TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Wirtschaftswissenschaften

Lehrstuhl für Operations Management

Behavioral Operations: Experimental Insights into Inventory, Health Care, and Portfolio Planning

Dipl.-Kfm. Sebastian Moritz Schiffels

Vollständiger Abdruck der von der Fakultät für Wirtschaftswissenschaften der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Martin Grunow
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Rainer Kolisch
	2.	Univ.-Prof. Dr. Stefan Minner

Die Dissertation wurde am 24.04.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Wirtschaftswissenschaften am 15.02.2015 angenommen.

Acknowledgements

First of all, I would like to thank my advisor, Prof. Dr. Rainer Kolisch, who encouraged me to conduct research in a fascinating and promising research domain and who guided and always supported me throughout my dissertation. Furthermore, I would like to express my gratitude to the other coauthors of my research projects, Prof. Dr. Jens Brunner, Thomas Fliedner, and Dr. Andreas Fügener. The way we complemented each other was fantastic. I would also like to thank Prof. Dr. Mirko Kremer, who invited me to the Pennsylvania State University and who gave me the opportunity to temporarily join the research group at the Smeal College of Business. The time as a visiting researcher opened new perspectives into my research and was a great experience for me personally. Very special thanks go to all my former and current colleagues: Claus Brech, Prof. Dr. Jens Brunner, Dr. André Dahlman, Alexander Döge, Dr. Jia-Yan Du, Thomas Fliedner, Markus Frey, Dr. Andreas Fügener, Dr. Daniel Gartner, Ferdinand Kiermaier, Dr. Philipp Melchiors, Dr. Anulark Naber, Christian Ruf, Dr. Hans-Jörg Schütz, and Martin Tritschler. All of them contributed to a brilliant work environment with many valuable discussions as well as a great social environment. Furthermore, I am grateful to Prof Dr. Stephan Minner, who fed my enthusiasm for this field of research during my studies at the University of Mannheim and to Dr. Jan Arnold who inspired me to conduct research projects at a university while supervising my diploma thesis. Finally, I thank my girlfriend Sabrina Seibert and my parents Brigitte Schiffels-Pitz and Prof. Dr. Edmund Schiffels for supporting me with heart and soul all the way.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Recent Findings in Behavioral Operations

Behavioral issues in the domain of operations research and operations management are the focus of a relatively new stream of research at the intersection of analytical and behavioral disciplines. There has been an increasing number of papers in behavioral operations in the past ten years and, while early research focused especially on inventory, production, and supply chain management, there is a large breadth of behavioral operations today (see Croson et al. [33]). It has become an accepted sub-field (see Sodhi and Tang [128]) and researchers are being encouraged to further engage in taking into account human behavior in operations management (see Gino and Pisano [58]) as well as operations research settings (see Hämäläinen et al. [64]), as there are manifold opportunities in this young but continuously growing domain. The importance of the field is also shown by various special issues as well as a number of review articles in top tier operations management journals. One of the earliest reviews discussing the role of experimental psychology in the context of operations research was written by Bearden and Rapoport [4], pointing out that both streams could enhance each other. Besides individual decision making, Loch and Wu [93] emphasize the influence of group dynamics, emotions, and culture in behavioral operations. Bendoly et al. [7] discuss the importance of controlled experiments, providing a review based on a framework taking into account behavioral assumptions made in analytical operations management models. They find increasing evidence that analytical models fail to reflect decision makers' actual behavior and goals in many contexts. Reviewing numerous studies, Gino and Pisano [58] discuss the implications of incorporating behavioral and cognitive factors as well as general opportunities for behavioral research in operations management. Bendoly et al. [6] review the bodies of knowledge in four areas, cognitive psychology, social psychology, group dynamics, and system dynamics, which provide a foundation for behavioral operations research. Katsikopoulos and Gigerenzer [75] encourage researchers to devote more attention to models of heuristics in the future since the descriptive power of the utility theory is limited in many settings. While operations management uses operations research methods to improve operations, Hämäläinen et al. [64] argue that there is considerably more behavioral research in operations management than in operations research, and they emphasize the need for more behavioral research in advancing the practice of operations research.

