and Individual Differences*, 38(8), pp.1781–1791.

Soros, G., 2009. *The Crash of 2008 And What it Means: The New Paradigm for Financial Markets*, New York:

PublicAffairs.

Spearman, C., 1904. “General Intelligence,” Objectively Determined and Measured. *The American Journal of Psychology*, 15(2), pp.201–292.

Srivastava, J. & Raghubir, P., 2002. Debiasing Using Decomposition: The Case of Memory-Based Credit Card Expense Estimates. *Journal of Consumer Psychology*, 12(3), pp.253–264.

Stadtler, H., 2005. Supply chain management and advanced planning––basics, overview and challenges. *European Journal of Operational Research*, 163(3), pp.575–588.

Stadtler, H., Kilger, C. & Meyr, H., 2015. *Supply Chain Management and Advanced Planning* 5th ed. H. Stadtler & C. Kilger, eds., Springer Berlin Heidelberg.

Stanovich, K.E. & West, R.F., 2000. Individual differences in reasoning: implications for the rationality debate? *The Behavioral and brain sciences*, 23(5), pp.645-665-726.

Starmer, C., Loomes, G., Bardsley, N., Moffatt, P., Sugden, R. & Cubitt, R., 2009. *Experimental Economics: Rethinking the Rules*, Princeton: Princeton University Press.

Sterman, J.D. & Dogan, G., 2015. “I’m not hoarding, I’m just stocking up before the hoarders get here.”: Behavioral causes of phantom ordering in supply chains. *Journal of Operations Management*, 39–40, pp.6–22.

Stigler, G., 1961. The economics of information. *The journal of political economy*, 69(3), pp.213–225.

Stiglitz, J.E., 2010. *Freefall: America, Free Markets, and the Sinking of the World Economy*, New York: W. W. Norton & Company.

Strohhecker, J. & Größler, A., 2013. Do personal traits influence inventory management performance? - The case of intelligence, personality, interest and knowledge. *International Journal of Production Economics*, 142(1), pp.37–50.

Su, X., 2008. Bounded Rationality in Newsvendor Models. *Manufacturing & Service Operations Management*, 10(4), pp.566–589.

Sun, J. & Xu, X., 2015. Coping with Loss Aversion in the Newsvendor Model. *Discrete Dynamics in Nature and Society*, 2015, pp.1–11.

Surowik, D., 2002. Leonard Savage’s Mathematical Theory of Decision. *Studies in Logic, Grammar and Rhetoric*, 5(18), pp.65–75.

Sutter, M., 2007. Are teams prone to myopic loss aversion? An experimental study on individual versus team investment behavior. *Economics Letters*, 97(2), pp.128–132.

Sweeney, L.B. & Sterman, J., 2000. Bathtub Dynamics : Initial Results of a Systems Thinking Inventory Bathtub Dynamics : Initial Results of a Systems Thinking Inventory. *System Dynamics Review*, 16(4), pp.249–286.

Syntetos, A.A., Babai, Z., Boylan, J.E., Kolassa, S. & Nikolopoulos, K., 2016. Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, 252(1), pp.1–26.

Syntetos, A.A., Nikolopoulos, K., Boylan, J.E., Fildes, R. & Goodwin, P., 2009. The effects of integrating management judgement into intermittent demand forecasts. *International Journal of Production Economics*, 118(1), pp.72–81.

Taleb, N.N., 2012. *Antifragile: things that gain from disorder*, New York: Random House.

Tan, L. & Ward, G., 2000. A recency-based account of the primacy effect in free recall. *Journal of experimental psychology. Learning, memory, and cognition*, 26(6), pp.1589–625.

Taylor, P.F. & Thomas, M.E., 1982. Short term forecasting: horses for courses. *Journal of the Operational Research Society*, 33(8), pp.685–694.

Thaler, R., 2000. From homo economicus to homo sapiens. *The Journal of Economic Perspectives*, 14(1), pp.133–141.

Thaler, R., Tversky, a., Kahneman, D. & Schwartz, a., 1997. The effect of myopia and loss aversion on risk taking: An experimental test. *The Quarterly Journal of Economics*, 112(2), pp.647–661.

Thaler, R.H., 1999. Mental accounting matters. *Journal of Behavioral decision making*, 12, pp.183–206.

Theocharis, Z. & Harvey, N., 2016. Order effects in judgmental forecasting. *International Journal of Forecasting*, 32(1), pp.44–60.

Tiwari, D., Patil, R. & Shah, J., 2011. Unreliable newsboy problem with a forecast update. *Operations Research Letters*, 39(4), pp.278–282.

Todd, P. & Benbasat, I., 1999. Evaluating Effort , and of DSS , Impact Cognitive Incentives on Strategy Selection. *Information Systems Research*, 10(4), pp.356–374.

Todd, P., Gigerenzer, G. & ABC Research Group, 2012. *Ecological rationality: Intelligence in the world*, Oxford: Oxford University Press.

Todd, P.M., 2007. How much information do we need? *European Journal of Operational Research*, 177(3), pp.1317–1332.

Todd, P.M. & Gigerenzer, G., 2007. Environments That Make Us Smart. *Association for Psychological Science*, 16(3), pp.167–171.

Todd, P.M. & Gigerenzer, G., 2000. Précis of Simple heuristics that make us smart. *The Behavioral and brain sciences*, 23(5), pp.727-41-80.

Tokar, T., 2010. Behavioural research in logistics and supply chain management. *The International Journal of Logistics Management*, 21(1), pp.89–103.

Tsay, a. a., Nahmias, S. & Agrawal, N., 1999. Modelling supply chain contracts: a review. *Quantitative Models for Supply Chain Management*, pp.300–336.

Tversky, A. & Kahneman, D., 1974. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), pp.1124–1131.

Tversky, A. & Kahneman, D., 1991. Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106, pp.1039–1061.

Tversky, A. & Kahneman, D., 1983. Probabilistic Reasoning. , 1131(1974), pp.1124–1131.

Tversky, a & Kahneman, D., 1992. Advances in Prospect-Theory - Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5(4), pp.297–323.

De Véricourt, F., Jain, K., Bearden, J.N. & Filipowicz, A., 2013. Sex, risk and the newsvendor. *Journal of Operations Management*, 31(1–2), pp.86–92.

