5-2021

Debiasing a decision maker facing supply uncertainties in a newsvendor setting

Alan William ZELLER
Singapore Management University

DEBIASING A DECISION MAKER FACING SUPPLY UNCERTAINTIES IN A NEWSVENDOR SETTING

ALAN WILLIAM ZELLER

SINGAPORE MANAGEMENT UNIVERSITY

2021

Debiasing a decision maker facing supply uncertainties in a newsvendor setting

By

Alan William Zeller

Submitted to School of Business in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Business (General Management)

<u>Dissertation Committee:</u>

Pascale Crama (Chair)

Associate Professor of Operations Management

Singapore Management University

Shantanu Bhattacharya

Professor of Operations Management

Singapore Management University

Niyazi Taneri

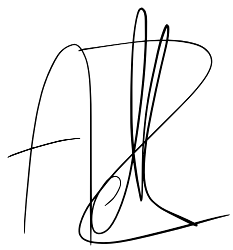
Senior Lecturer in Operations & Technology Management

Cambridge Judge Business School

Singapore Management University
2021

I hereby declare that this PhD dissertation is my original work, and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in this dissertation.

This PhD dissertation has also not been submitted for any degree in any university previously.



Alan William Zeller

17 May 2021

Debiasing a decision maker facing supply uncertainties

in a newsvendor setting

Alan William Zeller

ABSTRACT

Companies must be prepared to manage uncontrollable events that will disrupt their supply chain and add uncertainty to their inventory models. This thesis first studies the effect of different types of supply disruption risks on the ordering performance of profit-maximizing decision makers in a newsvendor setting. Then, this thesis aims at extending the literature on the newsvendor model in studying the effect of a Decision Support System and the effect of a Secondary Task on the ordering performance of profit-maximizing decision makers who face supply uncertainties in a newsvendor setting. Finally, implications for scholars and practitioners are discussed.

TABLE OF CONTENTS

ACKNOWLEDGEMENT

I would like to thank my dissertation supervisor Dr Pascale Crama and my committee members Dr Shantanu Bhattacharya and Dr Niyazi Taneri for their guidance and support throughout my studies and for providing me the feedback that allowed me to sharpen my thinking and aim higher. Their expertise was invaluable at each step of the dissertation and I have learned tremendously by their side.

DEDICATION

This work is dedicated to my wife Albena who supported me unconditionally throughout this journey and to the professors who inspired me since I am a child.

1. Statement of the research problem

Supply chains become more and more complex (Gurnani et al., 2014) and more and more dependent to events that companies cannot control (Simchi-Levi, 2015). In this context, companies must be prepared to manage and handle events that will inevitably disrupt their supply chain.

Perrow (1984) coined the term "Normal Accidents" in 1984 to describe those events that are inevitable and that will occur regularly due to the complexity of balancing the supply chain. More recently, Taleb (2007) introduced the term "Black Swans" to describe those catastrophic events that are extremely rare and hard to predict. Several Black Swans such as the trade-war between the USA and China or the COVID-19 pandemic have recently disrupted heavily the supply chains and in turn impacted significantly the profitability of companies. Indeed, according to the McKinsey Global Institute (2020, p. 12), "Supply-Chain-Disruption losses equal forty two percent of one year's earnings before interest, taxes, depreciation and amortization on average over a decade". As a consequence, supply chain managers must consider carefully supply uncertainties when making ordering decisions.

Supply uncertainties add complexity and subjectivity to inventory models and can lead to large variances and deviations from normative values as decision makers will apprehend supply uncertainties differently according to their goals and their preferences. In situations where the inventory is to be used in a single selling period and cannot be carried over to the next selling period, such as for perishable products, the newsvendor model is an inventory model widely used by practitioners. The newsvendor model is also one of the most used models in operations management in the fields of inventory management, sourcing and pricing. In its traditional form, it assumes that the supply is deterministic and that decision makers are rational. While the model has a well-known profit maximizing order quantity, its users generally exhibit long lasting biases and place sub-optimal orders.

The Assumption in the traditional newsvendor model that the supply is deterministic makes the model less relevant in our VUCA (Bennis & Nanus, 1985) world. Several authors have therefore changed this assumption and added supply uncertainties in the traditional newsvendor model. They have observed that the addition of supply uncertainties has led to worse decision making and to an amplification of the biases observed in the traditional newsvendor model.

In this context, it has become more and more important to understand the psychological process which creates the biases in the newsvendor model and to find nudges and strategies aiming at debiasing the decision makers facing supply uncertainty in a newsvendor setting.

Decision Support Systems are computerized programs that synthetize data and provide insightful reports to its users. They are used by companies to help their employees make better decisions. “Qlikview”, “SAP Business Object” and “Salesforce Analytics Could” are examples of Decision Support Systems with a large installed base that are used by companies across industries and functions.

This thesis first aims at understanding better how a Decision Support System can improve the ordering performance of a decision maker facing supply uncertainties in a newsvendor setting. Then, this research focuses on the cognitive load associated with the Decision Support System and aims at understanding better how it can affect the biases traditionally observed in a newsvendor setting.

Those findings will benefit the practitioners who face newsvendor problems while making ordering, pricing or sourcing decisions in a business environment that is more and more competitive, stressful and uncertain.

2. Literature review

The literature review has three sections. In the first section, I cover the literature on the newsvendor model, starting from its traditional normative form and continuing with its behavioral form. I then complement the review in covering articles on the newsvendor model with supply uncertainties. In the second section, I start with the literature on the classical Cognitive Load Theory and I subsequently narrow down the review to cover specifically the limitations of the working memory and the impact that a primary task can have on secondary task. In the third and final section, I review the conditions under which behavioral operations management experiments can be conducted on a web platform with online workers.

2.1. The newsvendor model

The newsvendor model is one of the most important models in operations management (Becker-Peth et al., 2018). It has been introduced in 1888 when Edgeworth (1888) studied the inventory policy for bank notes withdrawn by clients at a bank. The setting of the newsvendor model is simple: a decision-maker must place an order for a product that will be sold during a single selling period with stochastic demand. In this model, the decision maker faces a trade-off: order too little and incur missed sales, or order too much and incur left-over inventory at the end of the period.

The profit maximizing quantity, or “nominal order quantity’ is well known (Arrow et al., 1951): it balances the expected cost of ordering too much inventory and the expected cost of not ordering enough inventory. One example of the newsvendor model in real life is the newsboy (Morse & Kimball, 1951) who sells the daily newspaper in the street and who gave his/her name to the model. The newsboy must decide how many newspapers to buy from the printing company before he/she starts his/her milk run. If the newsboy does not buy enough newspaper, some clients will be disappointed because the newspaper will be out-of-stock. However, if the newsboy buys too many newspaper, he/she will incur the loss of the left-over inventory since the newspaper will be outdated and will have no or little value at the end of the day.

The traditional newsvendor model has been extended in the literature in the 1970s-1980s (Chen et al., 2016) and its extensions can be classified into eleven categories (Khouja, 1999). For example, Jucker and Rosenblatt (1985) extended the model to include supplier pricing and quantity discounts, and Lau and Lau (1988) extended the model to include price dependent demand distribution. More recently, Ray and Jenamani (2016) introduced a newsvendor setting were both supply and demand are uncertain and proposed two models for optimal order allocation.

While the traditional newsvendor model assumes that the decision maker is rational and seeks to improve the profit generated, it is possible that the newsboy would display a behavior that is different from the profit maximizing one, and that could be triggered by unconscious factors, such as the desire to please more clients.

Schweitzer and Cachon (2000) were the first to consider non-rational decision makers in a newsvendor setting. In their experiment, they asked participant to place order decisions during 15 periods for a low (or high) margin product followed by 15 periods for a high (or low) margin product. The price for both products was the same at 12 dollars per unit, but the low margin product cost 9 dollars per unit and the high margin product cost 3 dollars per unit. The demand for both products was uniformly distributed between 1 and 300 units. In this case, the nominal order quantity for the low margin product was 75 units, the nominal order quantity for the high margin product was 225 units, and the average demand was expected to be 150 units. In this setting, they noticed that the participants had a tendency to order between the mean and the nominal order quantity in both low and high margin conditions and they introduced the concept of "pull to center" effect. They concluded their paper in considering possible biases that could explain their findings: (i) Participants were anchoring on the average demand and would not adjust sufficiently; (ii) Participants were chasing the demand of the previous period; and (iii) participants were minimizing ex-post inventory error. Later research by Benzion et al. (2008) found that the pull-to-center effect is robust to the demand distribution and similarly exists for normally distributed demand.

The seminal article from Schweitzer and Cachon (2000) has led to further research in the field of Behavioral Operations Management. To study the pull-to-center effect at the subject pool level, Feng et al. (2011) repeated the experiment of Bolton and Katok (2008) with Chinese and American participants and found that the Chinese participants displayed a stronger pull to center effect. However, in their experiment, Ozer et al. (2014) found no significant difference in behavior between Chinese and American participants. Bolton et al. (2012) compared the ordering performance of students and professional buyers and found no significant difference either, with both students and professional buyers displaying the pull-to-center effect. Bolton et al. (2012) also introduced task training in the form of a 60 minutes video that explains in details the nominal order calculations and the fact that people wrongly chase the mean and the demand of the previous period. The training improved performance of both the students and the professional buyers but did not eliminate the pull-to-center effect. Taken together, those findings clearly demonstrate that the pull-to-center effect is sticky and very hard to correct.

In an effort to alleviate the pull-to-center effect, Bolton et al. (2012) conducted an experiment on the newsvendor model in which they provided some participants with a Decision Support System computing the expected profit according to the quantity ordered. They found that participants still displayed the pull-to-center effect, even when they were given the optimal solution.

Decision Support Systems are "computer based systems intended to help decision makers utilize data and models to identify and solve problems" (Rauscher, 1999). Ford (1985) explained that Decision Support Systems reduce the cognitive cost of collecting more information and are useful for managers that need to make difficult decisions under uncertainty. He also explained that for Decision Support Systems to be effective, their interface needs to be user friendly they need to adjust dynamically when the parameters change. Gonul et al. (2006) observed that explaining the rationale of the Decision Support System to the participants increases its acceptance and reduces algorithm aversion (Dietvorst et al., 2015). Ye and Johnson (1995) gave examples of rationale such as explaining the line of reasoning or explaining the overall strategy.

More recently, Lee and Siemsen (2017) found that the introduction of a Decision Support System was effective in improving performance in their newsvendor setting but that the decision makers were still displaying the traditional newsvendor biases.

In an attempt to explain the pull-to-center effect, Bolton and Katok (2008) and Lau et al. (2014) observed that individual orders displayed large variations that extended beyond the optimal order quantity and the mean and they proposed that the pull-to-center effect should be studied at the subject level.

In this spirit, Croson et al. (2008) studied the effect of overconfidence on the pull to center effect and found that overconfident decision makers placed sub-optimal orders. Ren and Croson (2012) built on this study and found that over-precise people who “believe that their estimate are more accurate than they truly are” (Moore & Healy, 2008) could underestimate the variance of the demand and are more likely display the pull-to-center effect. Moritz et al. (2013) studied the individual differences in ordering decision between intuitive and cognitive thinkers. They categorized the participants using the Cognitive Reflection Test of Frederick (2005) and found that cognitive thinkers chase the demand less and are less sensitive to the pull to center effect in the high margin condition only. Simulating the payoffs of the newsvendor model in a lottery setting, Kremer et al. (2010) concluded that demand chasing can be significant at individual level due to the psychological cost associated with leftovers and stock-outs.

Demand and mean chasing are the most common explanation given in the literature to explain the pull-to-center effect. To measure the extent to which people chase demand and adjust their previous order according to the actual prior demand, several indicators have been developed and compared in the literature:

Schweitzer and Cachon (2000) introduced a "Change Frequency" indicator that measures how often decision makers adjust their order quantity towards actual prior demand. Using this indicator in a newsvendor setting, they observed that decision makers change their order quantity more often towards the prior actual demand than away from the prior demand.

In the same article, to reinforce the validity of their finding, Schweitzer and Cachon (2000) introduced a different method to measure demand chasing and developed the "adjustment score" δ defined as follows:

$$\delta\mathrm{t} = \frac{q(t) - q(t-1)}{d(t-1) - q(t-1)}$$

The adjustment score is based on the assumption that decision makers start with the previous demand and adjust towards the nominal order quantity such as if they do not adjust at all, their order quantity will be equal to the previous demand and such as if they adjust fully, their order quantity will be equal to the optimal order quantity.

In a first step, two adjustment scores are computed, one adjustment score when the adjustment is towards the previous demand, and one adjustment score when the adjustment is away from the previous demand. In a second step, the 2 adjustment scores are compared. If the adjustment score towards the previous demand is superior to the adjustment score away from the previous demand, then the presence of a demand chasing behavior is determined. In their experiment, Schweitzer and Cachon (2000) observed a demand chasing behavior using this methodology in their high margin setting but not in their low margin setting. Bostian et al. (2008) have used linear regressions to derive the adjustment factor and observed that participants adjusted more towards the nominal order quantity over the 30 periods of their experiment.

Bolton and Katok (2008) measured demand chasing using a simple correlation between the current order and the previous demand. They found that thirty percent of decision makers are demand chasers in the low margin newsvendor setting and ten percent of decision makers are demand chasers in the high margin newsvendor setting.

To compare the validity of those demand chasing metrics, Lau and Bearden (2013) have simulated order patterns that exhibit chasing demand behavior and have concluded that the correlation analysis should be chosen to measure demand chasing because it has the "lowest false positive rate and an acceptable false negative rate"

To measure how much decision makers adjust from the mean, Bolton et al. (2012) have used a regression and developed a mean chasing anchoring factor (α). It is based on the assumption that decision makers start with the mean (μ) and adjust toward the nominal order quantity (q^*). If they don't adjust at all, then their order (q) would be equal to the mean (μ). And if they adjust fully, then their order (q) would be equal to the nominal order quantity (q^*).

The ordering quantity can therefore be represented as:

$$q = \mu + (1-\alpha)(q^* - \mu)$$

and the anchoring factor can be represented as:

$$\alpha = 1 - \frac{q-\mu}{q^* - \mu}$$

This measurement of how much decision makers anchor to the mean has the advantage to be simple. In their experiment, using this indicator, Bolton et al. (2012) found that the anchoring effect depends on the subject pool.

Researchers in the Behavioral Operations Management field have studied several strategies to reduce the pull-to-center effect. They realized the significant impact such a successful strategy would have in many of the models used by practitioners. A commonly studied question revolves around the ability of decision makers to learn and improve their decision making over time as they make the decisions. During the 15 periods of their experiment, Schweitzer and Cachon (2000) did not observe any improvement of the decision makers as (q^*-q) did not reduce significantly over time. Bolton and Katok (2008) ran their experiment over 100 periods and observed a small learning over time. However, the learning induced was insufficient to eliminate the pull to center effect.

Lurie and Swaminathan (2009) studied the effect of feedback frequencies on the ordering decision. They observed that less frequent feedback improved the ordering in a high demand variability setting and concluded that frequent feedback can actually decrease the ordering performance because the decision makers were attaching too much importance on the data of the previous period. Bolton and Katok (2008) also concluded that demand chasing behavior can be corrected when participants are asked to place orders for multiple periods at once.

The traditional newsvendor model assumes that the decision makers want to maximize the expected profit. Schweitzer and Cachon (2000) simulated many other individual goals and demonstrated that the pull to center effect is still observed when the decision makers are driven by those other individual goals.

Ruling out individual goals to explain the pull to center effect, some authors have focused their attention on individual preferences.

In a recent study, Becker-Perth et al. (2018) have studied the effect of individual risk preferences on the ordering performance. Using the scale of Holt and Laury (2002) to measure the risk aversion of the participants, they observed that risk averse decision makers ordered significantly less than the risk seeking ones and found a significant correlation between risk preferences and order quantities. In another recent study, Schultz et al. (2018) framed the setting of their experiment to display gains and losses and they did not observe any significant effect of individual loss preferences on the ordering performance.

The baseline assumption in the newsvendor model is that the lead time is deterministic and that the products ordered are available (Serel, 2008). Gallego and Moon (1993) were the first to introduce supply uncertainty in the normative model. In their experiment, a unit from a supplier's production was proper or improper for use according to a binomial random variable yield.

The supply uncertainty has been modelized in five different ways in the literature to reflect different situations, ranging from an operational risk where only a fraction of the order is received to a disruption risk where the order is not received at all (Yano & Lee, 1995):

1. The number of good units provided in a batch has a binomial distribution and each unit has an independent probability to be good or bad. In this

simple model, the fraction of good units is independent from the order size.

2. The number of good units provided in a batch is a fraction of the order size. In this simple model, the mean, the variance and the type of distribution can be adjusted, but the distribution is independent from the batch size.
3. The time until the supply becomes unreliable is randomly distributed.
4. The probability that the supply becomes unreliable increases or decreases with time
5. The probability that the supply becomes unreliable increases or decreases according to the batch size.

According to Yano and Lee (1995), the second model is appropriate for situations where the production system cannot be adapted easily to changes in the environment and for situations where the yield loss might be predictable for any particular set of conditions, but where the conditions themselves are not predictable.