A problem of special interest in the field of behavioral operations management is the newsvendor problem, which addresses one of the fundamental questions of inventory management, how much to order. While it is extensively studied in model based research (see Khouja [77]), it has also become one of the most investigated topics in behavioral operations. Since the seminal paper of Schweitzer and Cachon [122], more than 20 experimental studies on the newsvendor problem have been published considering, e.g. individual differences (see Moritz et al. [104]), the value of information availability (see Kremer et al. [83]), or gender differences (see de Véricourt et al. [34]) to mention just a few. The experimental studies have in common that robust behavioral patterns are observed, and furthermore they demonstrate decision maker behavior which underlies systematic deviations of normative predictions. The importance of the newsvendor problem for operations management lies in its practical relevance as well as its manifold extensions. It serves as a base for a great number of related problems in classical domains like inventory (see Sachs and Minner [117]), production (see Chod et al. [27]), or procurement management (see Budde and Minner [20]), but also in areas such as health care management (see Olivares et al. [108]) or project management (see Trietsch [133]). A review of experimental research investigating the human behavior in newsvendor situations is provided by Kremer and Minner [81], distinguishing between decision biases, antecedent sources of biases, as well as debiasing strategies.

While there are many behavioral studies in the areas inventory, production, and supply chain management, the service sector and its sub-areas have not been so thoroughly investigated (see Croson et al. [33]). The service sector in general, and the health care sector in particular, have very strong behavioral influences (see Brailsford and Schmidt [19]), which comes as no surprise since people are vital for health care services. Behavioral health care operations management can be classified with respect to the people involved in hospitals, who can be divided into staff (service providers) and patients (service receivers). Several studies discuss patients' behavior which typically relate to patient satisfaction, e.g. perceived waiting time (see Huang [69]) or general patient satisfaction (see Bleich et al. [13]). Nurses and surgeons comprise the majority of the staff in hospitals. Nurses' behavior is often discussed in the context of job satisfaction (see Chang et al. [25], Irvine and Evans [70], Jamal and Baba [71]). While surgeons make both medical and management decisions, in the literature their behavior is mainly discussed in the context of medical decision making, where surgeons' biased decision making is observed (see Bornstein and Emler [17], Bland and Altman [12], Moskowitz et al. [105]). Wachtel and Dexter [135] discuss the application of the newsvendor problem to determine time periods, where staff is required in the operating room and they expect decision biases known from the experimental newsvendor studies to also be present in the operating room setting. It remains an open question whether surgeons exhibit similar biases as inventory managers.

Another research stream in behavioral operations is related to project management topics (see Loch and Wu [93], Bendoly et al. [7]), including studies focusing on project portfolio selection as well as the planning and scheduling of projects. As it is well-known that projects are almost never completed on schedule and budget (see Loch and Wu [93]), experimental studies can help to better understand project manager behavior, considering for example behavioral effects of project prioritization (see Bendoly et al. [10]) or managers' willingness to share resources in multi-project settings (see Bendoly et al. [8]). In order to investigate how decision maker behavior influences performance, Gino and Pisano [57] conduct a simulation based study, assuming mangers employ simple heuristics to decide which projects to fund and which projects to terminate. Moreover, there are behavioral studies considering project selection settings with project-specific risks (see Chow and

Haddad [28]), the willingness to spend resources for a project related to the budget already invested (see Garland [51]), as well as the influence of group decisions on the effect of sunk costs (see Whyte [136]). Furthermore, there is some literature investigating project selection in the context of decision support (see Ghasemzadeh and Archer [54]). A review of experimental studies dealing with behavioral issues in portfolio decision analysis, in particular resource allocation problems, is written by Fasolo et al. [44]. Typically, resource allocation decisions are about time or money (see Langholtz et al. [89]) and allocating resources to projects in order to increase the value of a project portfolio or to reduce the completion time of a project are fundamental problems in operations research. While several studies consider the human element in these contexts, Bendoly and Swink [9] argue that there is still a lack of behavioral studies taking into account resource interdependent settings such as project management.