Vitasek, B.K.L., Manrodt, K.B. & Kelly, M., 2003. Solving the supply-demand mismatch. *Supply Chain Management Review*, pp.58–64.

Vranas, P.B., 2000. Gigerenzer's normative critique of Kahneman and Tversky. *Cognition*, 76(3), pp.179–93.

Vriend, N., 1996. Rational behavior and economic theory. *Journal of Economic Behavior & Organization*, 29, pp.263–285.

Wachtel, R.E. & Dexter, F., 2010. Review article: review of behavioral operations experimental studies of newsvendor problems for operating room management. *Anesthesia and Analgesia*, 110(6), pp.1698–710.

Wagner, S.M. & Bode, C., 2006. An empirical investigation into supply chain vulnerability. *Journal of Purchasing and Supply Management*, 12(6), pp.301–312.

Wakker, P.P., 2010. Prospect Theory.

Wang, C. & Webster, S., 2009. The loss-averse newsvendor problem☆. *Omega*, 37(1), pp.93–105.

Webby, R. & O'Connor, M., 1996. Judgemental and statistical time series forecasting: A review of the literature. *International Journal of Forecasting*, 12(1), pp.91–118.

Webby, R., O'Connor, M. & Edmundson, B., 2005. Forecasting support systems for the incorporation of event information: An empirical investigation. *International Journal of Forecasting*, 21(3), pp.411–423.

Weber, E.U. & Milliman, R.A., 1997. Perceived risk attitudes: Relating risk perception to risky choice.

Management Science, 43(2), pp.123–144.

Weinman, J., Elithorn, A. & Cooper, R., 1985. Personality and problem-solving: The nature of individual differences in planning, scanning and verification. *Personality and Individual Differences*, 6(4), pp.453–460.

Whitin, T.M., 1955. Inventory Control and Price Theory. *Management Science*, 2, pp.61–68.

Whybark, D.C. & Williams, J.G., 1976. Material requirements planning under uncertainty. *Decision sciences*, 7(4), pp.595–606.

Wilke, A. & Todd, P.M., 2010. Past and present environments: the evolution of decision making. *Psicothema*, 22(1), pp.4–8.

Williams, B., Myerson, J. & Hale, S., 2008. Individual Differences, Intelligence, and Behavior Analysis. *Journal of the Experimental Analysis of Behavior*, 90(2), pp.219–231.

Wold, H., 1954. *A Study in the Analysis of Stationary Time Series* 2nd ed., Uppsala: Almqvist and Wiksell Book Co.

Wu, D.Y. & Katok, E., 2006. Learning, communication, and the bullwhip effect. *Journal of Operations Management*, 24(6), pp.839–850.

Wübben, M. & Wangenheim, F., 2008. Instant customer base analysis: Managerial heuristics often "get it right." *Journal of Marketing*, 72(May), pp.82–93.

Xie, J., Zhao, X. & Lee, T.., 2003. Freezing the master production schedule under single resource constraint and demand uncertainty. *International Journal of Production Economics*, 83(1), pp.65–84.

Yang, T., Wen, Y.-F. & Wang, F.-F., 2011. Evaluation of robustness of supply chain information-sharing strategies using a hybrid Taguchi and multiple criteria decision-making method. *International Journal of Production Economics*, 134(2), pp.458–466.

Zhao, X. & Lam, K., 1997. Lot-sizing rules and freezing the master production schedule in material requirements planning systems. *International Journal of Production Economics*, 53(3), pp.281–305.

Zhao, X. & Lee, T.S., 1993. Freezing the master production schedule for material requirements planning systems under demand uncertainty. *Journal of Operations Management*, 11(2), pp.185–205.

Zheng, M., Wu, K. & Shu, Y., 2016. Newsvendor Problems with Demand Forecast Updating and Supply Constraints. *Computers & Operations Research*, 67, pp.193–206.

Zmud, R.W., 1979. Individual Differences and MIS Success: A Review of the Empirical Literature. *Management Science*, 25(10), p.966.

APPENDICES

Appendix A Mini-IPIP scale

20 Item Mini-IPIP (Donnellan et al. 2006)

Item	**Factor**	**Text** (1 = Very Inaccurate to 5 = Very Accurate)	**Score**
1	Extroversion	Am the life of the party.	
2	Agreeableness	Sympathize with others' feelings	
3	Conscientiousness	Get chores done right away.	
4	Neuroticism	Have frequent mood swings (rapid changes).	
5	Intellect/Imagination	Have a vivid imagination.	
6	Extroversion	Don't talk a lot.	Reverse
7	Agreeableness	Am not interested in other people's problems.	Reverse
8	Conscientiousness	Often forget to put things back in their proper place.	Reverse
9	Neuroticism	Am relaxed most of the time.	Reverse
10	Intellect/Imagination	Am not interested in abstract ideas.	Reverse
11	Extroversion	Talk to a lot of different people at parties.	
12	Agreeableness	Feel others' emotions.	
13	Conscientiousness	Like order.	
14	Neuroticism	Get upset easily.	
15	Intellect/Imagination	Have difficulty understanding abstract ideas.	Reverse
16	Extroversion	Keep in the background.	Reverse
17	Agreeableness	Am not really interested in others.	Reverse
18	Conscientiousness	Make a mess of things.	Reverse
19	Neuroticism	Rarely feel blue (sad).	Reverse
20	Intellect/Imagination	Do not have a good imagination.	Reverse

Appendix B EPO scale

Elaboration on Potential Outcomes (EPO) by Nenkov et al. (2008)

Sub-scales	Items (1 = strongly disagree to 7 = strongly agree)
Generation/evaluation dimension	Before I act I consider what I will gain or lose in the future as a result of my actions.
	I try to anticipate as many consequences of my actions as I can.
	Before I make a decision I consider all possible outcomes.
	I always try to assess how important the potential consequences of my decisions might be.
	I try hard to predict how likely different consequences are.
	Usually I carefully estimate the risk of various outcomes occurring.
Positive outcome focus dimension	I keep a positive attitude that things always turn out all right.
	I prefer to think about the good things that can happen rather than the bad.
	When thinking over my decisions I focus more on their positive end results.
Negative outcome focus dimension	I tend to think a lot about the negative outcomes that might occur as a result of my actions.
	I am often afraid that things might turn out badly.
	When thinking over my decisions I focus more on their negative end results.
	I often worry about what could go wrong as a result of my decisions.