In a single supplier setting where the capacity is uncertain, the newsvendor should not order less than the quantity it would have ordered if the supply was deterministic (Dada et al., 2007). Moreover, if the order quantity does not affect the capability of the supplier, the newsvendor should not modify its decision from the base case in which the supply is deterministic (Dada et al., 2007; Ciarallo et al., 1994). However, if the order quantity affects the capability of the supplier, the newsvendor should order more (Dada et al., 2007; Henig & Gerchak, 1990).

Swaminathan and Shanthikumar (1999) have introduced a setting with 2 suppliers that deliver all or nothing according to a probability and have provided example to show that under certain conditions, ordering from the more expensive and reliable supplier is optimal.

The model has been further complexified by the introduction of multiple suppliers with different yield uncertainties. In his experiment, Giri (2011) proposed a choice between a reliable but expensive supplier and a cheaper but unreliable alternative that would deliver only a fraction of the order according to a yield and showed that the dual sourcing strategy depends on the risk profile of the newsvendor. Merzifonluoglu and Feng (2014) developed a setting with multiple suppliers, including some unreliable with normally distributed random yields, and challenged the idea that when faced with multiple unreliable suppliers, "cost is the order qualifier and reliability is the order winner".

The existing literature on the newsvendor model shows that the pull-to-center is a robust effect at the collective level, that the ordering performance is deteriorated by complexity and that the Decision Support Systems have a limited impact on the ordering performance.

Kaki et al. (2015) introduced a treatment with an operational supply uncertainty in a behavioral setting. They found that participants struggled to process this additional parameter and placed orders that deviated even more from the optimal order quantity. The mental effort required to handle the newsvendor model with supply uncertainty drained the cognitive capacity of the participants and had a significant impact on their ordering performance. In this context, a review of the Cognitive Load Theory and the limitation of the working memory is required to find methodologies and strategies that improve the performance of the newsvendor model.

2.2. Cognitive Load Theory

The classical Cognitive Load Theory (Sweller, 1988) builds on the findings of memory research to improve learning, focusing on “the cognitive processes that occur during interactions between working memory and long term memory” (Ayres & Paas, 2012). It is based on the assumption that working memory resources are finite and information cannot be stored in long-term memory if it is not processed first by working memory.

De Groot (1965) and Chase and Simon (1973) have established the role of long term memory in learning and problem solving in studying chess Grand-Masters. The authors concluded that the superiority of Grand-Masters at playing chess did not come from their ability to compute more possible moves, but from their ability to memorize and replicate patterns from previous parties that can be applied to the game they are playing. Chi et al. (1981) coined the term schema to define a pattern or framework that has been learned and that can be replicated without effort to solve a problem. The acquisition of a new schema requires effort but once assimilated, a schema can be used automatically without effort (Kotovsky et al., 1985). Reading is a good example of a schema: it takes effort when we first learn how to read but reading becomes automatic with practice. Competence can therefore be achieved through the assimilation of domain specific schemas, which makes long term memory essential for problem solving. Similar to other mental shortcuts, schemas are very useful for thinking fast in general, but they can also be used wrongly and lead to systematic errors in problem solving in some cases (Sweller & Gee, 1978). The newsvendor model is a schema that requires a substantial effort to assimilate. Competence on the newsvendor model comes only after a deep understanding of its underlying principles, with experience and repetitive feedback, and with an active attention to detect and override the biases that it generates.

For a novel piece of information to be stored in the long-term memory, a number of cognitive processes must happen (Atkinson & Shiffrin, 1968). First, the information must be captured by sensory systems. Penney (1989) has identified that auditive and visual pieces of information are processed independently. Then, the novel piece of information must be processed by the working memory, recalling information from long-term memory and consolidating it with the new piece of information to create a new piece of knowledge (Baddeley, 1992). Finally, this new piece of knowledge must be processed to be stored in the long-term memory (Sweller, 1988).

The working memory has limited storage capacity. Miller (1956) had first established that it can only recall a maximum of seven items of novel information before Cowan (2001) reduced this number even further to four items of novel information. Not only the working memory has limited storage capacity, but it is also limited in duration. Peterson and Peterson (1959) mentioned that working memory can recall a new piece of information for a few seconds only before it is generally forgotten after twenty seconds.

Sweller et al. (1998) studied the different types of cognitive load that are affecting the working memory. They have concluded that cognitive load can be generated by the level of difficulty of a task (intrinsic cognitive load), by the way the information is presented to the learner (extraneous cognitive load) and by the effort generated to transform the newly acquired information into knowledge and to store it in the long-term memory (germane cognitive load). By definition, the intrinsic cognitive load can therefore only be reduced by the act of learning itself.

To measure the cognitive load on working memory, several methodologies have been developed in the literature. Owen and Sweller (1985) and Sweller and Cooper (1985) used the error rate of a task as a proxy to measure cognitive load while Chander and Sweller (1991) used the time spent on the task as a proxy. Moving away from indirect measurements, Paas (1992) developed a nine points Likert scale asking participants to rate their perceived amount of mental effort from very very low, to very very high. This subjective measurement was found to be reliable (Paas et al., 1994) and simple to administer without interfering with the task at hand (Sweller et al., 2011). However, Van Gog and Pass (2008) mentioned that too many variations of the scale have been used in the literature making comparisons difficult and Kirschner et al. (2006) mentioned that the scale has been used to measure different types of cognitive load.

Brunken et al. (2003) used a dual-task methodology to measure cognitive load and concluded that it is superior to the subjective method. The underlying assumption of the dual-task methodology is that as the primary task becomes more difficult, the secondary task performance reduces if the same cognitive resources are used. Similarly, as the secondary task becomes more difficult, the primary task performance reduces if the same cognitive resources are used.

Chandler and Sweller (1991), Halford et al. (1986) and Ayres (2001) used a problem solving secondary task and they found that it generated a high level of cognitive load on working memory. Paas (1992) compared the cognitive load generated by different ways of presenting a problem to the participants and concluded that the traditional way of presenting the problem and asking for the solution without guiding the participants to work-out the solution generated the most cognitive load.

The intrinsic cognitive load generated by the newsvendor model is high, and so is the germane cognitive load required to assimilate it. The extraneous cognitive load level depends on the way the model is presented to the participants, but in its simple form, it still requires a few steps and explanations. Overall, the newsvendor model generates a high level of cognitive load that saturates the cognitive capacity of its users that have not achieved competency on the model and that can therefore not beneficiate from the input of their long-term memory. The limitations of the working memory prevent the participants from processing the information and trigger the use of heuristics such as chasing the demand and chasing the mean.

The vast majority of experiments on the newsvendor model and on the cognitive load theory have been conducted in a laboratory setting. In the context of limited gathering options during the COVID-19 pandemic, the use of a web platform recruiting online workers is considered. The following literature review shows the conditions under which behavioral operations experiments can be conducted online and deliver equivalent validity of results than laboratory experiments.

2.3. Online experiments

Several classical studies in the field of behavioral economics and decision-making using students in a laboratory have been replicated using online workers on web platforms. Paolacci et al. (2010) replicated the Asian Disease Problem (Tversky & Kahneman, 1981), the Linda Problem (Tversky & Kahneman, 1983) and the Physician Problem (Baron & Hershey, 1988) on Amazon Turk and found no significant differences between the original findings and the findings in the replicated studies.

Lee and al. (2018) replicated an inventory management problem (Bolton & Katok, 2008), a procurement auction problem (Engelbrecht-Wiggans & Katok, 2008) and a supply chain contracting problem (Loch & Wu, 2008) on Amazon Turk to assess whether online workers on web platform can be used in a behavioral operations management context. They found no significant difference between the original findings and the findings in the replicated studies but noticed that the learning was slower on web platforms.

Several authors in the field of Behavioral Operations Management have used online workers and web platforms to run their experiments such as Dixon et al. (2017), Hutchison-Krupat and Chao (2014), Lau et al. (2014) and de Véricourt et al. (2013).

Online platforms offer an inexpensive access to a large subject pool that is more representative of the population than students. They allow for fast recruitment and increase the speed at which experiments can be run. Online platforms are now in a competitive market and each platform tries to differentiate itself in offering and developing features that allow researchers to select participants according to some criteria, pay bonuses, withhold payments or simply block online workers that have already participated in another treatment.

Rand (2012) compared the IP address of the participants and their declared country and found, similar to Farrell et al. (2017), that respondents answer demographics questions honestly. Goodman and al. (2013) concluded that online workers on web platforms can be trusted. Chandler et al. (2014) criticized the validity of using online workers on web platforms as they are unsupervised, anonymous and less attentive but Hauser and Schwartz (2016) reconciled both views in studying the characteristics of the respondents. They found that high reputation participants rarely fail attention checks and recommended to select participants with high previous ratings. Lee et al. (2018) added that it is possible to mitigate the risk of having inattentive respondents in asking questions that check that the respondents have understood the task at hand and was paying attention.

59% of Indian workers and 69% of US workers agree that “Amazon Turk is a fruitful way to spend some free time and get some cash” (Ipeirotis, 2010) and less than 8% reported making more than USD 50 / week on the site. Goodman et al. (2013) observed that participants will try to find answers even if there is no incentive to do so.

Yin et al. (2013) defined a Performance Contingent Reward as a reward for a task that will depend on quality, the quality being measured by key performance indicators of interest for the researcher. They found that a Performance Contingent Reward on online platforms does not affect the quality of the response nor the effort of the participant.

Mason and Watts (2009) observed similar behaviors and concluded that increased financial incentive will increase the quantity but not the quality of the work. They added that when intrinsic motivation is possible, non-paid work will be as good as paid, and that when intrinsic motivation is not possible, then the researcher's interest is to offer as little as possible when the subject pool is large enough to find participants. In other words, paying more would increase the pool of participants but not the quality of the answers.

3. Research hypotheses

3.1. Hypothesis 1

Supply disruptions are "unforeseen events that interfere with the normal flow of materials and/or goods within the supply chain" (Craighead et al., 2007). The overall supply disruption risk depends on the magnitude and on the probability of the supply disruption and is defined as "an individual's perception of the total potential loss associated with the disruption of supply of a particular purchased item from a particular supplier" (Ellis et al., 2010, p. 36).

A supply disruption risk that happens occasionally with a moderate magnitude is a normal accident (Perrow, 2011). Normal accidents are set to happen in the context of supply chains that become more and more complex (Gurnani et al., 2014). Supply chain managers prepare for the normal accidents and build buffers to absorb the variances in supply. By contrast, a supply disruption risk that happens rarely with a high magnitude is a black swan (Taleb, 2007). Black swans come as a surprise and supply chain managers struggle to prepare for them (Mitroff & Alpaslan, 2003). In this context, supply chain managers perceive the total potential loss associated with a black swan to be more important than the total potential loss associated with a normal accident and as a consequence, black swans represent a higher overall supply disruption risk than normal accidents do.

Kaki et al. (2015) introduced a newsvendor setting with a normal supply disruption that was set to happen at every selling period with a random magnitude. In this setting, compared to a traditional model with deterministic supply, participants deteriorated their ordering performance and mentioned risk to explain the rationale of their ordering decisions. In a setting with a higher level of overall supply disruption risk, the ordering performance will deteriorate even more, and the gap between the average order quantity and the optimal order quantity will widen.

Hypothesis 1a. Compared to a setting with a deterministic supply, the ordering performance of newsvendors in a normal accident supply disruption risk setting deteriorates, widening the gap between their average order quantity and the optimum order quantity.

Hypothesis 1b. Compared to a setting with a deterministic supply, the ordering performance of newsvendors in a black swan supply disruption risk setting deteriorates, widening the gap between their average order quantity and the optimum order quantity.

Hypothesis 1c. Compared to a normal accident supply disruption risk setting, the ordering performance of newsvendors in a black swan supply disruption risk setting deteriorates, widening the gap between their average order quantity and the optimum order quantity.

3.2. <u>Hypothesis 2</u>

Decision Support Systems are "computer based systems intended to help decision makers utilize data and models to identify and solve problems" (Rauscher, 1999). Lee and Siemsen (2017) introduced a Decision Support System in the form of suggested quantities in the traditional newsvendor model and found that it can improve the effectiveness of the ordering decision if the demand uncertainty is not too high and if the cost structure does not emphasize the cost of not ordering enough inventory. Bolton et al. (2012) introduced a Decision Support Systems in the form of feedback in a traditional newsvendor model and found that it improved the ordering performance effectively. While the Decision Support System that suggested quantities was prescriptive and while the Decision Support System that gave feedback was not prescriptive, they both provided insight on the order quantity, whose optimization was the task at hand.

Kaki et al. (2015) observed that ordering decisions are much harder to make when a supply uncertainty is introduced in the traditional newsvendor model. In the newsvendor model with stochastic supply, the potential impact of the disruption is an important additional information that is required to compute the optimal order quantity. Ford (1985) explained that Decision Support Systems reduce the cognitive cost of collecting more information and are especially useful for managers that need to make difficult decisions under uncertainty. A Decision Support System that informs on the potential impact of the disruption will therefore help the newsvendors optimize the order quantity, even if such a

Decision Support System does not provide insight on the order quantity when the risk does not materialize.

> ***Hypothesis 2. The introduction of a Decision Support System that provides insight on the risk but not on the order quantity improves the ordering performance of a decision maker facing supply uncertainties in a newsvendor setting.***

3.3. Hypothesis 3

The newsvendor model with supply uncertainty generates a high level of cognitive load (Käki et al., 2015a). Its assimilation requires a substantial mental effort and does not eliminate the need to go through a complex multiple steps process to compute the optimum order quantity.

The working memory has limited storage capacity (Miller, 1956 ; Cowan, 2001) and is limited in duration (Peterson & Peterson, 1959). Therefore, the mental effort required to solve the model saturates the working memory and triggers the use of mental shortcuts (Tversky & Kahneman, 1973). In the newsvendor model, the mean and the prior demand are salient pieces of information that are easily recalled and that anchor the decision maker in absence of active adjustment, explaining the pull-to-center effect (Schweitzer & Cachon, 2000b).

The introduction of a Secondary Task that calls on short term memory before the ordering decision is made will replace the information stored in the working memory by the Primary Task (Computing the order quantity) after the feedback from the previous period is known. This will artificially reduce the recalling and salience of the mean and of the prior demand and will lead to a reduction of the pull-to-center effect.

Hypothesis 3: The introduction of a secondary task that calls on short term memory, after the feedback from the previous period is known and before the ordering decision for the period is made, will reduce the pull-to-center effect and will improve the ordering performance of a decision maker facing supply uncertainties in a newsvendor setting.

4. Methodology

To test the different hypotheses, different groups with different treatments are compared to a control group in a traditional newsvendor setting according to Table 1.

Table 1

Different settings of the experiment

High/Low Margin (HM/LM)	**No Supply Risk (NR)**	**Normal Accident (NA)**	**Black Swan (BS)**
No Decision Support System **nor** Secondary Task **(NT)**	Control Group HM-NR-NT LM-NR-NT	Group with NA Disruption Risk HM-NA-NT LM-NA-NT	Group with BS Disruption Risk HM-BS-NT LM-BS-NT
With Secondary Task **(ST)**	N/A	Group with NA Disruption Risk and Secondary Task HM-NA-ST LM-NA-ST	Group with BS Disruption Risk and Secondary Task HM-BS-ST LM-BS-ST
With Decision Support System **(DSS)**	N/A	Group with NA Disruption Risk and Decision Support System HM-NA-DSS LM-NA-DSS	Group with BS Disruption Risk and Decision Support System HM-BS-DSS LM-BS-DSS

4.1. Control Group

The control group setting replicates a traditional newsvendor setting with a random linear demand between 0 and 500 over 8 periods. The selling price is set at 150 dollars per unit and the unsold units have no salvage value. In the low margin condition, the purchase cost is set at 100 dollars per unit, so that the profit for each unit sold is 50 dollars and the nominal order quantity is 167 units. In the high margin condition, the purchase cost is set at 50 dollars per unit, so that the profit for each unit sold is 100 dollars and the nominal order quantity is 333 units.

Those values are selected to simplify the calculation of the profit made for each unit sold and to create a sufficient gap between the nominal order quantity and the average expected demand. The actual demand is generated randomly once and is used in all the treatments.

The following instructional text is given to the participants:

> *You are in the business of buying and selling fresh organic apples.*
>
> *You buy the apples weekly from an orchard for 50/100 dollars per crate and sell them for 150 dollars per crate, which means you will earn 100/50 dollars per crate sold.*

The expected demand for organic apples is 250 crates per week and the actual demand is uniformly distributed between 0 and 500.

If you receive more apples than the demand, you will have some leftovers at the end of the week that you will have to discard. If you receive less apples than demand, you will incur missed sales.

The high season for organic apples lasts 8 weeks. For each of those 8 weeks, you have to order apples from the orchard, who delivers them instantly.

Your goal is to maximize your profit during those 8 weeks.

The instructional text, similar to the one used by Käki et al. (2015), makes the profit for each unit sold explicit and explains the trade-off between "missed-sales" and "leftover inventory" faced by the participants.

The duration of 8 weeks is chosen to allow for sufficient rounds of decision making before and after a potential supply shortage, while minimizing the cost of the experiments.

After each ordering decision, and before making the ordering decision for the next round, similar to the experiment of Käki et al. (2015), the participants are given a feedback including:

- The quantity ordered
- The quantity received
- The actual demand
- The profit generated in a summary table and in an explanatory text format.