One can conclude that behavioral operations is a growing field with numerous sub-areas like inventory, health care, and project management. A common denominator is the use of laboratory experiments as a major method since they provide an important opportunity to obtain new insights and to better link theory and practice (see Katok [73]). While laboratory experiments are common practice in many disciplines, like finance, accounting, and marketing, providing groundbreaking achievements, they are still in the early stages of development in operations research and operations management, even if it is well known that laboratory experiments are a major source of knowledge in the social sciences (see Falk and Heckman [43]). At the cost of lower external validity, the power of laboratory experiments is the control of situational factors (see Kremer [80]). They are a good first step to obtain fundamental insights into problem settings and to understand at least some of the behavioral factors that cause suboptimal behavior. Since the human element is a common influencing factor in most operations processes (see Bendoly et al. [7]), analytical models are often based on assumptions about decision maker behavior, or their goals, to make the mathematics tractable. Experiments can either validate or challenge the assumptions and implications of analytical models, and therefore experiments complement analytical approaches by bridging the gap between models and real business problems (see Gans and Croson [50]). Laboratory experiments are a powerful tool to create better operations management models (see Katok [73]).

1.2 Structure of the Dissertation

This dissertation presents experimental investigations of human decision making in three specific settings, inventory management, health care management, and portfolio planning, undertaking a set of laboratory experiments. The experimental findings should help to improve the understanding of human decision making in operations management and operations research contexts. One of the most studied operations management problems under uncertainty is the newsvendor model. Whereas previous studies only focused on opportunity-based settings, in Chapter 2 a laboratory study is presented in order to provide insight into decision biases in general penalty-based newsvendor problems. A penalty-based newsvendor problem is inherent in many practical applications involving, e.g. contractual penalties or penalized reorders. While behavioral aspects in health care operations management are widely ignored in the literature, Chapter 3 addresses surgeons behavior in operating room planning. Since planning surgery durations, involving penalties for staff overtime, is a typical application of a penalty-based newsvendor problem, Chapter 2 provides the theoretical foundation for Chapter 3. While most previous newsvendor studies were conducted with students, a study with experienced surgeons was conducted in order to increase practical acceptance and to reveal differences in behavior. The findings demonstrate identical biases in the health care setting as in the classical newsvendor studies, along with context related behavior. In addition to the newsvendor problem, the knapsack problem is arguably one of the most important problems in operations research. It serves as a foundation for manifold applications as well as more complex models considering, e.g. multiple objectives or project dependencies. In Chapter 4, an experimental investigation of behavioral decision making in the knapsack problem is provided, which is based on an experimental framework where subjects may dynamically select and deselect alternatives to create their portfolio. The experimental findings verify suboptimization in portfolio decision making as well as adherence to simple heuristics. The dissertation concludes with Chapter 5.

The dissertation is based on three for the most part independent research projects, i.e. Schiffels et al. [120], Fügener et al. [48], Schiffels et al. [119]. All research projects have in common that they are based on experimental studies of human decision making focusing on a fundamental problem in operations research, either the newsvendor problem or the knapsack problem. The structure of the dissertation should allow the reader to understand each chapter separately, as they are designed to be publishable as stand-alone journal articles. Notations and symbols are provided for each chapter separately in Appendix B and a more detailed summary of the three research projects is provided in the following.

On the Assessment of Costs in a Newsvendor Environment: Insights from an Experimental Study

Chapter 2 is based on Schiffels et al. [120] and addresses the question of how the assessment of costs influences decisions in a newsvendor setting. We expect that different cost types lead to different behavior. In our investigation, we consider a newsvendor problem with opportunity costs and a newsvendor problem with penalty costs. In addition, we differentiate between three cases with different margins for each of the two problems. In an experimental study, we observe that the average order quantities in the newsvendor problem with penalty costs exceed the average order quantities in the newsvendor problem with opportunity costs and that a mean anchor effect, familiar from a number of previous studies, exists. A different weighting of costs can be seen as the main driver for the different order quantities. Thus, a biased perception of different cost types exists and decision makers are more sensitive to penalty costs than to opportunity costs. Based on our observations, we can identify situations where the cost weighting and the mean anchor effect compensate for each other and thus lead to "good" decisions as well as situations where the two effects compound and therefore lead to "bad" decisions. As penalty costs are present in many newsvendor situations, our insights allow us to apply the findings from behavioral studies of the newsvendor problem to a broader context.