Appendix C BIS scale

BIS-11 Patton et al. (1995)

BIS-11 Questions (scale:1-4 from rarely/never - Almost Always/Always)		
1. Attentional Impulsiveness	1. Attention	11. I “squirm” at plays or lectures. 28. I am restless at the theatre or lectures. 5. I don’t “pay attention.” 9. I concentrate easily. (R) 20. I am a steady thinker. (R)
	6. Cognitive instability	6. I have “racing” thoughts 24. I change hobbies.
2. Motor Impulsiveness	5. Perseverance	21. I change residences. 16. I change jobs. 30. I am future oriented. (R) 23. I can only think about one problem at a time. 26. I often have extraneous thoughts when thinking.
	2. Motor Impulsiveness	17. I act “on impulse.” 19. I act on the spur of the moment. 22. I buy things on impulse. 3. I make up my mind quickly. 2. I do things without thinking. 25. I spend or charge more than I earn. 4. I am happy-go-lucky
3. Non-planning impulsiveness	4. Cognitive complexity	15. I like to think about complex problems. (R) 29. I like puzzles. (R) 10. I save regularly. (R) 27. I am more interested in the present than the future. 18. I get easily bored when solving thought problems.
	3. Self-control	12. I am a careful thinker. (R) 1. I plan tasks carefully. (R) 8. I am self-controlled. (R) 7. I plan trips well ahead of time. (R) 13. I plan for job security. (R) 14. I say things without thinking.

Appendix D GDMS scale

GDMS - General Decision Making Style Items

Sub-Scales	Items (1 = strongly disagree to 5 = strongly agree)
Rational	I double-check my information sources to be sure I have the right facts before making decisions
	I make decisions in a logical and systematic way.
	My decision making requires careful thought.
	When making a decision, I consider various options in terms of a specific goal.
	I explore all of my options before making a decision.
Avoidant	I avoid making important decisions until the pressure is on.
	I postpone decision making whenever possible.
	I often procrastinate when it comes to making important decisions.
	I generally make important decisions at the last minute.
	I put off making many decisions because thinking about them makes me
Dependent	I often need the assistance of other people when making important decisions.
	I rarely make important decisions without consulting other people.
	If I have the support of others, its easier for me to make important decisions.
	I use the advice of other people in making my important decisions.
	I like to have someone to steer me in the right direction when I am faced with important decisions.
Intuitive	When making decisions, I rely upon my instincts.
	When I make decisions, I tend to rely on my intuition.
	I generally make decisions that feel right to me.
	When I make a decision, it is more important for me to feel the decision is right than to have a rational reason for it.
	When I make a decision, I trust my inner feelings and reactions.
Spontaneous	I generally make snap decisions.
	I often make decisions on the spur of the moment.
	I make quick decisions.
	I often make impulsive decisions.
	When making decisions, I do what seems natural at the moment.

Appendix E Participant targeting groups

Group	Members	Obs
Demand Planning, Sales Forecasting, IBP and Supply Chain Optimization	34,908	https://www.linkedin.com/groups/1808515
Demand Planners, Supply Chain Planners, Forecasters	13,112	https://www.linkedin.com/groups/1064377/profile
S&OP, Demand Planning, Supply Chain in Germany	233	https://www.linkedin.com/groups/8192918/profile
Supply Chain, Demand Planning, Production Planning, Supply Network, Logistics, Transportation	141	https://www.linkedin.com/groups/6710984/profile
Demand Planners,Supply Chain Planners,Forecasters	13,111	https://www.linkedin.com/groups/1064377/profile
Sales & Operations Planning Network	8,361	https://www.linkedin.com/groups/1812222/profile
Sales & Operations Planning / CPFR	8,191	https://www.linkedin.com/groups/1524967/profile
S&OP – Sales and Operations Planning, Forecasting, Demand Management and Supply Planning	13,323	https://www.linkedin.com/groups/3989507/profile
Total Estimated Reach	**Aprox. 35,000**	Estimate based on largest group assuming overlap of members
SCIP Collaborators Forum	29	Actively spread the survey within the company
SCIP Network	115	
SCIP Mailing	1455	
Total mailing	**1772**	
Incentivised students across Warwick University registered to SONA system (https://warwick.sona-systems.com)	1639	Pre-requisites: • Active students • Based in the UK • Fluent in English • Normal vision
Non incentivised students (MSc students attending supply chain modules and others)	Aprox. 1500	Estimate based on total students at WMG