4.2. Black Swan Supply Disruption Risk Treatment

In the black swan supply disruption risk treatment, a risk of supply disruption is added to the conditions of the control group. The disruption is set to happen with a probability of 1% and with a significant impact on the business since only 30% of the goods ordered are delivered in case of disruption. The rarity of the disruption and the severity of its consequences make the disruption an abnormal event, but an event that can happen at some point during the course of the business.

In this treatment, the instructional text is modified to reflect the inclusion of the black swan supply disruption risk and includes the following sentence between the last two paragraphs:

> *There is a 1% probability that the orchard can't fulfill your order in full because of bad weather conditions. When this happens, you receive only 30% of your order.*

The addition of this sentence is the only difference with the instructional text given to the participants in the control group of the experiment. The nominal order quantity for this setting is 167 units for the low margin condition and 334 units for the high margin condition.

4.3. Normal Accident Supply Disruption Risk Treatment (NA)

In the normal accident supply disruption risk treatment, the supply shortage is set to happen with a probability of 10% and with a moderate impact on the business since 90% of the goods ordered are delivered when the supply shortage occurs. The frequency of the risk and the effect of its consequences make the supply shortage a normal accident (Perrow, 1984), an event that is likely happen regularly during the course of the business.

In this treatment, the instructional text is modified to reflect the inclusion of the normal accident disruption risk and includes the following sentence between the last two paragraphs:

> *There is a 10% probability that the orchard can't fulfill your order in full because of bad weather conditions. When this happens, you receive only 90% of your order.*

Similar to in the black swan setting, the addition of this sentence is the only difference with the instructional text given to the participants in the control group of the experiment. The nominal order quantity for this setting is 168 units in the low margin condition and 336 units in the high margin condition.

4.4. Decision Support System Treatment (DSS)

To assess the risk of events that may disrupt the supply chain, risk managers traditionally assign a probability and a severity to those events. This process leads to a theoretical relative ranking of identified risks. However, in practice, managers face challenges to assess high impact - low probability risks (Mitroff & Alpaslan, 2003).

Acknowledging those challenges and limitations in the risk assessment phase and noting that risk mitigation activities are often the same regardless of the cause of the disruption, Simchi-Levy et al. (2015) have developed a Decision Support System that analyses the impact of a supply chain supply risk regardless of its cause but according to its impact. To measure the impact, the authors have developed a Performance Impact (PI) indicator, which measures the cost the company will incur if it recovers optimally from the disruption using available alternatives.

As a Performance Impact indicator is relevant and used by practitioners in the context of supply chain disruptions (Simchi-Levi, 2015), it is given to the participants in this treatment as a Decision Support System at the end of each period and before they have to make the ordering decision for the following period. The Performance Impact displays the expected loss in profit when the orchard faces bad weather and can only deliver a fraction of the previous week's order. The Decision Support System therefore provides the participants with an additional piece of information that helps them make a better decision, but that does not give them the answer to the task at hand.

The rationale of the Decision Support System is given to the participants (Gönül et al., 2006) and the Performance Impact is updated every week (Ye & Johnson, 1995) to increase the acceptance of the Decision Support System.

The Decision Support System is displayed as followed:

> *To help you decide how many apple crates you should order, a Performance Impact value is given to you each week.*
>
> *The Performance Impact is your expected loss in profit when the orchard faces bad weather and can only deliver 30/70% of the order*
>
> *The Performance impact for your Week X order of Y apple crates is: Z dollars*

4.5. <u>Secondary Task Treatment (ST)</u>

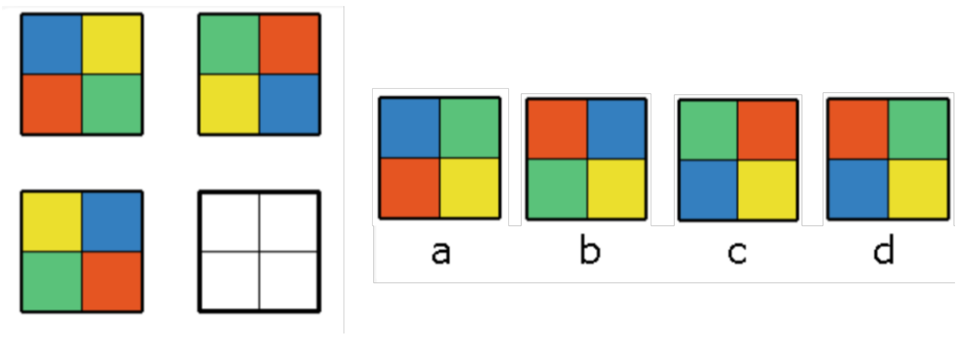
In the Secondary Task treatment, a secondary task is introduced to reduce the recalling of the mean and of the previous week's actual demand. To be effective, the Secondary Task must deplete the cognitive load of the participants.

To this effect, a secondary task is designed with the following 3 characteristics:

1. The secondary task must generate a low intrinsic cognitive load so that all participants can complete it successfully and rapidly. It must therefore be easy to complete.
2. The secondary task must generate a low extraneous cognitive load so that participants can focus on the task at hand without spending mental load in figuring out what needs to be done.
3. The secondary task must generate high germane cognitive load and deplete the working memory of the participants.

The secondary task is presented to the participants after the feedback on the current period has been given and before the ordering decision for the following period must be made.

In this treatment, the secondary task is displayed as follows:



Solve the puzzle, enter the correct letter […], and click next

This brainteaser is used in elementary schools to develop the working memory of the students as it requires them to remember the position of a few colors before they are in a position to find the solution.

At each period, the colors of the squares are shuffled so that the mental effort required to solve the puzzle remains constant over the eight periods of the experiment.

4.6. Post Survey questions

4.6.1. Post-Survey Question 1: Attention and Comprehension check

According to Lee et al. (2018), the risk of having non-attentive participants can be mitigated by asking attention and comprehension check questions. To this effect, the following post-survey 1 questions are developed:

- In the No Tool and in the Secondary Task treatments, a simple multiple-choice question asking about the cost of the left-over inventory is given to the participants. It is expected that participants who answer correctly this question have understood the task at hand and were paying attention to the experiment since the cost of the left-over inventory was not given in the instructions and since it is needed to solve the problem.

  The Post-Survey 1 question for those two treatments reads as follows:

What is your cost per crate of apples that are left over?

1) 0 Dollars

2) 50 Dollars

3) 100 Dollars

4) 150 Dollars

4 possible answers are given to reduce the number of false-positive answers. The correct answer is 50 dollars per crate in the high margin condition and 100 dollars per crate in the low margin condition.

- In the Decision Support System treatment, a simple multiple-choice question asking about the purpose of the decision tool is given to the participants in order to eliminate the respondents who have not understood the tool or paid attention to it. Four possible answers are also given to the participants for consistency. The Post-Survey 1 question for the Decision Support System treatment reads as follows:

The "Performance Impact" value given by the decision tool is:

1) The expected cost of the leftover inventory

2) The expected value of the missed sales

3) The expected loss in profit when the orchard faces bad weather and can only deliver the order partially

4) The expected profit

The answer to this question is actually given to the participants in the text of the Decision Support System at each period. It is therefore expected that participants who answers the question correctly have paid attention to the experiment and understood the task at hand.

4.6.2. Post-Survey Question 2: Risk profile of the participants

Individual preferences such as risk aversion are often cited in the literature to explain the differences in ordering performance. As different levels of supply risk are introduced in the treatments, the participants could react to them differently according to their risk profile. It is therefore relevant to measure the risk aversion of the participants and to study whether it moderates the effect of the Decision Support System and/or the effect of the Secondary Task.

To measure the risk aversion of the participants in their experiment, Holt and Laury (2002) have developed a scale that has already been used by several other authors in the behavioral operational management field (Becker-Peth, Thonemann, & Gully, 2018; Tuncel et al., 2019). In this scale, the participants must make ten choices between two lotteries, one safer than the other. As the payoff for the riskier lottery increases, the participants should eventually switch to it and the time it takes for the participants to switch will be the measurement of their risk aversion. A participant who chooses the safest lottery less than 4 times will be risk seeking, a participant who chose the safest lottery 4 times will be risk neutral, and a participant who chooses the safest lottery more than 4 times will be risk averse.

The format of the Post-Survey question 2 is as follows:

You are presented with 2 options - please select your preferred option (A or B) for each row

For example, the cell in purple means you have 10% chance of receiving 2 dollars and 90% chance of receiving 1.6 dollars

A	B	Choice "A" or "B"
1/10 of \$2 , 9/10 of \$1.6	1/10 of \$3.85 , 9/10 of \$0.1	
2/10 of \$2 , 8/10 of \$1.6	2/10 of \$3.85 , 8/10 of \$0.1	
3/10 of \$2 , 7/10 of \$1.6	3/10 of \$3.85 , 7/10 of \$0.1	
4/10 of \$2 , 6/10 of \$1.6	4/10 of \$3.85 , 6/10 of \$0.1	
5/10 of \$2 , 5/10 of \$1.6	5/10 of \$3.85 , 5/10 of \$0.1	
6/10 of \$2 , 4/10 of \$1.6	6/10 of \$3.85 , 4/10 of \$0.1	
7/10 of \$2 , 3/10 of \$1.6	7/10 of \$3.85 , 3/10 of \$0.1	
8/10 of \$2 , 2/10 of \$1.6	8/10 of \$3.85 , 2/10 of \$0.1	
9/10 of \$2 , 1/10 of \$1.6	9/10 of \$3.85 , 1/10 of \$0.1	
10/10 of \$2 , 0/10 of \$1.6	10/10 of \$3.85 , 0/10 of \$0.1	

4.7. Format of the experiment

The participants of the 14 experiments are online workers recruited on the Prolific web platform (www.prolific.co). The Prolific platform is chosen for its large pool of participants, its ability to select participants according to their level of education and according to their previous passing rate, and for its feature that allows researchers to block participants to enroll in a study if they have already participated in another one.

The minimum requirements to be eligible to enroll in the experiment are:

- An undergraduate university level
- A high previous passing rate
- No participation in another treatment of the experiment
- Access to a computer with Microsoft Excel installed

The experiments are run over several weeks on Saturday morning Singapore Time, which corresponds to Friday evening time in the USA. Due to the time difference, it is expected that participants from Europe or Africa are not very representative in the study. However, this time zone allows for a large number of respondents in Asia, Pacific, and Americas to participate in the experiment.

50 participants are hired for each treatment. 700 different participants are therefore hired to complete the experiment. As Prolific has a pool of more than

50,000 workers that meet the minimum requirements of the study, no shortage of participants is experienced.

The experiment takes 12-15 minutes to complete, and each worker is paid a flat fee upon submission of his/her answer according to the remuneration policy of Prolific. The flat fee option is preferred to a success fee option as the success fee option can eliminate a category of participants that are risk averse. Moreover, since the demand is randomly generated over 8 periods, it is possible that the participants making the nominal ordering decision do not generate the best profit.

To ensure a high validity of the answers and to mitigate the risk of having un-attentive workers, the responses of the participants who fail to answer the post-survey question 1 successfully are discarded.

A supply shortage is introduced in period 5 for all participants that are not in the baseline setting for comparison of the effect of the different treatments before and after the supply shortage actually happens. In the Normal Accident treatment, the probability for the supply shortage to happen is 10%. Over an 8-weeks period, a supply shortage should therefore statistically happen 67% of the time and the participants should not be deceived when the supply disruption is introduced.

In the Black Swan treatment (BS), the probability for the supply shortage to happen is 1%. Over an 8-weeks period, a supply shortage should therefore statistically happen 7.7% of the time. While participants could expect the supply shortage not to happen over the 8 weeks period, there is definitely a chance that it could happen. Viewed at the participant level, they can conclude that they are unlucky if the disruption happens, but they cannot conclude that they are being deceived. Since the participants cannot communicate with one another and since the participation to the study is anonymous, the participants cannot access to a subject pool view on the probability of the actual disruption.

5. Analysis

5.1. Data description and validation

In this section, the aggregate average Order Quantity observed in the different settings is analysed. The pull-to-centre effect is observed in the low margin condition in all settings but in none of the settings in the high margin condition. Some theories supported by the literature are provided to explain those observations. Replicating the methodology used by Lau et al. (2014), the average demand is also analysed at subject level in this section.

5.1.1. Low Margin. No Risk. No Tool.

The setting with No Risk and No Tool is the baseline setting. In this section, the findings observed in the low margin baseline setting are analysed and compared with the findings in the paper of Schweitzer and Cachon (2000).

Figure 1 shows that at aggregate level, the mean order quantity (208) is in the pull-to-centre zone, between the expected average demand (250) and the optimum order quantity (167), consistent with the observation of Schweitzer and Cachon (2000).

Figure 1

LM.NR.NT - Frequency, Mean and Standard deviation

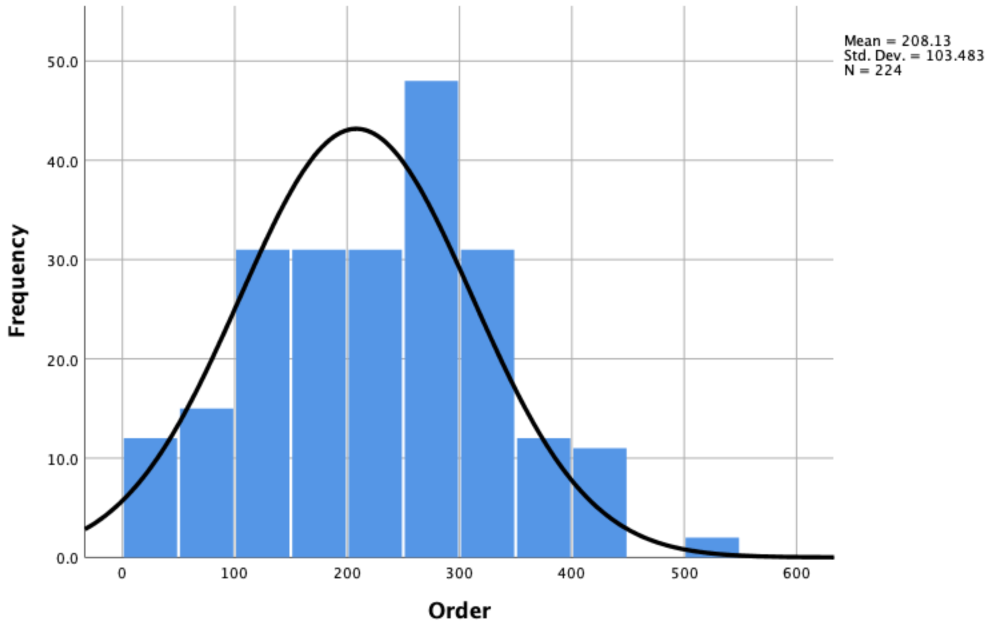
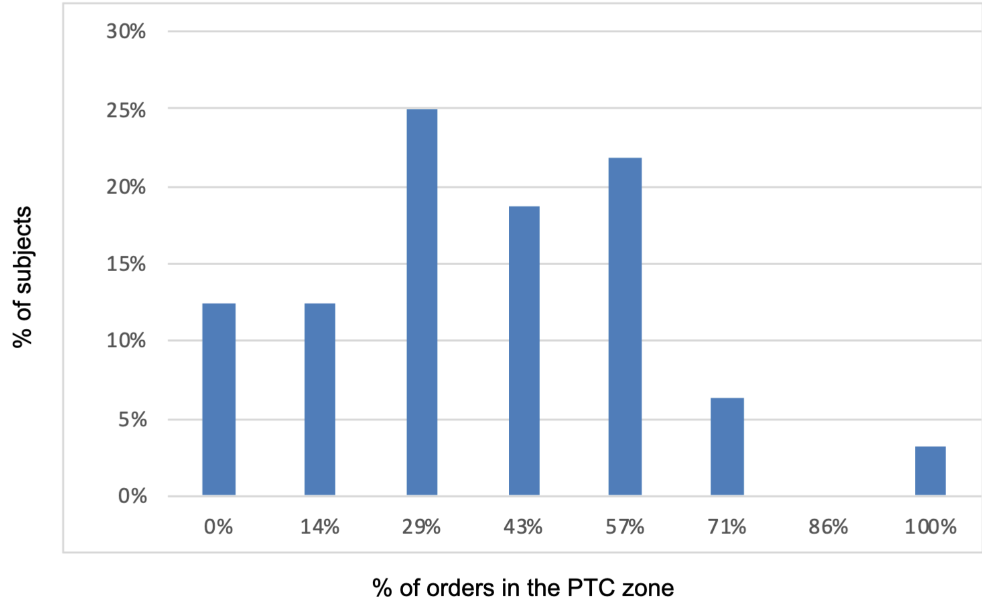


Figure 2 shows the histogram of subject by their percentage of orders in the pull-to-centre zone, replicating the Figure 2 in the paper of Lau et al. (2014). According to the authors, "if the pull-to-centre is present in individual ordering, then all the mass should be on the right of each graph". Using the same definition as the authors that "pull-to-centre effect applies to an individual when 50% or more of their orders are in the pull-to-centre zone", then only 31% of the subjects of this experiment can be said to exhibit the pull to centre effect.

Figure 2

LM.NR.NT - Percentage of orders in the pull-to-centre zone



The percentage of subjects said to display the pull-to-centre effect (31%) is consistent with the number of participants said to display the pull-to-centre effect (35%) in the low margin experiment of Bolton and Katok (2008).