Over- and Under-Utilization of Operating Rooms: Insights from Behavioral Health Care Operations Management

Chapter 3 is based on Fügener et al. [48] and considers the planning of surgery durations which is a crucial task for efficient usage of operating theaters. Both planning too long and too short durations for surgeries lead to operating room inefficiency, e.g. idle time, overtime, or rescheduling of surgeries. The overall objective of planning surgery durations is to minimize the expected operating room inefficiency, since surgery durations are stochastic. While most health care studies assume rational behavior of decision makers, experimental studies have shown that decision makers often act nonrational. Based on insights from health care operations management, medical decision making, behavioral operations management, as well as empirical observations, we derive hypotheses that surgeons' behavior deviates from rational behavior. To investigate this, we undertake an experimental study where experienced surgeons were asked to plan surgeries with uncertain durations. We discover systematic deviations from optimal decision making and offer behavioral explanations for the observed biases. Our research provides new insights to tackle a major problem in hospitals, i.e. low operating room utilization going along with staff overtime.

Behavioral Portfolio Decision Making: Insights from an Experimental Study

Chapter 4 is based on Schiffels et al. [119] and addresses the question of how human decision makers behave in the context of portfolio decision making. Choosing the right projects from a set of alternatives is a key driver of success and failure for organizations. We set up an experimental study based on the knapsack problem to investigate human portfolio selection. Decision makers select suboptimal portfolios consistently and independent of the considered experimental treatment. Based on subjects portfolio construction we identify two decision heuristics which partially explain observed decision maker behavior during the selection process. Furthermore, we demonstrate that decision makers typically consider only a subset of all alternatives even for small problem instances. Our findings demonstrate that people try to adhere to simple heuristics but that the problem complexity limits their application to a subset of alternatives.

Chapter 2

On the Assessment of Costs in a Newsvendor Environment: Insights from an Experimental Study

2.1 Introduction

In the newsvendor problem, a decision maker has to decide on the number of ordered products under stochastic demand. Once the uncertainty is resolved, the costs incurred from the mismatch between the decision and the realization become apparent. The decision maker observes that his decision was too "high" or too "low". The newsvendor model provides a theoretically grounded approach to determine the optimal order quantity, i.e. the order quantity that minimizes the expected mismatch costs.[1] However, experimental studies show that decision makers systematically deviate from the optimal order quantity. In their seminal paper, Schweitzer and Cachon [122] observe a pattern of behavior where subjects order too few high margin products and too many low margin products. According to the anchoring and adjustment heuristic (see Tversky and Kahneman [134]), this too low/too high pattern can be explained by the fact that individuals anchor on the mean

[1] A minimization of the expected mismatch costs is equivalent to a maximization of the expected profit, see Silver et al. [125] or Khouja [77].

demand and insufficiently adjust toward the optimal order quantity. A number of follow-up studies have confirmed the too low/too high pattern, e.g. in experimental newsvendor studies considering doubled payoffs and reduced order frequency (see Bostian et al. [18]), the effect of learning (see Bolton and Katok [14]), different demand distributions (see Benzion et al. [11]), participants with different educational backgrounds (see Bolton et al. [15]), different frames (see Kremer et al. [82], Schultz et al. [121]), multilocation inventory systems (see Ho et al. [66]), different payment schemes (see Chen et al. [26]) as well as cross-cultural differences between Western and Eastern countries (see Feng et al. [45]). Order decisions in the newsvendor problem tend to be biased towards the anchor of mean demand, which we call the "mean anchor effect". For a recent review considering experimental studies of the newsvendor problem, see Kremer and Minner [81].