Appendix F Experiment code (JavaScript)

```
var demand = [227, 241, 113, 228, 169, 149, 169, 198, 171, 93, 53, 291, 242, 23, 83, 82, 99, 193, 47, 76,
190, 161, 212, 21, 233, 242, 54, 298, 211, 241, 178, 195, 198, 187, 162, 17, 113, 24, 236]; //Uniform
pre-generaged signal
var demandNow = 0;
var satisfied = 0;
var overage = 0;
var shortage = 0;
var maxPeriod = demand.length; //change back to normal demand
var forecast = 225; // optimal order quantity
var planNow = 0;
var plan = []; // array plotting player's answers
var demandPast = []; // array plotting known demand
var price = 12;
var manufCost = 3;
var roundManufCost = 0;
var savCost = 0;
var lostSalesCost = 0;
var roundRevenue = 0;
var totalProfit = 0;
var roundProfit = 0;
var round = 1; // substitute of the i in the loop, starts with 1, so must be always compensated wuth -1
when used in arrays
var roundNow = 1;
var step = 3;
var practiceRounds = 10; //exclusive the last e.g., if 4 training rounds then the payer has 3 steps
var wastedProductValue = 0;
var disposalProductValue = 0;
var foregoneSalesValue = 0;
var min = 1; // number of treatments lower bound inclusive
var max = 3; // number of treatments higher bound inclusive

var $ =jQuery.noConflict(); // makes the Plot.ly work

//Graphical interface
var trace1 = {
 x: ["P1", "P2", "P3", "P4", "P5", "P6", "P7", "P8", "P9", "R1", "R2", "R3", "R4", "R5", "R6", "R7", "R8", "R9",
"R10", "R11", "R12", "R13", "R14", "R15", "R16", "R17", "R18", "R19", "R20", "R21", "R22", "R23", "R24",
"R25", "R26", "R27", "R28", "R29", "R30"
],
 y: demandPast,
 name:"Demand",
 type:"bar"
};

var trace2 = {
 x: ["P1", "P2", "P3", "P4", "P5", "P6", "P7", "P8", "P9", "R1", "R2", "R3", "R4", "R5", "R6", "R7", "R8", "R9",
"R10", "R11", "R12", "R13", "R14", "R15", "R16", "R17", "R18", "R19", "R20", "R21", "R22", "R23", "R24",
"R25", "R26", "R27", "R28", "R29", "R30"
],
 y: plan,
 name:"Plan",
 type:"bar"
};

var data = [trace1, trace2];
var layout = {
	title: 'Results',
	bargap:0.1,
	bargroupgap:0.1,
	xaxis:{
		title:"Periods",
		fixedrange:true
	},
	yaxis:{
		title:"Volume",
		fixedrange:true
	},
```

```
        autosize: true,
        margin:{
                b:30,
                l:50,
                r:0,
                pad:0,
                t:30,
                autoexpand:true
        },
        legend:{
                bordercolor:"rgba(0, 0, 0, 0)",
                yanchor:"auto",
                traceorder:"normal",
                xanchor:"auto",
                bgcolor:"rgba(255, 255, 255, 0)",
                borderwidth:1,
                y:-0.3,
                x:1,
                font:{
                color:"",
                family:"",
                size:11
                },
        },
}
//================================================================
var treatmentGenerator = function(min, max) { //this generates a uniform random number between min and max so a treatment can be assigned
  return treatment = Math.floor(Math.random() * (max - min +1)) + min;
};

var treatment = treatmentGenerator(min, max); //this is the number of the treatment from min to max in an uniform distribution
//================================================================
if (treatment === 1) {
                        document.getElementById("planQuestion").innerHTML = "What is the planned quantity for round " + round + "?";
} else if(treatment === 2) {
                if ((round-1)%step === 0) { // to modify the step, just devide by the commitment length
                        document.getElementById("planQuestion").innerHTML = "What is the planned quantity for round " + round + "? Note that you must order the same quantity for rounds " + (round + 1) + " and " + (round + 2);
                }
} else if(treatment === 3) {
                if ((round-1)%step === 0) { // to modify the step, just devide by the commitment length
                        document.getElementById("planQuestion").innerHTML = "What is the planned quantity for round " + round + "? Note that you should order the same quantity for rounds " + (round + 1) + " and " + (round + 2);
                }
}

var resultsRound = function(){
  roundManufCost = manufCost * planNow;
  if (demand[round-1]<=planNow) { // if the demand is less than planned
    satisfied = demand[round-1]; //satisfied demand if demand is lower than planned
    overage = planNow - demand[round-1]; // overage a.k.a. unsold inventory is null
    shortage = 0 // shortage of product
  } else { // if the demand is more than planned
    satisfied = planNow; // satisfied demand if demand is greater than planned
    overage = 0; // overage a.k.a. unsold inventory
    shortage = demand[round-1] - planNow // shortage of product
  }
  roundRevenue = satisfied * price; // value of sales for the round
  wastedProductValue = overage * manufCost; // the cost of wasted products, e.g., disposal cost
  disposalProductValue = overage * savCost; // cost of disposing product
  foregoneSalesValue = shortage * lostSalesCost; //the cost of not meeting the demand in full
```

```
  roundProfit = roundRevenue - disposalProductValue - foregoneSalesValue - roundManufCost; // profit of the round including all revenue minus all the costs
  totalProfit =  totalProfit + roundProfit; //updates the total profit with the new profit
  plan[round-1] = planNow; // shows the planned orders so far by the player
  demandNow = demand[round-1]; // updates the demand for the current round
  demandPast[round-1] = demand[round-1]; // shows known demand to the present round
  round++; // this is what advances the round counter
};

//===============================================================
var getPlanNow = function(){ //Function that picks the order value and processes the results fields, this is activates with the submit button
        if (round < practiceRounds) {
                roundNow = round;
                document.getElementById("practice").innerHTML = "Practice round " + round + " out of " + (practiceRounds - 1); // initiates
                //document.getElementById("round").innerHTML = "Practice round: " + round;
        } else {
                roundNow = round - practiceRounds + 1;
                document.getElementById("practice").innerHTML = "Real round: " + roundNow  + " out of " + (maxPeriod - practiceRounds + 1); // otherwise the round is 0 and must be 1
        }

        planNow = document.getElementById("planNowField").value;
        document.getElementById("planNow").innerHTML = "Your planned quantity: " + planNow;
        //document.getElementById("graphic").src= graphics[round-1]; // shows the graphic

        resultsRound(); // runs computation of the results
        Plotly.newPlot('myDiv', data, layout);
        // updating the fields that show the results
        document.getElementById("roundDemand").innerHTML = "Round demand: " + demandNow + " units";
        document.getElementById("roundProfit").innerHTML = "Round profit: " + roundProfit + " tokens";
        document.getElementById("roundManufacturingCost").innerHTML = "Round manufacturing cost: " + roundManufCost + " tokens";
        document.getElementById("roundSatisfiedDemand").innerHTML = "Round satisfied demand: " + satisfied + " units";
        document.getElementById("roundRevenue").innerHTML = "Round revenue: " + roundRevenue + " tokens";
        document.getElementById("roundExcessProduct").innerHTML = "Round excess product: " + overage + " units";
        document.getElementById("roundWaste").innerHTML = "Round waste value: " + wastedProductValue + " tokens";
        document.getElementById("roundShortageProduct").innerHTML = "Round shortage product: " + shortage + " units";
        document.getElementById("totalProfit").innerHTML = "Total cumulative profit: " + totalProfit + " tokens";
        //document.getElementById("plan").innerHTML = "So far orders: " + plan;
        //document.getElementById("demandPast").innerHTML = "So far demand: " + demandPast;

//===============================================================
// Game flow for different treatments and practice vs. real rounds
//===============================================================
        if(treatment === 1) {
                document.getElementById("planQuestion").innerHTML = "What is the planned quantity for round " + (roundNow + 1) + "?";
                document.getElementById("planNowField").value = 0; //reset the order field to zero
        } else if(treatment === 2) {
                        if ((round-1)%step === 0) {
                                document.getElementById("planQuestion").innerHTML = "What is the planned quantity for round " + (roundNow + 1) + "? Note that you must order the same quantity for rounds " + (roundNow + 2) + " and " + (roundNow + 3);
                                document.getElementById("planNowField").disabled = false;  // shows the order field
                                document.getElementById("planNowField").value = 0; //reset the order field to zero
                        } else {
```

```
                    document.getElementById("planQuestion").innerHTML =
"Resubmit the previous planned quantity of " + planNow + " units for the next round";
                    document.getElementById("planNowField").disabled = true; //
hides the order field
                }
    } else {
                if ((round-1)%step === 0) {
                    document.getElementById("planQuestion").innerHTML = "What
is the planned quantity for round " + (roundNow + 1) + "? Note that you should order the same quantity
for rounds " + (roundNow + 2) + " and " + (roundNow + 3);
                    document.getElementById("planNowField").value = 0; //reset
the order field to zero
                } else {
                    document.getElementById("planQuestion").innerHTML = "Do
you want to keep the previously planned quantity of " + planNow + " units for the next round?";
                }
    }
//===================================================================
    if(round < practiceRounds){ // tests if we reached the end of the practice
                //nothing
            } else if (round === practiceRounds) {
                document.getElementById("practice").innerHTML = "Practice round: " +
(round - 1) + " - End of practice rounds, total profit will be reset after this round";
                totalProfit = 0; // resets the profit
                if (treatment === 1) {
                    document.getElementById("planQuestion").innerHTML = "What
is the planned quantity for real round 1?";
                } else if (treatment === 2) {
                    document.getElementById("planQuestion").innerHTML = "What
is the planned quantity for real round 1? Note that you must order the same quantity for real rounds 2 and
3";
                } else if (treatment === 3) {
                    document.getElementById("planQuestion").innerHTML = "What
is the planned quantity for real round 1? Note that you should order the same quantity for real rounds 2
and 3";
                }

            } else {
                //normal game
            }

//===================================================================
    if(round <= maxPeriod){ // tests if we reached the end of the game
                // nothing happens, game continues
            } else {
                document.getElementById("buttonSubmitPlan").style.visibility = 'hidden';
                document.getElementById("planNowField").style.visibility = 'hidden';
                document.getElementById("planQuestion").innerHTML = "Task complete -
Proceed by hitting 'Next' at the bottom of the page";
                console.log("TotalProfit = " + totalProfit);
            }
 }; // end of getPlanNow function

//*******************************************************************
// STATIC STUFF: first run Populate interface with values that are static and dont change over the game
//*******************************************************************
    document.getElementById("practice").innerHTML = "Practice round 0 out of " +
(practiceRounds - 1); // initiates
    //document.getElementById("round").innerHTML = "Round: " + 0; // initial first round that will
be updated
    document.getElementById("price").innerHTML = "Price: " + price + " tokens / unit";
    document.getElementById("cost").innerHTML = "Cost: " + manufCost + " tokens / unit";
    document.getElementById("forecast").innerHTML = "Forecast: " + forecast + " units";

Qualtrics.SurveyEngine.addOnload(function()
{
    // saves the variables into the response DB
            Qualtrics.SurveyEngine.setEmbeddedData("plan", plan);     //first field is the
embeded object in qualtrics, second field is the variable I want to save
```

```
                              Qualtrics.SurveyEngine.setEmbeddedData("treatment", treatment);

               // Hide the Next Button. Show it again when the user reaches round 40
                              this.disableNextButton();
    this.questionclick = function(event,element)
               {
        if (round > maxPeriod)
        {
            this.enableNextButton();
                                         Qualtrics.SurveyEngine.setEmbeddedData("totalProfit", totalProfit);
                              }
               }


});
```