Figure 3 and Figure 4 show the cumulative distribution of subject's average order quantities using the mean and the median, replicating the figure 3 in the paper of Lau et al. (2014). Vertical lines on the plot indicate the boundaries of the pull-to-centre zone between the expected average order and the optimum order quantity. According to the authors, compared to the previous definition in Figure 2, "a looser definition of the pull-to-centre would be that an individual's average order quantity lies in the pull-to-centre zone". However, 56% of participants in the LM.NR.NT experiment have mean orders outside of the pull-to-centre zone and 44% of participants have median orders outside of the pull-to-centre zone. Those percentages represent participants who do not display the pull-to-centre effect according to the looser definition of Lau et al. (2014). Those findings are consistent with the findings of Lau et al. (2014) who found that 30% of subjects have a mean order outside the pull-to-centre zone and 55% of subjects have a median order outside the pull-to-centre zone in the low margin experiment of Bolton and Katok (2008).

Figure 3

LM.NR.NT - Mean orders

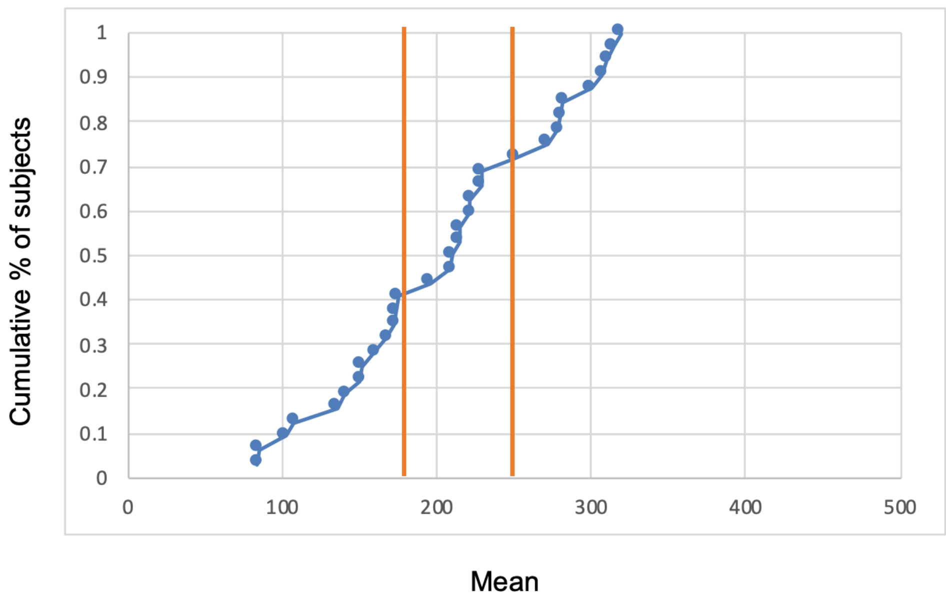
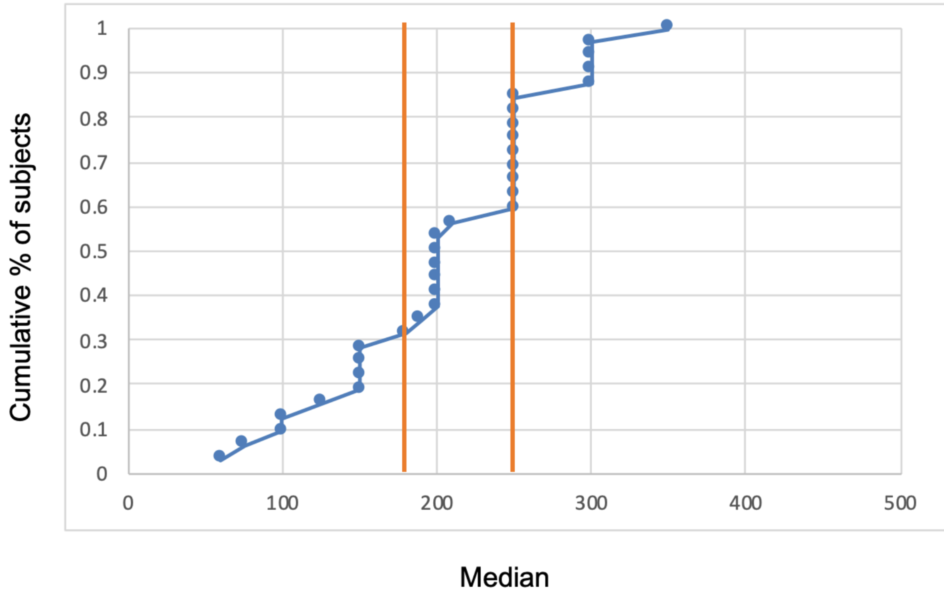


Figure 4

LM.NR.NT - Median orders



The study of data at subject level in the low margin condition shows that there is heterogeneity in the newsvendor's behaviour and that conclusions about individual behaviours cannot be drawn from aggregate data. Nevertheless, at aggregate level, the pull-to-centre effect is a very stable observed phenomenon (Zhang & Siemsen, 2019).

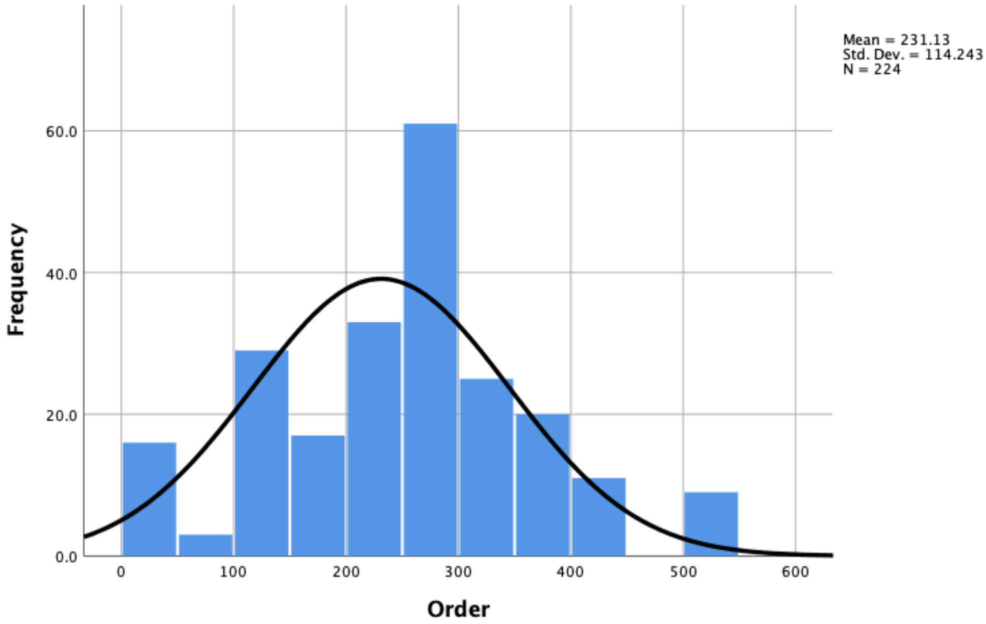
5.1.2. <u>High Margin. No Risk. No Tool.</u>

In this section, the findings observed in the high margin baseline setting are analysed.

Figure 5 shows that at aggregate level, the mean order quantity (231) is below the pull-to-center zone which is defined as the interval between the expected average demand (250) and the optimum order quantity (366). The mean order quantity is in fact further away from the optimum quantity than what is traditionally observed in the newsvendor model (Schweitzer & Cachon, 2000a).

Figure 5

HM.NR.NT - Frequency, Mean and Standard deviation



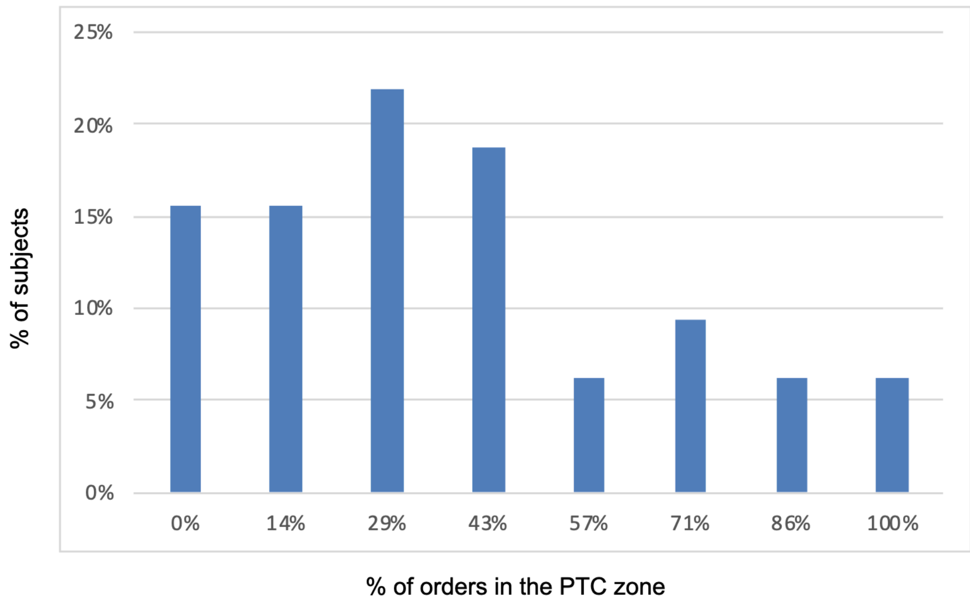
Zhang and Siemsen (2019) have demonstrated that there is an asymmetry between the low and high margin conditions in the newsvendor model and that "framing the newsvendor problem using a procurement cost and sales price emphasizes overage cost, leading to a stronger Pull-to-centre tendency in high-margin condition". The design of the experiment does indeed frame the newsvendor problem using a procurement cost and a sale price as the first phrase of the instruction given to the participants starts with: "You buy the apples weekly from an orchard for 50 dollars per crate and you sell them for 150 dollars per crate"

Ho et al. (2010) have also observed that the asymmetry between low and high margin conditions depends on the salience of the underage cost. Zhang and Siemsen (2019) added that “newsvendor experimental designs which do not effectively communicate the loss of underage to participants will likely see a stronger pull-to-centre effect in high-margin conditions”. In the experiment, the instruction given to the participants reads “if you receive less apples than demand, you will incur missed sales” and does not explicitly communicate the loss of underage. As a consequence, the average order quantity is likely to be further away from the optimum order quantity in the high margin condition than in the low margin condition.

Similar to Figure 2 in the LM.NR.NT setting, Figure 6 shows the histogram of subject by their percentage of orders in the pull-to-centre zone.

Figure 6

HM.NR.NT - Percentage of orders in the pull-to-centre zone



According to the pull-to-centre definition of Lau et al. (2014), 28% of participants can be said to exhibit the pull-to-centre affect. This is consistent with their finding that 33% of the participants can be said to exhibit the pull-to-centre effect in the high margin experiment of Bolton and Katok (2008).

Similar to Figure 3 and Figure 4 in the LM.NR.NT experiment, Figure 7 and Figure 8 show the cumulative distribution of subject's average order quantities using the mean and the median. According to the looser pull-to-centre definition of Lau et al. (2014), 66% of the participants do not display the pull-to-centre effect because their mean individual order is outside of the pull-to-centre zone and 56% of the participants do not display the pull-to-centre if their median orders are considered instead of their mean orders. Those findings are consistent with the findings of Lau et al. (2014) who found that 45% of participants have mean orders outside the pull-to-centre zone and 39% of participants have median orders outside the pull-to-centre zone.

Figure 7

HM.NR.NT - Mean orders

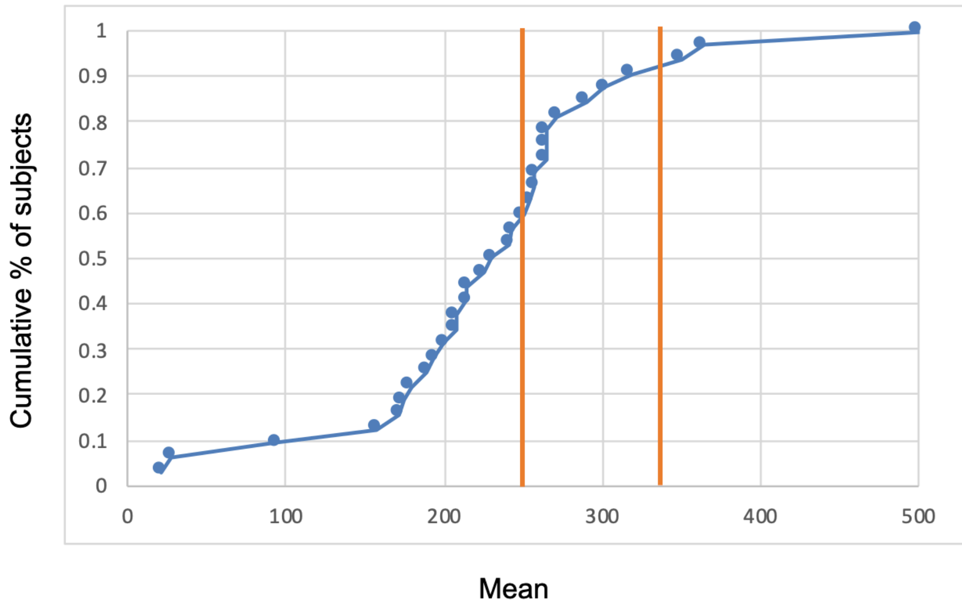
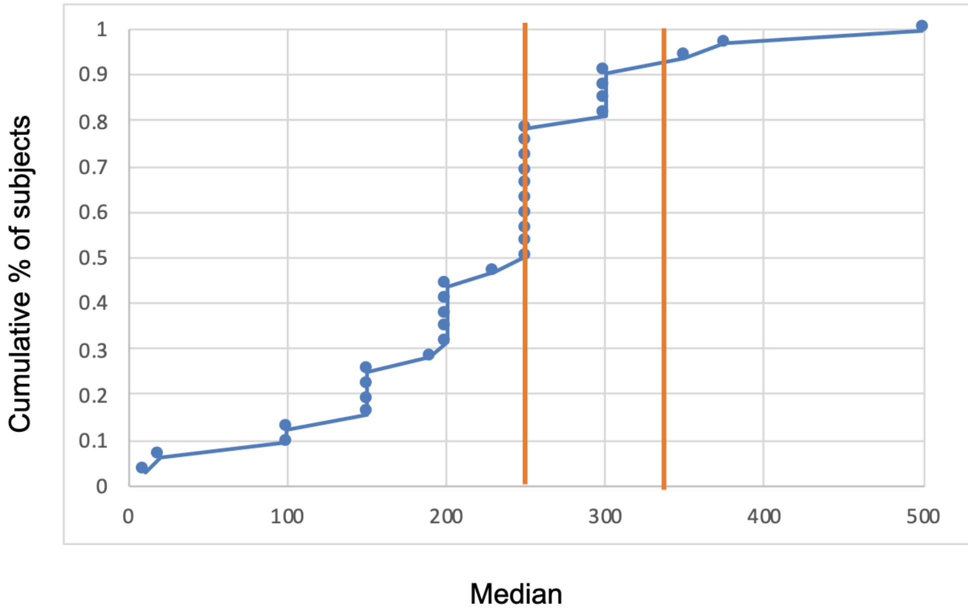


Figure 8

HM.NR.NT - Median orders



The study of data at subject level in the high margin condition supports the findings in the low margin condition and confirms that conclusions about individual behaviours cannot be drawn from aggregate data.

5.1.3. Other settings

In the low margin condition, the mean order is inside the pull-to-centre zone in all settings, consistent with the observation of Schweitzer and Cachon (2000). However, in all settings in the high margin condition, the mean order is below the pull-to-center zone, further away from the optimum quantity. This may be due to the design of the experiment as explained in chapter 5.1.2.

Table 2

Mean Orders

	No Risk	Normal Accident	Black Swan
NT	LM Mean: 208.13* (Figure 1) HM Mean: 231.13 (Figure 5)	LM Mean: 242.98* (Figure 9) HM Mean: 219.23 (Figure 10)	LM Mean: 237.15* (Figure 11) HM Mean: 214.13 (Figure 12)
ST		LM Mean: 231.18* (Figure 17) HM Mean: 238.86 (Figure 18)	LM Mean: 217.47* (Figure 19) HM Mean: 227.53 (Figure 20)
DSS		LM Mean: 238.4* (Figure 13) HM Mean: 220.58 (Figure 14)	LM Mean: 222.74* (Figure 15) HM Mean: 226.97 (Figure 16)

Note. * inside the pull-to-centre zone

5.2. **Hypothesis 1 Analysis (No Tool)**

In this section of the analysis, the effect of a Normal Accident and a Black Swan supply disruption risks on the Order Quantity is analyzed in the low and high margin conditions.

5.2.1. Correlations tables

Table 3 shows the correlations in the low margin setting between:

- The dependent variable "Order Quantity".
- The independent variables "Normal Accident" and "Black Swan" that returns 1 if the participants are in the "Normal Accident" and "Black Swan" setting respectively, and that returns 0 otherwise.
- The independent variable "D-1" that represents the demand of the previous period.
- The control variables "Risk Averse", Risk Neutral" and "Risk Seeking" that return 1 when the participants are "Risk Averse", "Risk Neutral" and "Risk Seeking" respectively as determined by their answer to the second post-survey question. The participants are either "Risk Averse", "Risk Neutral" or "Risk Seeking".
- The control variable "Over-order" that returns 1 if the order of the previous period is superior to the demand of the previous period and that returns 0 otherwise.

The Order Quantity is significantly correlated to the presence of a Normal Accident (0.097) and to the demand of the previous period (0.27), the latter implying that the participants display a demand chasing behavior. The demand of the previous week is also correlated to the presence of over-ordering in the previous week (-.767), which is logical since the lower the demand of the previous week, the more likely the ordering quantity will be higher than the demand of the previous week. The variables Normal Accident and Black Swan are correlated (0.122 and -0.1 respectively) to Risk Aversion.

Table 3

Correlations – Low Margin

	Mean	SD	Order	Norm. Accid.	Black Swan	Risk Av.	Risk Neutr.	Risk Seek.	D-1	Over-order
Order	229.7	103.30	1							
Normal Accident	0.36	0.481	.097*	1						
Black Swan	0.31	0.463	.048	-.50**	1					
Risk Averse	0.07	0.259	-.015	.122**	-.10**	1				
Risk Neutral	0.18	0.38	.046	.049	-.015	-.12**	1			
Risk Seeking	0.73	0.443	.004	-.078*	.052	-.46**	-.76**	1		
D - 1	206.1	142.99	.27**	.000	.000	.000	.000	.000	1	
Over-order	0.49	0.5	.021	.040	.004	-.021	.031	-.009	-.77**	1

Note: N = 679.

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4 shows the same correlations as Table 3 but in the high margin setting. The Order Quantity is significantly correlated to the demand of the previous period (0.215), which also implies a demand chasing behavior of the participants. Similar to in the low margin setting, the demand of the previous week is also correlated to the presence of over-ordering the previous week (-.763). No correlation between the Order Quantity and the presence of Normal Accident or Black Swan is observed. The variables Normal Accident and Black Swan are significantly correlated to Risk Aversion (-.192 and .116 respectively)

Table 4

Correlations – High Margin

	Mean	SD	Order	Norm. Accid.	Black Swan	Risk Av.	Risk Neutr.	Risk See.	D-1	Over-order
Order	221.87	110.30	1							
Norm. Accident	0.33	0.472	-.017	1						
Black Swan	0.31	0.463	-.047	-.47**	1					
Risk Averse	0.12	0.328	.027	-.19**	.12**	1				
Risk Neutral	0.18	0.383	.059	.29**	-.18**	-.17**	1			
Risk Seeking	0.69	0.463	-.069	-.085*	.037	-.55**	-.69**	1		
Demand -1	206.14	142.99	.21**	.000	.000	.000	.000	.000	1	
Over-order	0.48	0.5	.062	-.007	-.009	.019	.018	-.026	-.76**	1

Note. N = 630

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

5.2.2. No Risk vs Normal Accident

In this section of the analysis, the effect of the Normal Accident supply disruption risk on the order performance is analysed, compared to a setting with No Risk (Hypothesis 1a)

Table 5 shows the result of a linear regression analysis in the low and high margin settings with the Order Quantity as the dependant variable and with the demand of the previous week, the over-order variable and the risk profile of the participants as independent variables.

Table 5

Regression analysis using Order Quantity as a criterion (NR vs NA)

	Low Margin	High Margin
(Constant)	67.172***	76.820***
	(16.262)	(19.064)
Normal Accident	29.786***	-14.018
	(8.878)	(10.028)
Risk Averse	-18.176	-9.268
	(15.656)	(16.671)
Risk Neutral	-9.115	12.441
	(11.593)	(11.912)
Demand-1	.456***	.491***
	(.047)	(.0.54)
Over-order	108.60***	107.805***
	(13.390)	(15.359)

*: significant at 0.1, **: significant at 0.05; ***: significant at 0.01

Table 5 shows that the inclusion of a Normal Accident supply disruption risk in the low margin setting increased significantly the mean order quantity, away from the optimum quantity. This finding is consistent with the findings of Kaki et al. (2015) who found that the inclusion of a supply disruption risk increased significantly the mean order quantity in the low margin condition and deteriorated the ordering performance. Additionally, the control variable D-1 has a significant effect on the order quantity: the higher the demand of the previous period, the higher the order quantity of the current period.

Table 6 shows the results of a Wilcoxon signed-ranked test comparing the Order Quantities under the No Risk and the Normal Accident settings in the low margin setting. The test shows that the Order Quantities in the two different settings have a different distribution.

Table 6

Ranks

		N	Mean Rank	Sum of Ranks
Normal Accident – No Risk	Negative Ranks	72[a]	85.78	6176.50
	Positive Ranks	128[b]	108.78	13923.50
	Ties	24[c]		
	Total	224		

Note. a. NA < NR; b. NA > NR; c. NA = NR
Z = -4.732 based on negative ranks; Asym. Sig. (2 tailed) = 0.000

Those findings show that compared to a setting with a deterministic supply, decision makers in a Normal Accident supply disruption risk setting deteriorate their ordering performance, widening the gap between their average quantity and the optimum order quantity and therefore they support Hypothesis 1a in the low margin setting.

In a high margin setting, Table 5 shows that the inclusion of a Normal Accident supply disruption risk decreased the mean order quantity, away from the optimum quantity, but that the decrease is not statistically significant.

However, the average order quantity in the No Risk setting is already below the pull-to-centre zone and while the introduction of a Normal Accident deteriorates even further the ordering performance, it may not have a significant effect due to the fact that the order quantity in the No Risk setting is already below the pull-to-centre zone.

Table 5 also shows that, similar to in the low margin setting, the control variable D-1 has a significant effect on the Order Quantity. The control variable "Over-order" has also a significant effect on the Order Quantity and shows that over-ordering in the previous period significantly increases the Order Quantity of the current period. Finally, this analysis shows that the risk profile of the participant has no significant impact on the Order Quantity.

Table 7 shows the results of a Wilcoxon signed-ranked test comparing the Order Quantities under the high margin Normal Accident and the No Risk settings. The test shows that the Order Quantities in the two different settings do not have a different distribution.

Table 7

Ranks

		N	Mean Rank	Sum of Ranks
Normal Accident – No Risk	Negative Ranks	101[a]	89.55	9044.50
	Positive Ranks	84[b]	97.15	8160.50
	Ties	25[c]		
	Total	210		

Note. a. NA < NR; b. NA > NR; c. NA = NR
Z = -0.607 based on positive ranks; Asym. Sig. (2 tailed) = 0.544

The results of the regression analysis and the Wilcoxon signed-rank test do not support Hypothesis 1a in the high margin setting.

5.2.3. <u>No Risk vs Black Swan</u>

In this section of the analysis, the effect of a Black Swan supply disruption risk on the ordering performance is analysed in both the low and high margin settings (hypothesis 1b).

Table 8 shows the result of a linear regression analysis in the low and high margin settings with the Order Quantity as the dependant variable and with the demand of the previous week, the over-order variable and the risk profile of the participants as independent variables..

Table 8

Regression analysis using Order Quantity as a criterion (NR vs BS)

	Low Margin	High Margin
(Constant)	33.561**	69.578***
	(17.073)	18.484
Black Swan	25.321***	-12.854
	(8.517)	(10.180)
Risk Averse	29.230	10.407
	(19.898)	(13.742)
Risk Neutral	33.480***	43.847**
	(11.592)	(17.149)
Demand-1	.548***	.459***
	(.050)	(.053)
Over-order	119.630***	121.939***
	(14.247)	(15.067)

*: significant at 0.1, **: significant at 0.05; ***: significant at 0.01

Table 8 shows that the inclusion of a Black Swan supply disruption risk in the low margin setting increased the mean order quantity significantly, compared to the No Risk setting, and away from the optimum quantity. This finding is consistent with the findings of Kaki et al. (2015) detailed in 5.2.2. Similar to in the previous setting, the variable D-1 has a significant effect on the Order Quantity. We also observe that the participants that are Risk Neutral order significantly more than those than are Risk Seeking. However, Risk Aversion has no significant impact on Order Quantity.

Table 9 shows the results of a Wilcoxon signed-ranked test comparing the Order Quantities under the Black Swan and the No Risk settings. The test shows that the Order Quantities in the two different settings have a different distribution.

Table 9

Ranks

		N	Mean Rank	Sum of Ranks
Black Swan – No Risk	Negative Ranks	81[d]	90.22	7308.00
	Positive Ranks	114[e]	103.53	11802.00
	Ties	15[f]		
	Total	210		

Note. d. BS < NR; e. BS > NR; f. BS = NR

Z = -2.853 based on negative ranks; Asym. Sig. (2 tailed) = 0.004

Those findings show that compared to a setting with a deterministic supply, newsvendors in a Black Swan supply disruption risk setting deteriorate their ordering performance, widening the gap between their average order quantity and the optimum order quantity and therefore they support Hypothesis 1b in the low margin setting.

In the high margin setting, Table 8 shows that the inclusion of a Black Swan supply disruption risk decreased the mean order quantity, away from the optimum quantity, but that the decrease is not statistically significant.

Similar to the analysis in 5.2.2 with the introduction of a Normal Accident supply disruption risk, the introduction of a Black Swan supply disruption risk deteriorates the ordering performance, albeit non significantly, and this may be explained by the fact that the average Order Quantity is already below the pull-to-centre zone in the high margin baseline setting.

Table 8 also shows that the control variable D-1 has a significant effect on the order quantity. The control variable Over-order has also a significant effect on the order quantity and shows that over-ordering in the previous period significantly increases the order quantity of the current period.

Table 10 shows the results of a Wilcoxon signed-ranked test comparing the Order Quantities under the high margin Black Swan and the No Risk settings. The test shows that the Order Quantities in the two different settings do not have a different distribution.

Table 10

Ranks

		N	Mean Rank	Sum of Ranks
Black Swan – No Risk	Negative Ranks	101[d]	90.25	9115.00
	Positive Ranks	78[e]	89.68	6995.00
	Ties	17[f]		
	Total	196		

Note. d. BS < NR; e. BS > NR; f. BS = NR
Z = -1.530 based on positive ranks; Asym. Sig. (2 tailed) = 0.126

The results of the regression analysis and the Wilcoxon signed-rank test do not support Hypothesis 1b in the high margin setting.

5.2.4. Normal Accident vs Black Swan

In this section of the analysis, the effect of a Black Swan supply disruption risk on the ordering performance is analysed compared to the effect of a Normal Accident disruption risk. (Hypotheses 1c)

Table 11 is a linear regression analysis with the Order Quantity as a dependant variable. The independent variable is the presence of the Black Swan supply disruption risk, or alternatively the presence of a Normal Accident supply disruption risk.

Table 11

Regression analysis using Order Quantity as a criterion (NA vs BS)

	Low Margin	High Margin
(Constant)	91.083***	57.213***
	(16.356)	(18.966)
Black Swan	-3.787	-6.988
	(8.728)	(10.492)
Risk Averse	-5.166	22.751
	(16.461)	(16.566)
Risk Neutral	-1.540	4.606
	(11.205)	(12.762)
Demand-1	.478***	.489***
	(.045)	(.053)
Over-order	105.417***	123.788***
	(13.004)	(15.269)

*: significant at 0.1, **: significant at 0.05; ***: significant at 0.01

Table 11 shows that there is no significant difference in ordering performance between participants in a Normal Accident setting and participants in a Black Swan setting in both the low and high margin settings.

Table 12 shows the results of a Wilcoxon signed-ranked test comparing the Order Quantities under the low margin Black Swan and the Normal Accident settings. Table 13 shows the same but in a high margin setting. The test is inconclusive in both setting in showing that the Order Quantities in the two different settings have a different distribution.

Table 12

Ranks (Low Margin)

		N	Mean Rank	Sum of Ranks
Black Swan –	Negative Ranks	90[g]	104.47	9402.50
Normal	Positive Ranks	99[h]	86.39	8552.50
Accident	Ties	21[i]		
	Total	210		

Note. g. BS < NA; h. BS > NA; i. BS = NA

Z = -2.853 based on negative ranks; Asym. Sig. (2 tailed) = 0.004

Table 13

Ranks

		N	Mean Rank	Sum of Ranks
Black Swan –	Negative Ranks	90[g]	81.13	7302.00
Normal	Positive Ranks	82[h]	92.39	7576.00
Accident	Ties	24[i]		
	Total	196		

Note. g. BS < NA; h. BS > NA; i. BS = NA

Z = -.210 based on negative ranks; Asym. Sig. (2 tailed) = .834

Those findings do not provide support for Hypothesis 1c in both the low and high margin setting as the Ordering Performance in the Black Swan setting is not significantly different from the Ordering Performance in the Normal Accident setting in the high margin setting.

5.2.5. Summary of findings

Table 14 summarizes the findings for the hypothesis 1. In the low margin setting, the introduction of a Normal Accident or a Black Swan deteriorates the ordering performance compared to a No Risk setting. However, a Black Swan setting does not deteriorate the ordering performance significantly further compared to a Normal Accident setting.

In the high margin condition, the average order quantity in the No Risk setting is already below the pull-to-center zone and the introduction of a Normal Accident or a Black Swan does not deteriorate the order quantity significantly further.

Table 14

Hypothesis 1: Summary of findings

	No Risk	Black Swan
No Risk	N/A	Hypothesis 1b: Low Margin: **Supported** High Margin: Not Supported
Normal Accident	Hypothesis 1a: Low Margin: **Supported** High Margin: Not Supported	Hypothesis 1c: Low Margin: Not Supported High Margin: Not Supported

5.3. Hypothesis 2 Analysis (Decision Support System)

In this section of the analysis, the effect of the Decision Support System on the ordering performance is analyzed in the Normal Accident and in the Black Swan settings.

5.3.1. Normal Accident

Table 15 shows the correlations in a low margin condition between:

- The dependent variable “Order Quantity”
- The independent variable “DSS” that returns 1 if the participants have been given a Decision Support System and that returns 0 otherwise.
- The independent variable “D-1”
- The independent variable “Veryearly23” that returns 1 during the second and the third periods of the experiment and that returns 0 otherwise.
- The control variables “Risk Averse”, Risk Neutral” and “Risk Seeking”
- The control variable “Over-ordering”

Table 16 shows the same correlations but in the high margin condition.

In the low margin condition, the Order Quantity is significantly correlated to the demand of the previous week (.213). The Order Quantity and the variable

"Veryearly23" are also correlated, but the variable "veryearly23" is also correlated to the demand of the previous week. This is due to the random generation of the demand which actually gave the participants a very high demand of 423 units in period 1 and that in turn amplified the demand chasing behavior of the participants and triggered higher orders in the early stage of the experiment.

The demand of the previous week is significantly correlated to the presence of overordering the previous week (-.739), which was explained earlier in 5.2.1.

Table 15

Correlations – Normal Accident Low Margin Condition

	Mean	SD	Orde	DSS	Risk Av.	Risk Neutr.	Risk Seek.	D-1	Over-order	Veryearly23
Order	240.83	103.04	1							
DSS	0.47	0.5	-.022	1						
Risk Averse	0.14	0.344	-.029	.068	1					
Risk Neutral	0.2	0.398	-.010	-.008	-.197**	1				
Risk Seeking	0.67	0.472	.030	-.043	-.562**	-.700**	1			
D -1	206.14	143.04	.213**	.000	.000	.000	.000	1		
Over-order	0.51	0.5	.047	-.007	-.028	.016	.006	-.739**	1	
Very early 23	.29	.452	.159**	.000	.000	.000	.000	.360**	-.052	1

Note: N = 462.

**. Correlation is significant at the 0.01 level (2-tailed).

In the high margin condition, the Order Quantity is also significantly correlated to the demand of the previous period (.286), to the variable "Veryearly23" (.288) and the demand of the previous week is also correlated to the presence of overordering the previous week (-.788). In this setting, the risk profile of the participants is correlated to the presence of the Decision Support System, which can be due to the selection of the sample.

Table 16

Correlations – Normal Accident High Margin Condition

	Mean	SD	Order	DSS	Risk Av.	Risk Neutr.	Risk Seek.	D-1	Over-order	Veryearly23
Order	219.8	100.0	1							
DSS	0.45	0.499	.007	1						
Risk Averse	0.05	0.227	.030	.102*	1					
Risk Neutral	0.25	0.436	.025	-.198**	-.140**	1				
Risk Seeking	0.69	0.463	-.039	.136**	-.359**	-.874**	1			
D – 1	206.1	143.1	.286**	.000	.000	.000	.000	1		
Over-order	0.47	0.5	-.057	-.008	.002	.032	-.031	-.788**	1	
Veryearly23	.29	.452	.288**	.000	.000	.000	.000	.360**	.176	1

Note: N = 385.

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 17 shows the results of a nested regression analysis using the Order Quantity as a criterion.

“D-1” and “Over-Order” are used as control variables in 2 different models in the low and high margin conditions:

In Model 1, the same base model is used for the analysis first, before moving to a second model which adds more variables that are relevant to this particular setting.