Appendix G Graphical user interface

The planning task

Basic instructions

- This is an experiment in decision-making. You will function as a planner of a single product
- The experiment is composed of a number of rounds in which you will be asked to make planning decisions.
- In each round you are able to plan the manufacturing quantity of the product which has a cost to manufacture. The Sales department will then sell the product to customers at a higher price.
- Your manufacturing department is able to fulfil any quantity planned.
- Consumer demand in each round is randomly selected from a distribution.
- The system offers an estimate of the future demand, you are free to use it or not.
- The prices and profits in every round will be in experiment tokens.
- Your objective is to maximise the profit.

Possible scenarios

- **Excess of stock** – If fewer products are demanded than the quantity you planned, you will have to dispose of the unsold products (i.e. you cannot keep unsold products for future periods).
- **Shortage** – If more products are demanded than the quantity you planned, you will have to forgo some sales.

Data after each round

After providing the planned quantity to the manufacturing department in each round, that quantity will be manufactured and then delivered to the customer without any issues, the realized demand and the profit will be presented to you at the beginning of the next round. The results are displayed in an interactive chart at the bottom of the page.

The task flow

You will be given 9 rounds of practice before the actual game starts.

The flow between the practice rounds and the actual game is seamless, total cumulative profit is reset.

Task

Price: 12 tokens / unit

Cost: 3 tokens / unit

Forecast: 225 units

Real round: 3 out of 30

Your planned quantity: 150

Round demand: 291 units

Round profit: 1350 tokens

Round manufacturing cost: 450 tokens

Round satisfied demand: 150 units

Round revenue: 1800 tokens

Round excess product: 0 units

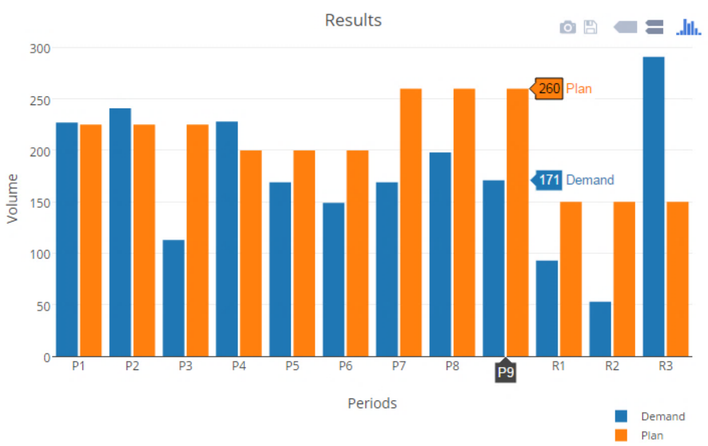
Round waste value: 0 tokens

Round shortage product: 141 units

Total cumulative profit: 2202 tokens

What is the planned quantity for round 4? Note that you must order the same quantity for rounds 5 and 6