In Model 2, the independent variable Veryearly23 has been added in the regression analysis to control if the effect of the Decision Support System evolves over time. The Veryearly23 variable returns 1 in the second and the third periods of the experiment, and returns 0 in the other periods of the experiment. Possible explanations for a different effect of the Decision Support System in the early stage of the experiment include the effect of learning and the need for the participants to gain experience on how to use the Decision Support System before it impacts the ordering performance. This possible explanation is supported by the findings of Bolton et al. (2012) who concluded that accumulated experience might improve ordering performance in a newsvendor setting. The interaction effect between the 2 independent variables Veryealry23 and Presence of a Decision Support System is also analyzed in the section.

In both the low and high margin conditions, the presence of the Decision Support System improves the ordering performance, albeit not statistically significantly, probably due to the small size of the sample. The demand of the previous period and the presence of over-ordering in the previous period both have a significant impact on the ordering performance, and the impact is similar in the low and high margin conditions. This indicates a demand chasing behavior of the participants in both low and high margin conditions. Risk Averse and Risk Neutral participants improve their ordering performance in both the low and high margin conditions, but not in a statistically significant manner. The variable “veryearly23” shows that the ordering performance is superior in periods 2 and 3 than in the rest of the experiment implying that there is no learning effect. This can be explained by the fact that participants chased the very high demand of 423 units that has been randomly generated in period 1 and by the fact that 2 periods did not generate sufficient learnings for the participants to improve their ordering performance. Indeed, the experiment of Bolton et al. (2012) lasted 100 periods and in the experiment of Kaki et al. (2015), no learning seems to have happened in period 2 and 3 in their Figure 10, even though some learning impact can be observed over the 15 periods of the experiment.

Table 17

Regression analysis using Order Quantity as a criterion – Normal Accident

	Low Margin		High Margin	
	Model 1	Model 2	Model 1	Model 2
(Constant)	115.84***	114.73***	83.654***	95.345***
	(16.814)	(17.584)	(18.745)	(19.589)
DSS	-3.648	-4.968	2.090	5.697
	(8.982)	(10.642)	(9.673)	(11.330)
Risk Averse	-5.998	-5.930	13.558	13.759
	(13.331)	(13.359)	(20.999)	(20.910)
Risk Neutral	-5.601	-5.628	4.055	4.586
	(11.471)	(11.494)	(11.124)	(11.079)
Veryearly23		-6.357		32.386**
		(14.734)		(15.491)
Veryearly23xDSS		4.657		-12.716
		(19.882)		(20.888)
Demand-1	.394***	.403***	.444***	.374***
	(.046)	(.053)	(.054)	(.062)
Over-order	92.889***	94.750***	88.706***	74.446***
	(13.297)	(14.178)	(15.332)	(16.502)

*: significant at 0.1, **: significant at 0.05; ***: significant at 0.01

5.3.2. Black Swan

Table 18 and Table 19 show the same correlations as Table 15 and Table 16 but in the Black Swan setting.

In the low margin condition, the Order Quantity is significantly correlated to risk aversion (-.104) and to the demand of the previous period (.286). Similar to in the Normal Accident setting, the risk profile of the participants and the presence of a Decision Support System are correlated, and the demand of the previous week is correlated to the presence of overordering the previous week (-.8).

Table 18

Correlations – Black Swan Low Margin Condition

	Mean	SD	Order	DSS	Risk Av.	Risk Neutr.	Risk Seek.	D-1	Over-order	Veryearly23
Order	229.8	95.15	1							
DSS	0.51	0.501	-.076	1						
Risk Averse	0.08	0.275	-.104*	.174**	1					
Risk Neutral	0.25	0.431	.033	.181**	-.17**	1				
Risk Seeking	0.62	0.485	.030	-.29**	-.38**	-.73**	1			
D – 1	206.1	143.0	.286**	.000	.000	.000	.000	1		
Over-order	0.48	0.5	-.071	-.030	-.011	.001	-.014	-.80**	1	
Veryearly23	.29	.452	.190**	.000	.000	.000	.000	.36**	-.031	1

Note: N = 427.

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

In the high margin condition, the Order Quantity is significantly correlated to risk-neutral (.167), to risk-seeking (-.145) and to the demand of the previous period (.214). Similar to the other settings, the demand of the previous week is correlated to the presence of over-ordering the previous week (-.758).

Table 19

Correlations – Black Swan High Margin Condition

	Mean	SD	Order	DSS	Risk Av.	Risk Neutr.	Risk Seek.	D-1	Over-order	Veryearly23
Order	220.19	109.94	1							
DSS	0.47	0.5	.058	1						
Risk Averse	0.13	0.339	.029	-.145**	1					
Risk Neutral	0.11	0.317	.167**	.140**	-.139**	1				
Risk Seek	0.74	0.441	-.145**	.052	-.651**	-.596**	1			
D – 1	206.14	143.07	.214**	.000	.000	.000	.000	1		
Over-order	0.49	0.5	.082	.023	.020	.011	-.018	-.758**	1	
Veryearly23	.29	.452	.231**	.000	.000	.000	.000	.360**	-.065	1

Note. N = 371

**. Correlation is significant at the 0.01 level (2-tailed).

Table 20 replicates the structure of Table 17 but in the Black Swan condition.

The presence of a Decision Support System improves the ordering performance in Model 1 and in Model 2 in the low and high margin conditions, but not in a statistically significant way. This may be due to the small sample of the participants in this experiment. The variables D-1 and Over-order have a significant effect on the ordering performance, which also implies a demand chasing behaviour in the Black Swan setting, similar to the one observed in the Normal Accident setting.

Risk Aversion significantly improves the ordering performance in the low margin condition and Risk Neutrality significantly improves the ordering performance in the high margin condition.

In both the low and high margin condition, participants with the Decision Support System improved significantly their ordering performance during periods 2 and 3. The beneficial effect of the Decision Support System faded out after period 3. A possible explanation could be that the participants stop paying attention to the Decision Support System after a few periods since it does not give them the information they need to solve the problem at hand.

Table 20

Regression analysis using Order Quantity as a criterion – Black Swan

	Low Margin		High Margin	
	Model 1	Model 2	Model 1	Model 2
(Constant)	110.88***	100.41***	46.596***	55.725***
	(17.460)	(19.008)	(19.089)	(20.061)
DSS	-10.152	1.722	6.453	-1.737
	(8.830)	(10.293)	(10.400)	(12.228)
Risk Averse	-29.397*	-29.357*	14.586	14.794
	(16.044)	(15.983)	(15.333)	(15.325)
Risk Neutral	6.219	6.190	56.459	56.554***
	(10.233)	(10.194)	(16.362)	(16.351)
Veryearly23		13.722		-2.034
		(14.750)		(16.622)
Veryearly23xDSS		-41.081**		29.157
		(18.694)		(22.585)
Demand-1	.418***	.439***	.494***	.466***
	(.049)	(.059)	(.055)	(.62)
Over-order	81.616***	86.185***	124.38***	118.66***
	(14.148)	(15.863)	(15.65)	(16.656)

*: significant at 0.1, **: significant at 0.05; ***: significant at 0.01

In both the Normal Accident and the Black Swan settings, and for both the low and the high margin conditions, the presence of a Decision Support System does not improve significantly the ordering performance of the participants over the duration of the experiment. However, In the low margin Black Swan conditions, the Decision Support System helps the participants improve the ordering performance in the early stage of the experiment, and its effect fades over time, indicating that the participants chose to ignore the Decision Support System at some point, likely because it does not provide them with the information that they need to solve the task.

5.4. Hypothesis 3 Analysis (Secondary Task)

In this section of the analysis, the effect of a Secondary Task that reduces the salience of the mean and of the previous demand, before the ordering decision for the next period is made, is analyzed in the Normal Accident and Black Swan settings.

5.4.1. Normal Accident

Table 21 and Table 22 show the same correlations as those analyzed in Table 18 and Table 19 but in the Secondary Task setting.

In the low margin condition, the Order Quantity is significantly correlated to the demand of the previous period (0.222). The demand of the previous week is also correlated to the presence of overordering the previous week (-.734). The variable Secondary Task is significantly correlated to Risk Averse (.110) and similar to the other previous settings, the variable “veryearly23” is correlated to the demand of the previous week (.360).

Table 21

Correlations – Normal Accident Low Margin Condition

	Mean	SD	Order	ST	Risk Av.	Risk Neutr.	Risk Seek.	D-1	Over-order	Veryearly23
Order	237.4	101.5	1							
ST	0.47	0.5	-.058	1						
Risk Averse	0.15	0.359	.013	.110*	1					
Risk Neutral	0.2	0.398	-.076	-.008	-.209**	1				
Risk Seeking	0.65	0.477	.053	-.076	-.578**	-.677**	1			
D – 1	206.1	143.0	.222**	.000	.000	.000	.000	1		
Over-order	0.5	0.501	.009	-.021	-.004	-.010	.011	-.734**	1	
Veryearly23	.29	.452	.174**	.000	.000	.000	.000	.360**	-.063	1

Note: N = 462

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

In the high margin condition, the Order Quantity is significantly correlated to the presence of a Secondary Task (.101), to the demand of the previous period (0.339) and to the presence of over-ordering (-.107). The demand of the previous week is also significantly and highly correlated to the presence of overordering the previous week (-.810). The variable Secondary Task is significantly correlated to Risk Aversion (.208), to Risk Neutral (-.343) and to Risk Seeking (.163).

Table 22

Correlations – Normal Accident High Margin Condition

	Mean	SD	Order	ST	Risk Av.	Risk Neutr.	Risk Seek	D-1	Over-order	Veryearly12
Order	229.36	96.81	1							
ST	0.52	0.5	.101*	1						
Risk Averse	0.1	0.296	.021	.208**	1					
Risk Neutral	0.19	0.396	-.021	-.34**	-.16**	1				
Risk Seeking	0.71	0.454	.005	.163**	-.51**	-.77**	1			
D – 1	206.1	143.0	.339**	.000	.000	.000	.000	1		
Over-order	0.49	0.5	-.107*	.024	-.008	-.012	.016	-.81**	1	
Veryearly23	.29	.452	.310**	.000	.000	.000	.000	.360**	-.036	1

Note: N = 434

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 23 shows the results of a nested regression analysis using the Order Quantity as a criterion. D-1 and Over-Order are used as control variables in the 2 different models presented in 5.3.1 in the low and high margin conditions.

In both the low and high margin conditions, the presence of the Secondary Task improved the ordering performance, and in a significant way in the high margin condition. The variable "veryearly23" has a significant impact on the ordering performance as the average demand in period 1 and 2 is higher than in period 3 and beyond. The control variables "D-1" and "Over-order" have a similar effect in both the low and high margin conditions. In this setting, the risk profile of the participants does not impact the ordering performance.

Table 23

Regression analysis using Order Quantity as a criterion – Normal Accident

	Low Margin		High Margin	
	Model 1	Model 2	Model 1	Model 2
(Constant)	135.51***	138.68***	71.228***	83.216***
	(16.625)	(17.256)	(17.772)	(19.402)
Secondary Task	-10.438	-13.760	18.256**	21.709**
	(8.978)	(10.616)	(9.003)	(10.416)
Risk Averse	1.491	1.481	2.843	2.514
	(12.774)	(12.794)	(14.478)	(14.451)
Risk Neutral	-18.255	-18.279	4.442	4.341
	(11.448)	(11.466)	(11.278)	(11.256)
Veryearly23		.010		26.005*
		(14.625)		(14.680)
Veryearly23xDSS		11.494		-10.892
		(19.792)		(18.418)
Demand-1	.351***	.340***	.494***	.432***
	(.046)	(.052)	(.050)	(.060)
Over-order	75.161***	73.341***	93.476***	79.718***
	(13.126)	(13.883)	(14.231)	(16.065)

*: significant at 0.1, **: significant at 0.05; ***: significant at 0.01

5.4.2. Black Swan

Table 24 and Table 25 show the same correlations as Table 18 and Table 19 but in a Secondary Task setting.

In the low margin setting, the Order Quantity is significantly correlated to the presence of a Secondary Task (-.093) and to the demand of the previous period

(0.276). The demand of the previous week is also significantly and highly correlated to the presence of over-ordering the previous week (-.779). The variable Secondary Task is also significantly correlated to the Risk Aversion (.226) and to Risk Seeking (-.109).

Table 24

Correlations – Black Swan Low Margin Condition

	Mean	SD	Order	ST	Risk Av.	Risk Neutr.	Risk Seek.	D-1	Over-order	Veryearly23
Order	226.0	104.7	1							
ST	0.57	0.496	-.093*	1						
Risk Averse	0.12	0.32	.106*	.226**	1					
Risk Neutral	0.16	0.366	-.021	-.017	-.158**	1				
Risk Seeking	0.71	0.454	-.040	-.109*	-.567**	-.682**	1			
D – 1	206.1	143.0	.276**	.000	.000	.000	.000	1		
Over-order	0.49	0.5	.005	-.003	.073	-.030	-.024	-.779**	1	
Veryearly23	.29	.452	.254**	.000	.000	.000	.000	.360**	-.050	1

Note: N = 483

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

In the high margin setting, the Order Quantity is significantly correlated to the demand of the previous period (0.177) and to the presence of over-ordering (.103). The demand of the previous week is also significantly and highly correlated to the presence of overordering the previous week (-.759). The variable Secondary Task is significantly correlated to Risk Neutral (.140) and the variable "veryearly23" is correlated to the demand of the previous week (.360).

Table 25

Correlations – Black Swan High Margin Condition

	Mean	SD	Order	ST	Risk Averse	Risk Neutral	Risk Seek	D-1	Over-order	Veryearly23
Order	220.4	103.4	1							
ST	0.47	0.5	.065	1						
Risk Averse	0.15	0.358	.034	-.082	1					
Risk Neutral	0.11	0.317	.105*	.140**	-.151**	1				
Risk Seeking	0.7	0.46	-.128*	-.037	-.641**	-.543**	1			
D – 1	206.1	143.1	.177**	.000	.000	.000	.000	1		
Over-order	0.49	0.5	.103*	.023	.043	-.023	-.008	-.759**	1	
Veryearly23	.29	.452	.218**	.000	.000	.000	.000	.360**	-.065	1

Note: N = 371

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Similar to Table 23, Table 26 shows the results of a nested regression analysis using the Order Quantity as a criterion. D-1 and Over-Order are also used as control variables in the 2 different models presented in 5.3.1 in the low and high margin conditions.

In both the low and high margin conditions, the Secondary Task improves the ordering performance of the participants, and in a significant manner in the low margin condition. In this setting, the control variables "D-1" and "Over-order" also impact the ordering performance in a similar way. In the high margin condition, Risk Neutral has a significant impact on the ordering performance. After further analysis of the raw data and considering the small sample of the experiment, the impact of Risk Neutrality on ordering performance is likely to be due to the sample selection.

Table 26

Regression analysis using Order Quantity as a criterion – Black Swan

	Low Margin		High Margin	
	Model 1	Model 2	Model 1	Model 2
(Constant)	75.235***	80.279***	65.108***	72.021***
	(17.006)	(18.129)	(18.366)	(19.342)
Secondary Task	-23.72***	-24.020**	7.906	3.555
	(8.798)	(10.342)	(9.931)	(11.715)
Risk Averse	30.387	30.989**	8.871	9.179
	(13.877)	(13.912)	(13.885)	(13.902)
Risk Neutral	2.324	2.196	38.184**	38.007**
	(11.751)	(11.767)	(15.767)	(15.783)
Veryearly23		9.232		3.752
		(15.453)		(16.000)
Veryearly23xDSS		.683		15.767
		(18.989)		(21.734)
Demand-1	.510***	.485***	.435***	.409***
	(.048)	(.056)	(.053)	(.060)
Over-order	113.03***	107.87***	115.86***	110.57***
	(13.638)	(14.884)	(15.089)	(16.092)

*: significant at 0.1, **: significant at 0.05; ***: significant at 0.01

In both the Normal Accident and the Black Swan settings, and in both the low and high margin conditions, the Secondary Task improves the ordering performance, regardless of the risk profile of the participants. And in all settings, the effect of the Secondary Task does not fade out over time. This is an important finding as it demonstrates that breaking down the task of the newsvendor facing supply uncertainties and inserting a secondary task that calls on short term memory between the feedback of the previous week and the ordering decision of the following week improves sustainably the performance of the decision makers that face supply uncertainties in a newsvendor setting.

6. Conclusion

This research first aimed to study the effect of two types of supply disruption risks on the ordering performance of a profit-maximizing decision maker using the newsvendor model. The Normal Accident supply disruption risk setting replicated a situation where the disruption is expected to happen with a low probability and with mild consequences and the Black Swan supply disruption risk setting replicated a situation where the disruption is expected to happen with a very low probability but with more severe consequences.

Compared to a setting with a deterministic supply, the introduction of both types of supply disruption triggered the decision makers to deviate further from the optimum order quantity and worsened the ordering performance. The asymmetry of results that is often observed in the literature was also observed in this setting of the experiment: In the low margin setting, those findings are statistically significant and in the high margin, the statistical significance could not be established.

The decision makers do not worsen significantly their ordering performance in a Black Swan setting compared to a Normal Accident setting, implying that the personal perception of the frequency and the magnitude of the supply disruption risk do not influence the ordering decision of a user of the model facing supply uncertainties.

Those findings contribute to a better understanding of the behaviour of practitioners that make ordering decisions using the newsvendor model: When confronted to a supply disruption risk, irrespective of its frequency or magnitude, the practitioners will deteriorate their ordering performance, unless nudges and strategies are developed to mitigate this effect. Those findings also complement the article of Kaki et al. (2015) in studying different types of supply disruption risks and their effect on the ordering performance.

While the effect of a supply disruption risk on the ordering performance of a user could be clearly illustrated in this setting, the asymmetry of findings between the low margin and high margin settings could be investigated in further research, for example with an experiment setting that does not emphasise the overage cost.

Once established that the decision makers deteriorate their ordering performance as they face supply uncertainties, the second part of this research aimed to test tools that can mitigate this behaviour.

Decision Support Systems reduce the cognitive load of gathering and analysing more complex data that are generated by the supply uncertainty. Even if they do not give directly the answer to the problem, they provide additional insight that can help users make a better decision. Specifically, this second part of the research studied the effect on the ordering performance of a Decision Support System that gives insight on the risk but not on the optimum order quantity.

In all settings of the experiment (Low & High Margin, Normal Accident & Black Swan), the Decision Support System improved the ordering performance. The demand of the previous week has a significant effect on the ordering performance, implying that the participants displayed the demand chasing behaviour traditionally observed in the newsvendor model. The risk profile of the participants did not have a material impact on the ordering performance, implying that the benefits of the Decision Support System apply generally to the decision makers independently of their risk appetite.

In all settings, the Decision Support System helped the participants make better ordering decisions in the early stage of the experiment but its effect faded away quickly after period 3. This is contrary to a learning effect that has been observed in the literature for longer term experiments. To explain this observation, we theorise that the participants chose to ignore the Decision Support System after a few periods as it did not give them the answer to the task at hand.

Those findings are very important for the practitioners that develop and use Decision Support Systems. While they reduce the cognitive load of its users in synthetising the additional data available, Decision Support Systems still create an incremental cognitive load and they can be ignored by their users if they do not generate direct enough information on how to solve the problem at hand. This is the case with the Decision Support System that has been designed and tested for this experiment as it gives participants insight on the risk, but not on the order quantity.

While the sample size limits the generalizability of the results, this research provides new insight into the types and limitations of Decisions Support Systems that have been studied in the literature in a behavioural newsvendor setting. To better understand the implications of these findings, future studies could be designed over more periods to understand the learning effect on the effectiveness of Decision Support Systems.

The third part of this research focused on the effect of a Secondary Task that replaced the information contained in the working memory of the decision makers after they receive the feedback of the previous period and before they make the ordering decision for the next period. The aim of introducing the Secondary Task was to reduce the salience of the mean and of the prior demand in the working memory of the decision makers and to weaken those anchors that are known to trigger the pull-to-centre effect.

The introduction of a Secondary Task improved the ordering performance of the participants in all settings of the experiment (Low and High Margin, Normal Accident & Black Swan) and the risk profile of the participants did not moderate significantly the results. Since the Secondary Task had to be performed at each period with no possible learning benefits over time, its effect was not different in the early stage of the experiment than in the later stage of the experiment.

By analysing the effect of a secondary task that calls on short term memory and that is introduced after the feedback of the previous period is given and before the ordering decision of the next period is made, this thesis has shown that participants beneficiate sustainably from a reset of the working memory that reduces the salience of the anchors traditionally observed in the pull-to-centre effect. This finding is very important for practitioners and scholars because it shows that the decision makers facing supply uncertainties in a newsvendor setting can be debiased in breaking down the ordering process and in inserting a non-related activity between the feedback from the previous period and the ordering decision for the next period.

This finding has broad implications for the practitioners, beyond Operations. For example, in Human Resources, the profile of those tasked with inventory management should include the ability to multi-task. This is tremendously important for recruitment and training purposes. Additionally, the organization of work of inventory managers, and the configuration of the software they use, should change to break down the ordering process and include a Secondary Task. This research also extends the paper of Lee and Siemsen (2017) in a behavioral newsvendor setting and contributes to the literature on Secondary Task in providing an operational setting in which it can be applied.

The experiments of this research were conducted using an online platform during various phases of the COVID-19 confinement. It is unclear how it has changed the nature of the sample available online and how it has impacted the validity of the findings. No study to my knowledge has been published to this date.

The complexity of the newsvendor problem for the general population, even with a master's degree, generated a high number of answers that could not be included in the dataset. Moreover, the number of settings (7) that had to be duplicated each time for the low and high margin condition also limited the number of respondents that qualified and that could be paid within the research budget and constraints. As a consequence, the statistical significance in some of the settings was hard to demonstrate. Further research with a bigger scale and in a laboratory can be developed to complement the findings of this thesis.

7. References

Arrow, K. J., Harris, T., & Marschak, J. (1951). Optimal inventory policy. *Econometrica: Journal of the Econometric Society*, 250–272.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation*, *2*(4), 89–195.

Ayres, P. L. (2001). Systematic mathematical errors and cognitive load. *Contemporary Educational Psychology*, *26*(2), 227–248.

Ayres, P., & Paas, F. (2012). Cognitive load theory: New directions and challenges. *Applied Cognitive Psychology*, *26*(6), 827–832.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*(4), 569.

Becker-Peth, M., Thonemann, U. W., Donohue, E., & Leider, S. (Eds). (2018). Behavioral inventory decisions. *The Handbook of Behavioral Operations*, 11393.

Becker-Peth, M., Thonemann, U. W., & Gully, T. (2018). A note on the risk aversion of informed newsvendors. *Journal of the Operational Research Society*, *69*(7), 1135–1145.

Bennis, W., & Nanus, B. (1985). The strategies for taking charge. *Leaders, New York: Harper. Row*, *41*.

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

Bolton, G. E., & Katok, E. (2008). Learning by doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback. *Manufacturing & Service Operations Management*, *10*(3), 519–538.

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

Bostian, A. A., Holt, C. A., & Smith, A. M. (2008). Newsvendor "pull-to-center" effect: Adaptive learning in a laboratory experiment. *Manufacturing & Service Operations Management*, *10*(4), 590–608.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 53–61.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112–130.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293–332.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81.

Chen, R. R., Cheng, T. C. E., Choi, T.-M., & Wang, Y. (2016). Novel advances in applications of the newsvendor model. *Decision Sciences*.

Chi, M. T., Glaser, R., & Rees, E. (1981). *Expertise in problem solving.* Pittsburgh Univ PA Learning Research and Development Center.

Ciarallo, F. W., Akella, R., & Morton, T. E. (1994). A Periodic Review, Production Planning Model with Uncertain Capacity and Uncertain Demand—Optimality of Extended Myopic Policies. *Management Science*, *40*(3), 320–332. https://doi.org/10.1287/mnsc.40.3.320

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.

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), 131–156.

Croson, D., Croson, R., & Ren, Y. (2008). How to manage an overconfident newsvendor. *Working Paper*.

Dada, M., Petruzzi, N. C., & Schwarz, L. B. (2007). A newsvendor's procurement problem when suppliers are unreliable. *Manufacturing & Service Operations Management*, *9*(1), 9–32.

De Groot, A. D. (2014). *Thought and choice in chess* (Vol. 4). Walter de Gruyter GmbH & Co KG.

de Véricourt, F., Jain, K., Bearden, J. N., & Filipowicz, A. (2013). Sex, risk and the newsvendor. *Journal of Operations Management*, *31*(1–2), 86–92. https://doi.org/10.1016/j.jom.2012.11.001

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114.

Dixon, M. J., Victorino, L., Kwortnik, R. J., & Verma, R. (2017). Surprise, anticipation, and sequence effects in the design of experiential services. *Production and Operations Management*, *26*(5), 945–960.

Edgeworth, F. Y. (1888). The mathematical theory of banking. *Journal of the Royal Statistical Society*, *51*(1), 113–127.

Ellis, S. C., Henry, R. M., & Shockley, J. (2010). Buyer perceptions of supply disruption risk: A behavioral view and empirical assessment. *Journal of Operations Management*, *28*(1), 34–46.

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

Farrell, A. M., Grenier, J. H., & Leiby, J. (2017). Scoundrels or stars? Theory and evidence on the quality of workers in online labor markets. *The Accounting Review*, *92*(1), 93–114.

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

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

Gallego, G., & Moon, I. (1993). The distribution free newsboy problem: Review and extensions. *Journal of the Operational Research Society*, *44*(8), 825–834.

Giri, B. C. (2011). Managing inventory with two suppliers under yield uncertainty and risk aversion. *International Journal of Production Economics*, *133*(1), 80–85.

Gönül, M. S., Önkal, D., & Lawrence, M. (2006). The effects of structural characteristics of explanations on use of a DSS. *Decision Support Systems*, *42*(3), 1481–1493.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224.

Gurnani, H., Ramachandran, K., Ray, S., & Xia, Y. (2014). Ordering Behavior Under Supply Risk:An Experimental Investigation. *Manufacturing & Service Operations Management*, *16*(1), 61–75. https://doi.org/10.1287/msom.2013.0453

Halford, G. S., Maybery, M. T., & Bain, J. D. (1986). Capacity limitations in children's reasoning: A dual-task approach. *Child Development*, 616–627.

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407.

Henig, M., & Gerchak, Y. (1990). The structure of periodic review policies in the presence of random yield. *Operations Research*, *38*(4), 634–643.

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

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

Hutchison-Krupat, J., & Chao, R. O. (2014). Tolerance for Failure and Incentives for Collaborative Innovation. *Production and Operations Management*, *23*(8), 1265–1285. https://doi.org/10.1111/poms.12092

Ipeirotis, P. G. (2010). *Demographics of mechanical turk*.

Jucker, J. V., & Rosenblatt, M. J. (1985). Single-period inventory models with demand uncertainty and quantity discounts: Behavioral implications and a new solution procedure. *Naval Research Logistics Quarterly*, *32*(4), 537–550.

Käki, A., Liesiö, J., Salo, A., & Talluri, S. (2015a). Newsvendor decisions under supply uncertainty. *International Journal of Production Research*, *53*(5), 1544–1560.

Käki, A., Liesiö, J., Salo, A., & Talluri, S. (2015b). Newsvendor decisions under supply uncertainty. *International Journal of Production Research*, *53*(5), 1544–1560.

Khouja, M. (1999). The single-period (news-vendor) problem: Literature review and suggestions for future research. *Omega*, *27*(5), 537–553.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology*, *17*(2), 248–294.

Kremer, M., Minner, S., & Van Wassenhove, L. N. (2010). Do Random Errors Explain Newsvendor Behavior? *Manufacturing & Service Operations*

Management, *12*(4), 673–681. https://doi.org/10.1287/msom.1100.0294

Lau, A. H.-L., & Lau, H.-S. (1988). The newsboy problem with price-dependent demand distribution. *IIE Transactions*, *20*(2), 168–175.

Lau, N., & Bearden, J. N. (2012). Newsvendor Demand Chasing Revisited. *Management Science*, *59*(5), 1245–1249. https://doi.org/10.1287/mnsc.1120.1617

Lau, N., & Bearden, J. N. (2013). Newsvendor demand chasing revisited. *Management Science*, *59*(5), 1245–1249.

Lau, N., Hasija, S., & Bearden, J. N. (2014). Newsvendor pull-to-center reconsidered. *Decision Support Systems*, *58*, 68–73.

Lee, Y. S., Seo, Y. W., & Siemsen, E. (2018). Running Behavioral Operations Experiments Using Amazon's Mechanical Turk. *Production and Operations Management*, *27*(5), 973–989. https://doi.org/10.1111/poms.12841

Lee, Y. S., & Siemsen, E. (2017). Task decomposition and newsvendor decision making. *Management Science*, *63*(10), 3226–3245.

Loch, C. H., & Wu, Y. (2008). Social Preferences and Supply Chain Performance: An Experimental Study. *Management Science*, *54*(11), 1835–1849. https://doi.org/10.1287/mnsc.1080.0910

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

Mason, W., & Watts, D. J. (2009). Financial incentives and the" performance of crowds". *Proceedings of the ACM SIGKDD Workshop on Human Computation*, 77–85.

McKinsey Institute. (2020). *Risk, Resilience and Rebalancing in global value Chain*.

Merzifonluoglu, Y., & Feng, Y. (2014). Newsvendor problem with multiple unreliable suppliers. *International Journal of Production Research*, *52*(1), 221–242. https://doi.org/10.1080/00207543.2013.835497

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Mitroff, I. I., & Alpaslan, M. C. (2003). *Preparing for evil*. Harvard Business School Pub.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502.

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

Morse, P. M., & Kimball, G. E. (1951). *Methods of operations research, New York and London*.

Nelson Ford, F. (1985). Decision support systems and expert systems: A comparison. *Information & Management*, *8*(1), 21–26. https://doi.org/10.1016/0378-7206(85)90066-7

Owen, E., & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Educational Psychology*, *77*(3), 272.

Özer, Ö., Zheng, Y., & Ren, Y. (2014). Trust, trustworthiness, and information sharing in supply chains bridging China and the United States. *Management Science*, *60*(10), 2435–2460.

Paas, Fred G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429.

Paas, Fred GWC, Van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*(1), 419–430.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, *5*(5), 411–419.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17*(4), 398–422.

Perrow, C. (1984). *Normal accidents: Living with high risk technologies-Updated edition*. Princeton university press.

Perrow, C. (2011). *Normal accidents: Living with high risk technologies-Updated edition*. Princeton university press.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193.

Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, *299*, 172–179.

Rauscher, H. M. (1999). Ecosystem management decision support for federal forests in the United States: A review. *Forest Ecology and Management*, *114*(2–3), 173–197.

Ray, P., & Jenamani, M. (2016). Sourcing decision under disruption risk with supply and demand uncertainty: A newsvendor approach. *Annals of Operations Research*, *237*(1–2), 237–262.

Ren, Y., & Croson, R. T. A. (2012). *Explaining biased newsvendor orders: An experimental study*. Working paper, University of Texas at Dallas, Richardson.

Schultz, K. L., Robinson, L. W., Thomas, L. J., Schultz, J., & McClain, J. O. (2018). The Use of Framing in Inventory Decisions. *Production and Operations Management*, *27*(1), 49–57. https://doi.org/10.1111/poms.12782

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

Schweitzer, M. E., & Cachon, G. P. (2000b). Decision Bias in the Newsvendor Problem with a Known Demand Distribution: Experimental Evidence. *Management Science*, *46*(3), 404–420. https://doi.org/10.1287/mnsc.46.3.404.12070

Serel, D. A. (2008). Inventory and pricing decisions in a single-period problem involving risky supply. *International Journal of Production Economics*, *116*(1), 115–128.

Simchi-Levi, D. (2015, June 9). Find the Weak Link in Your Supply Chain. *Harvard Business Review*. https://hbr.org/2015/06/find-the-weak-link-in-your-supply-chain

Simchi-Levi, D., Schmidt, W., Wei, Y., Zhang, P. Y., Combs, K., Ge, Y., Gusikhin, O., Sanders, M., & Zhang, D. (2015). Identifying Risks and Mitigating Disruptions in the Automotive Supply Chain. *Interfaces*, *45*(5), 375–390. https://doi.org/10.1287/inte.2015.0804

Swaminathan, J. M., & Shanthikumar, J. G. (1999). Supplier diversification: Effect of discrete demand. *Operations Research Letters*, *24*(5), 213–221.

Sweller, J., & Gee, W. (1978). Einstellung, the sequence effect, and hypothesis theory. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(5), 513.

Sweller, John. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285.

Sweller, John, Ayres, P., & Kalyuga, S. (2011). Measuring cognitive load. In *Cognitive load theory* (pp. 71–85). Springer.

Sweller, John, Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296.

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable* (Vol. 2). Random house.

Tuncel, O., Taneri, N., & Hasija, S. (2019). *Why are Minimum Order Quantity Contracts Popular in Practice? A Behavioral Investigation* (SSRN Scholarly Paper ID 3169112). Social Science Research Network. https://doi.org/10.2139/ssrn.3169112

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*(4), 293.

Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, *43*(1), 16–26.

Yano, C. A., & Lee, H. L. (1995). Lot sizing with random yields: A review. *Operations Research*, *43*(2), 311–334.

Ye, L. R., & Johnson, P. E. (1995). The Impact of Explanation Facilities on User Acceptance of Expert Systems Advice. *MIS Quarterly*, *19*(2), 157–172. https://doi.org/10.2307/249686

Yin, M., Chen, Y., & Sun, Y.-A. (2013). *The effects of performance-contingent financial incentives in online labor markets*.

Zhang, Y., & Siemsen, E. (2019). A meta-analysis of newsvendor experiments: Revisiting the pull-to-center asymmetry. *Production and Operations Management*, *28*(1), 140–156.