250 Submit



Appendix H Electronic information sheet (professionals)

Version 2
Date: 23-02-2016

THE UNIVERSITY OF WARWICK

Electronic Information Sheet

Principal Investigator:	Mr. Alexander Kharlamov
Project Title:	A Study on Planning and Personality
Participation Duration:	15 minutes

Contact	**Contact Type**	**E-mail**
Mr. Alexander Kharlamov	Principal Investigator	a.a.kharlamov@warwick.ac.uk
Prof. Dr. Janet Godsell	Supervisor	j.godsell@warwick.ac.uk
Dr. Ganna Pogrebna	Supervisor	g.pogrebna@warwick.ac.uk

Research Purpose

We would like to invite you to take part in a research study on planning and personality. This study will consist of a series of decisions you will make. This study will enhance our knowledge on planning decisions and may lead to the development of new decision aids for business and policy makers.

Information on Research

The study will consist of a series of decisions made by you. Should you choose to participate, you will be able to read the instructions describing all the aspects of the study. In this study, decisions will consist of selecting an option with a mouse or touch-screen on a computer/tablet/smartphone screen or typing a number in a designated field, moving a slider or organizing options in your preferred order by drag-and-drop tool. Once you have made all decisions, you will be asked to complete a simple demographic questionnaire for statistical purposes.

Risks

There are no physical risks of any kind involved in this study. There is a potential loss of confidentiality of the data you supply. Yet, we store all data on secure servers and this risk is minimal. You will not be asked to provide your name, address, or any other sensitive information as a part of this study. Should you decide to receive an individual feedback from this study, you **may opt** to provide us with your email address so this feedback could be sent to you. Your email information **will not be shared with any third parties** and will be used **only for sending you individual feedback**.
Once the questionnaire has been submitted, the data cannot be retrieved unless you opted to provide us with your email address so we can retrieve and delete the record upon request.

Benefits

The data from this study will help us to advance understanding of individual planning decisions. This will have important business and policy implications.

Anonymity

You will not receive any information about results and identities of other participants in this study. Likewise, other participants will not receive any information about your identity and your results in this study. Participants' email addresses (should such addresses be provided for individual feedback) will be kept separate from the study data. **All email addresses will be deleted after individual feedback is provided.** Please, note that once your data is collected we will not be able to retrieve your data should you decide to withdraw it at a later stage if you do not provide us with your email address. However, please, also note that your email address (if supplied) might reveal some of your personal information. The study data will include only an **identification number for each participant**. The data will be electronically stored on the University of Warwick secure server equipped with the latest safe authentication methodology and secure TLS tunnel.

Voluntary Participation

Participation in this study is voluntary. If at any point you wish to stop your participation in this study, you can do so.

Approval

This study was approved by BSREC under REGO-2015-1736

Complaints

If you have any comments or complaints about this study, you may contact the members of research team (listed above) as well as:

Director of Delivery Assurance
Registrar's Office
University House
University of Warwick
Coventry
CV4 8UW
Complaints@warwick.ac.uk
024 7657 4774

Appendix I Consent form (professionals)

Version 2
Date: 23-02-2016

Online Consent Form

Study: Why Are Some Planners More Susceptible to Myopic Loss Aversion than Others?

Research Group Contacts:

Alexander A. Kharlamov, PhD Researcher, International Institute for Product and Service Innovation, WMG, The University of Warwick, Coventry CV4 7AL UK, e-mail: a.a.kharlamov@warwick.ac.uk, tel: +44 (0)24 7657 3484

Professor Janet Godsell, Professor of Operations and Supply Chain Strategy, International Institute for Product and Service Innovation, WMG, The University of Warwick, Coventry CV4 7AL UK, e-mail: j.godsell@warwick.ac.uk, tel: +44 (0)24 7657 3482

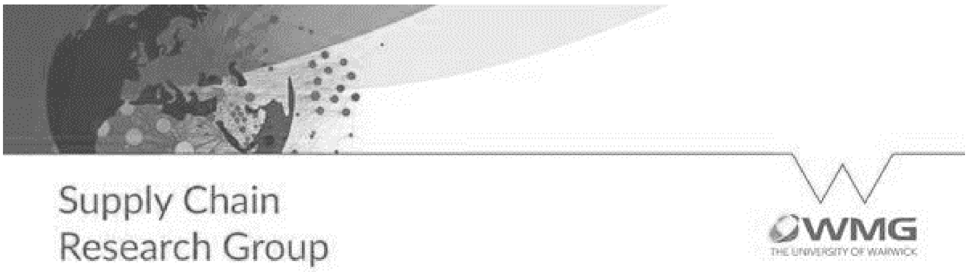
Dr. Ganna Pogrebna, Associate Professor, International Institute for Product and Service Innovation, WMG, The University of Warwick, Coventry CV4 7AL, e-mail: g.pogrebna@warwick.ac.uk, tel: +44 (0)24 7657 3530

1. I confirm that I am 18 years of age or older.
2. I confirm that I have read and understood the Electronic Information Sheet for this project.
3. I agree to take part in the above study and am willing to follow study instructions and procedures and complete all tasks.
4. I understand that my information will be held and processed for the purposes of publication in scientific journals and presentation on scientific conferences.
5. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without being penalised or disadvantaged in any way.

I understand that by clicking the PROCEED button below I agree with all of the above statements.

<PROCEED BUTTON>

Appendix J Invitations for participation



We are inviting you to take part in a study on Planning and Personality.

The aim of this study is to explore the relationship between planning decisions and personality preferences. Ultimately we want to understand how people behave to inform planning policy.

The study consists of one simple planning task followed by a series of questions to identify your personality preferences.

It takes on average 15 minutes to complete. To participate please follow the link: https://warwickwmg.eu.qualtrics.com/SE/?SID=SV_73wdOSHCRmfWG8J

This study is cross-industry and does not aim at any specific industry. The targeted population are professionals with planning experience.

Your participation is anonymous. However, if you are interested in your individual results or/and wish to have access to the research report, you can choose to submit your e-mail at the end of the study. Your individual results will not be shared with anyone except yourself. The individual results and the research report will be completely anonymised and will not include any identifying information that can be linked back to you or to your company. Attached is a sample Individual Report (please note that some items might be excluded from the final version) and the information sheet about the study.

Finally, in order to guarantee the success of this study we will be grateful if you could disseminate it through your professional network both inside and outside your company.

Thank you,

Appendix K Sample individual report

ID104

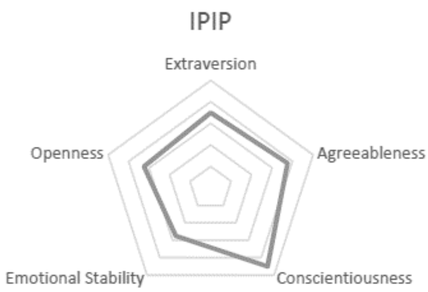
A Study on Planning and Personality

Individual Report

Scores below only represents your personality preferences derived from your answers to psychometric scales based on the Trait Theory. It should **not be interepreted as an classification or assessment tool as it does not measure your skill or ability in any area**. Individual preferences are highly subjective and the purpose of such assessment is to identify your particular style. There is no better or worse result and no trait is more desirable than other.

International Personality Item Pool

The International Personality Item Pool (IPIP) was designed to assess the constellation of traits defined by the Five Factor Theory of Personality (Gosling et al., 2003).



Extraversion is characterized by talkativeness, assertiveness, and energy.

Agreeableness is characterized by good-naturedness, cooperativeness, and trust.

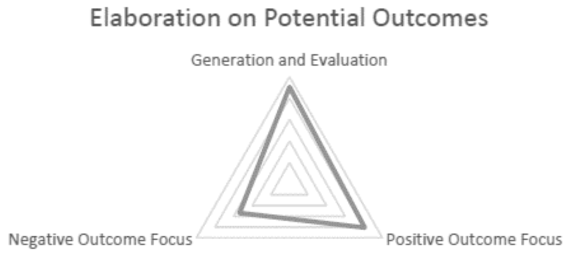
Conscientiousness is characterized by orderliness, responsibility, and dependability.

Emotional stability is characterised by how difficult is to upset the individual

Openness is characterized by originality, curiosity, and ingenuity.

Elaboration on Potential Outcomes Scale

EPO was designed to assess individuals' tendencies to generate and evaluate possible positive and negative consequences of their behavior (Nenkov et al., 2007)



Generation / evaluation of potential outcomes (General) assesses consideration and focus on potential outcomes in general

Positive outcome focus (Positive) assesses focus on potential positive outcomes

Negative outcome focus (Negative) assesses focus on potential negative outcomes

Barratt Impulsiveness Scale

The Barratt Impulsiveness Scale (BIS-11; Patton et al., 1995) is a questionnaire designed to assess the personality/behavioral construct of impulsiveness.

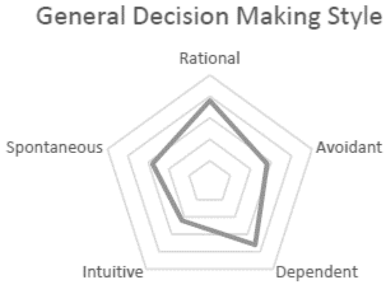
Attentional Impulsiveness assesses task-focus, intrusive thoughts, and racing thoughts.

Motor Impulsiveness assesses tendency to act on the spur of the moment and consistency of lifestyle.

Non-planning Impulsiveness assesses careful thinking and planning and enjoyment of challenging mental tasks.

General Decision Making Style Scale

The GDMS was designed to assess how individuals approach decision situations. It distinguishes between 5 decision styles. (Scott and Bruce, 1995)



A **rational** style emphasizes "a thorough search for and logical evaluation of alternatives."

An **avoidant** style emphasizes postponing and avoiding decisions.

A **dependent** style emphasizes "a search for advice and direction from others."

An **intuitive** style emphasizes "a reliance on hunches and feelings."

A **spontaneous** style emphasizes "a sense of immediacy and a desire to get through the decision-making process as soon as possible."

Further reading on personality metrics

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. Journal of Research in Personality, 37(6), 504-528

Nenkov, G. Y., Inman, J. J., & Hulland, J. (2007). Considering the future: The conceptualization and measurement of elaboration on potential outcomes. Journal of Consumer Research, 35, 126-141.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt Impulsiveness Scale. Journal of Clinical Psychology, 51(6), 768-774.

Scott, S. G., & Bruce, R. A. (1995). Decision-making style: The development and assessment of a new measure. Educational and Psychological Measurement, 55(5), 818-831

Contact	Contact Type	E-mail
Mr. Alexander Kharlamov	Principal Investigator	a.a.kharlamov@warwick.ac.uk
Prof. Dr. Janet Godsell	Supervisor	j.godsell@warwick.ac.uk
Dr. Ganna Pogrebna	Supervisor	g.pogrebna@warwick.ac.uk

Appendix L Summary statistics of individual differences by group and by treatment (normalised values)

Variable	Treatment	Naïve Students mean value	Sophisticated students mean value	Sophisticated non-students mean value	Total
Naïve interventionism	Short	0.49	0.34	0.37	0.42
	Long	0.07	0.06	0.10	0.08
	Hybrid	0.42	0.22	0.34	0.35
	Total	0.33	0.22	0.24	0.28
Big 5 Extroversion	Short	0.58	0.55	0.61	0.58
	Long	0.64	0.64	0.64	0.64
	Hybrid	0.61	0.62	0.67	0.63
	Total	0.61	0.60	0.64	0.62
Big 5 Agreeableness	Short	0.72	0.73	0.71	0.73
	Long	0.76	0.72	0.78	0.77
	Hybrid	0.77	0.76	0.81	0.77
	Total	0.75	0.74	0.77	0.76
Big 5 Conscientiousness	Short	0.71	0.71	0.74	0.71
	Long	0.69	0.65	0.77	0.72
	Hybrid	0.67	0.68	0.73	0.68
	Total	0.69	0.68	0.75	0.70
Big 5 Neuroticism	Short	0.57	0.64	0.57	0.58
	Long	0.57	0.64	0.50	0.55
	Hybrid	0.62	0.58	0.52	0.58
	Total	0.59	0.61	0.52	0.57
Big 5 Intellect/Imagination	Short	0.73	0.70	0.65	0.71
	Long	0.75	0.82	0.72	0.75
	Hybrid	0.74	0.71	0.78	0.74
	Total	0.74	0.73	0.72	0.73
EPO Generation & Evaluation	Short	0.75	0.66	0.76	0.74
	Long	0.74	0.75	0.80	0.76
	Hybrid	0.75	0.69	0.77	0.75
	Total	0.74	0.69	0.78	0.75
EPO Positive Outcome Focus	Short	0.63	0.61	0.67	0.63
	Long	0.67	0.74	0.65	0.68
	Hybrid	0.67	0.72	0.69	0.69
	Total	0.66	0.69	0.67	0.66