8. Appendices

Figure 9

LM.NA.NT - Frequency, Mean and Standard deviation

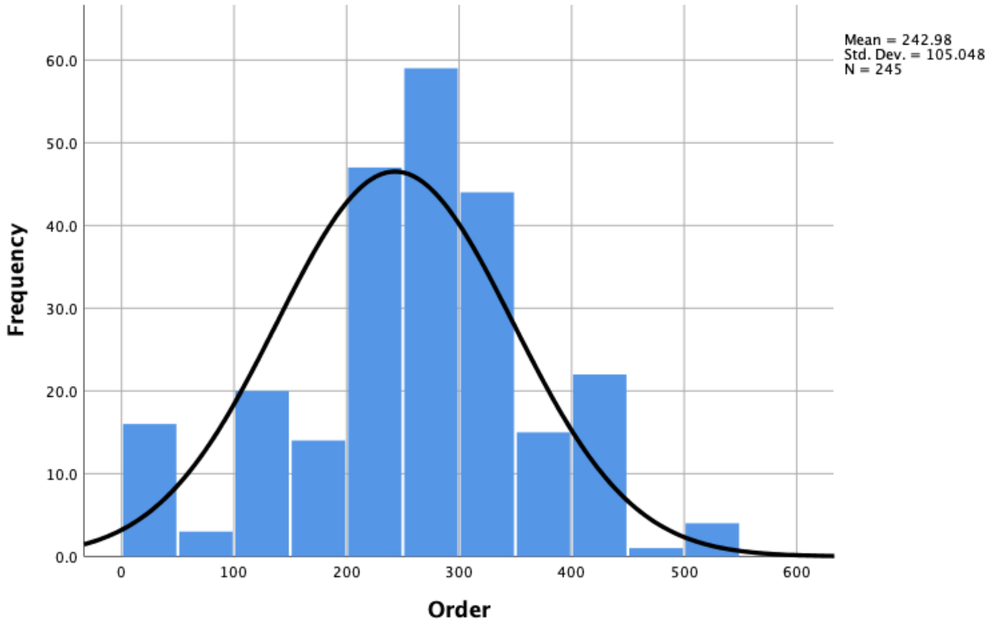


Figure 10

HM.NA.NT - Frequency, Mean and Standard deviation

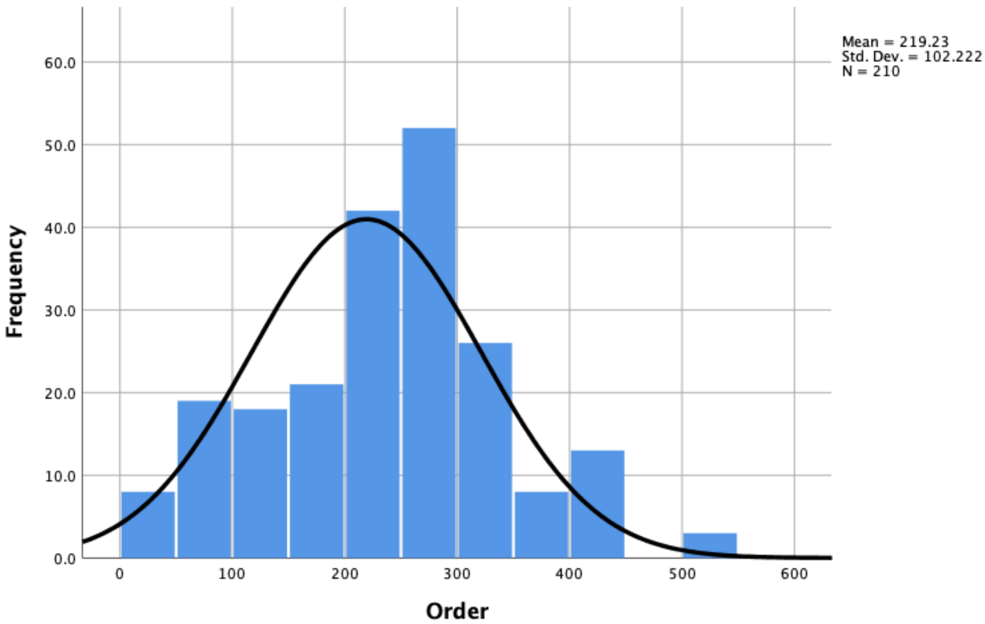


Figure 11

LM.BS.NT - Frequency, Mean and Standard deviation

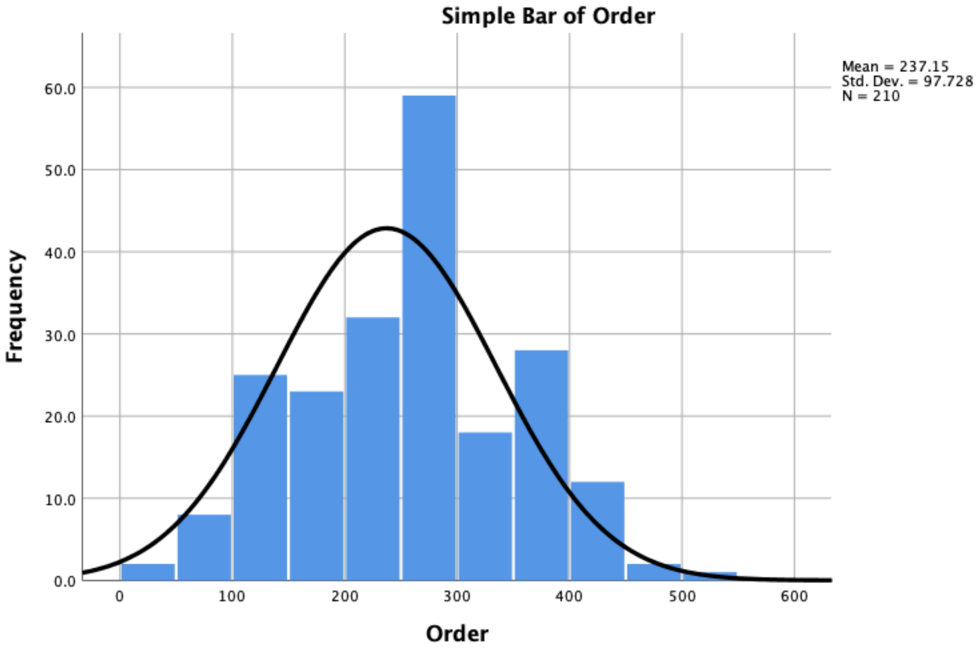


Figure 12

HM.BS.NT - Frequency, Mean and Standard deviation

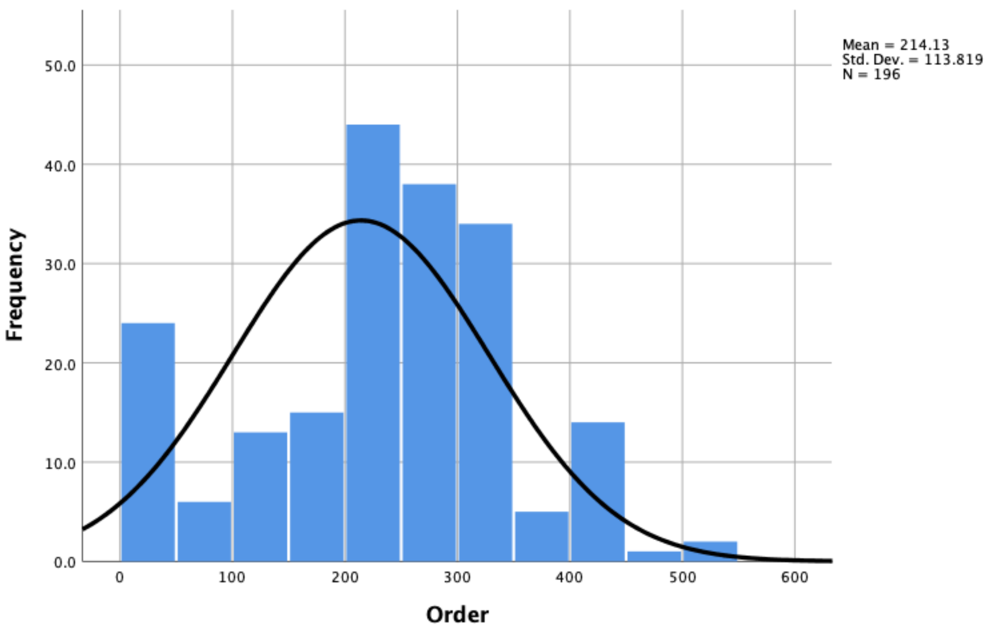


Figure 13

LM.NA.DSS - Frequency, Mean and Standard deviation

Figure 14

HM.NA.DSS - Frequency, Mean and Standard deviation

Figure 15

LM.BS.DSS - Frequency, Mean and Standard deviation

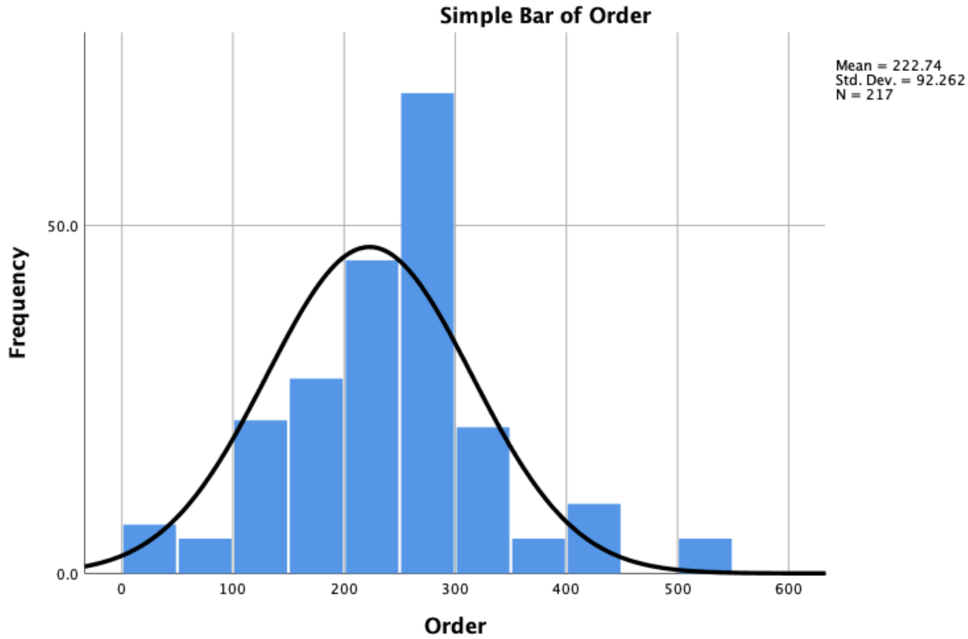


Figure 16

HM.BS.DSS - Frequency, Mean and Standard deviation

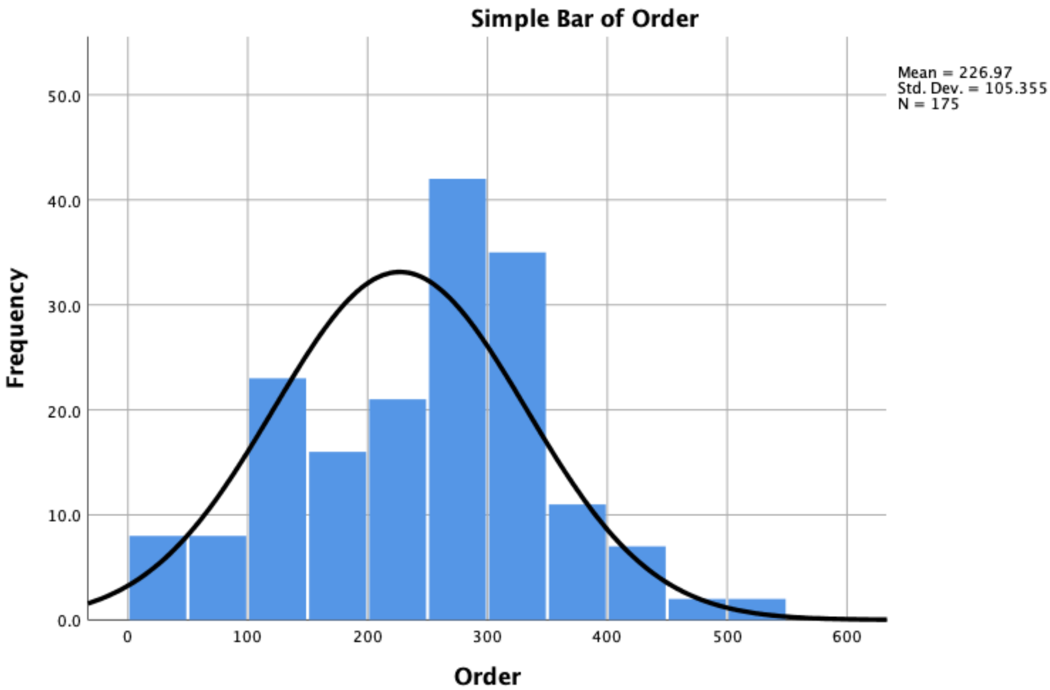


Figure 17

LM.NA.ST - Frequency, Mean and Standard deviation

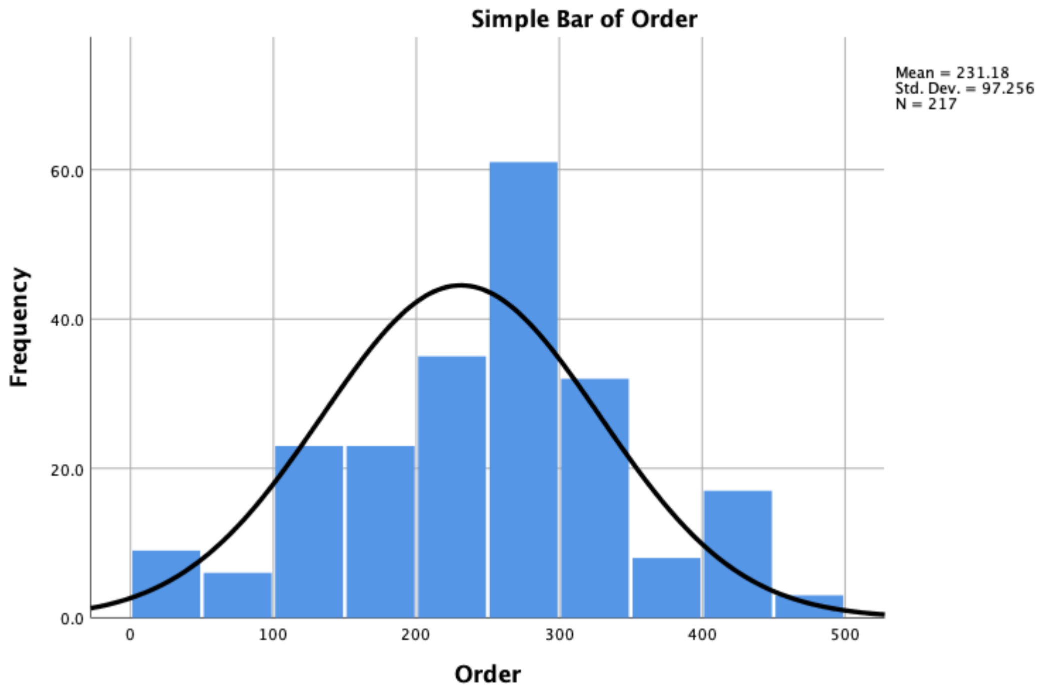


Figure 18

HM.NA.ST - Frequency, Mean and Standard deviation

Figure 19

LM.BS.ST - Frequency, Mean and Standard deviation

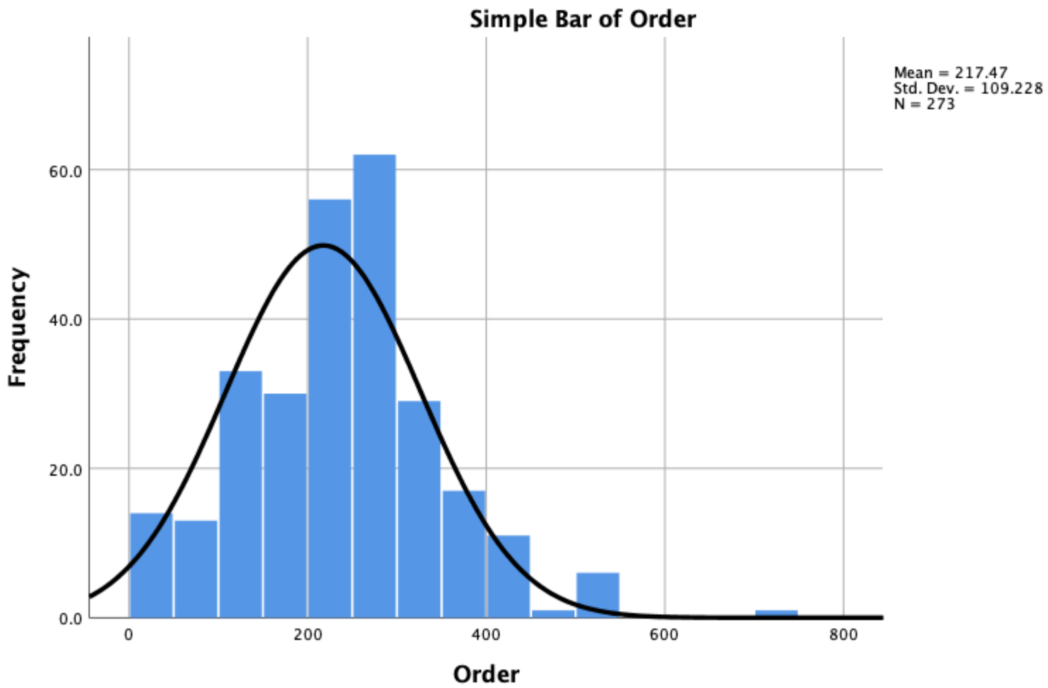


Figure 20

HM.BS.ST - Frequency, Mean and Standard deviation