EPO Negative Outcome Focus	Short	0.68	0.60	0.55	0.63
	Long	0.63	0.59	0.54	0.58
	Hybrid	0.68	0.55	0.53	0.62
	Total	0.67	0.58	0.54	0.61
BIS	Short	0.56	0.57	0.52	0.54
	Long	0.54	0.53	0.49	0.52
	Hybrid	0.54	0.56	0.53	0.54
	Total	0.54	0.56	0.51	0.54
GDMS Rational	Short	0.75	0.74	0.79	0.76
	Long	0.76	0.68	0.82	0.77
	Hybrid	0.76	0.73	0.76	0.75
	Total	0.76	0.72	0.79	0.76
GDMS Intuitive	Short	0.70	0.69	0.64	0.68
	Long	0.70	0.73	0.66	0.69
	Hybrid	0.70	0.68	0.71	0.70
	Total	0.70	0.70	0.67	0.69
GDMS Dependent	Short	0.71	0.67	0.66	0.70
	Long	0.72	0.58	0.71	0.71
	Hybrid	0.74	0.68	0.71	0.72
	Total	0.73	0.65	0.70	0.71
GDMS Avoidant	Short	0.64	0.58	0.55	0.60
	Long	0.63	0.53	0.46	0.56
	Hybrid	0.62	0.59	0.53	0.59
	Total	0.63	0.57	0.50	0.58
GDMS Spontaneous	Short	0.57	0.60	0.53	0.56
	Long	0.57	0.59	0.53	0.56
	Hybrid	0.58	0.60	0.59	0.59
	Total	0.58	0.60	0.55	0.57

Appendix M Stability analysis of personality scales

Scale	Cronbach alpha	Cronbach alpha squared	Measurement error
Big 5	0.80	0.64	0.36
BIS	0.85	0.72	0.28
GDMS	0.77	0.59	0.41
EPO	0.71	0.50	0.50

Note: EPO Cronbach alpha is the lowest in part because EPO is the smallest scale with only three constructs

Appendix N Correlation matrix of Personality Variables

Variable		NI	5E	5A	5C	5N	5I	GE	PF	NF	BIS	GR	GI	GD	GA
Big 5 Extroversion (5E)	coeff	-0.1080*	1												
	p	0.0476													
Big 5 Agreeableness (5A)	coeff	0.0558	0.2115*	1											
	p	0.3072	0.0001												
Big 5 Conscientiousness (5C)	coeff	0.0254	-0.0225	0.0484	1										
	p	0.6423	0.681	0.3759											
Big 5 Neuroticism (5N)	coeff	0.1261*	-0.1180*	0.0975	-0.1647*	1									
	p	0.0206	0.0303	0.0738	0.0024										
Big 5 Intellect/Imagination (5I)	coeff	-0.0494	0.1712*	0.2036*	0.0317	0.024	1								
	p	0.3655	0.0016	0.0002	0.5618	0.6603									
EPO Generation & Evaluation (GE)	coeff	0.0993	0.0008	0.0647	0.1959*	-0.0226	0.1158*	1							
	p	0.0694	0.9881	0.2376	0.0003	0.6804	0.0341								
EPO Positive Outcome Focus (PF)	coeff	0.0002	0.1422*	0.0279	0.1039	-0.3784*	0.0889	0.043	1						
	p	0.9966	0.0092	0.6105	0.0575	0.0000	0.1044	0.4324							
EPO Negative Outcome Focus (NF)	coeff	0.1244*	-0.1514*	0.0124	-0.1124*	0.4682*	-0.0707	0.2229*	-0.5288*	1					

	BIS		GDMS Rational		GDMS Intuitive		GDMS Dependent		GDMS Avoidant		GDMS Spontaneous	
p	coeff	p	coeff	p	coeff	p	coeff	p	coeff	p	coeff	p
0.0227	0.0161	0.7691	0.0287	0.6004	0.0478	0.3829	0.013	0.8129	0.015	0.7851	0.0579	0.2905
0.0055	0.0736	0.1791	0.1143*	0.0366	-0.0117	0.8311	0.0675	0.2182	0.0553	0.3133	-0.0339	0.5369
0.8212	-0.0257	0.6397	-0.1380*	0.0114	-0.0849	0.121	-0.0914	0.0948	0.0301	0.5827	-0.1479*	0.0067
0.0397	-0.2988*	0	-0.2804*	0	-0.3781*	0	-0.3312*	0	-0.1850*	0.0007	-0.2463*	0
0	0.2875*	0	-0.0037	0.9455	0.1538*	0.0048	0.1624*	0.0029	0.2941*	0	0.0827	0.1309
0.1968	0.0041	0.9403	-0.0302	0.582	-0.1845*	0.0007	-0.0693	0.2056	0.0527	0.336	-0.0993	0.0694
0	-0.0804	0.1421	-0.2486*	0	-0.3713*	0	-0.1893*	0.0005	0.0257	0.6387	-0.1553*	0.0044
0	-0.2087*	0.0001	0.0682	0.2132	-0.1362*	0.0126	-0.1094*	0.0453	-0.2190*	0.0001	-0.077	0.1596
	0.2817*	0	0.0063	0.9084	0.1081*	0.048	0.1793*	0.001	0.2717*	0	0.0571	0.2975
	1		0.5430*	0	0.2948*	0	0.7526*	0	0.8720*	0	0.4403*	0
			1		0.3447*	0	0.4566*	0	0.4396*	0	0.8131*	0
					1		0.4001*	0	0.1271*	0.0199	0.3199*	0
							1		0.3353*	0	0.3647*	0
									1		0.3623*	0

* significant at least at 5% level; p=0 in the table refers to p<0.0001