Although many studies discuss behavioral aspects in the newsvendor problem, there is hardly any research on the assessment of the different cost types. Since costs are one of the essential influencing variables in the newsvendor problem, the assessment of costs may have a strong effect on human decision making. Depending on the field of application, costs like out-of-pocket costs, opportunity costs, or penalty costs can be relevant when deciding on the order quantities. A detailed definition of the cost types in the context of our study will be given in Section 2.2. Previous studies have shown that these cost types may have a diverse influence on behavior in several situations. The indirect character of opportunity costs is a reason why they are often neglected in decision making. Northcraft and Neale [107] state that opportunity costs are abstract possibilities which can lead to a biased assessment of the cost/benefit picture of a decision maker. This biased opportunity cost perception is documented in numerous papers. The results of an experimental study by Becker et al. [5] suggest that decision makers consider opportunity costs as less important than out-of-pocket costs and even ignore them in some cases. A study by Friedman and Neumann [47] leads to consistent results. They conclude that decision makers underweight opportunity costs when only partial information is available. While Becker et al. [5] as well as Friedman and Neumann [47] investigate a setting with a certain environment, Hoskin [68] considers the assessment of opportunity costs in an uncertain environment. Seventeen years before the paper by Schweitzer and Cachon [122], the experimental study of Hoskin [68] had already addressed human behavior in the newsvendor problem. The results show that decision makers deviate from the order quantities that optimize expected profits. However, the study has a number of technical shortcomings which do not allow for deriving consistent and reliable results.[2] Since previous research has shown that decision makers underweight or even neglect foregone payoffs, Ho et al. [66] hypothesize that the psychological aversion to leftovers is greater than the disutility of stockouts. They develop and experimentally test a newsvendor framework where they add psychological costs of overordering and underordering. A main weakness of their additive approach is that an underweighting of foregone losses is modeled as additional (positive) psychological costs, which seems counterintuitive. Furthermore, the case in which decision makers neglect foregone payoffs is even incompatible. The question why in many situations decision makers underweight opportunity costs compared to out-of-pocket costs is addressed by Thaler [132]. He argues that the endowment effect supports the different weighting of these costs. While opportunity costs are often underweighted, other cost types tend to be overweighted by decision makers. McCaffery and Baron [101] refer to Richard Thaler's real-world observation: "when a gas station charged a 'penalty' for using credit cards ($2.00 versus $1.90, say), people paid cash; when a gas station across the street gave a 'bonus' for using cash ($1.90 versus $2.00), people used credit cards". McCaffery and Baron [101] state that, due to penalty aversion, individuals would rather avoid penalties than obtain bonuses. The tendency of people to avoid penalties is documented in several experimental studies and holds true in diverse economical contexts. For example, tax rules (see McCaffery and Baron [100]) or contracts (see Luft [94]) are less likely to be accepted when they are presented as penalties rather than as bonuses. The consequences of penalty aversion are decisions where penalty costs are higher weighted than out-of-pocket costs - another example that the different assessment of cost types can lead to a different behavior.

Involving different types of costs, a wide range of business decisions require that a decision is made before the occurrence of a random event. The underlying trade-off, concerning the costs of the mismatch between the de-

[2] The number of participants per experimental setting as well as the number of periods were too small, participants had to estimate the demand distribution based on the past data on demand, and some participants received changed information already after few periods. Furthermore, several product types had to be ordered and the margins of the products were chosen unfavorably.

cision and the realization, is captured by the newsvendor model. However, experimental studies of the newsvendor problem typically consider out-of-pocket costs (overage costs) and opportunity costs (underage costs) as mismatch costs. To investigate how the assessment of costs influences a decision maker, we consider two newsvendor situations involving different types of costs. Motivated by the literature, we expect an underweighting of opportunity costs and an overweighting of penalties. Therefore, we consider a situation where penalty costs (respectively additional reorder costs) instead of opportunity costs occur in the underage case. An example is a newsvendor situation involving a second order for an additional premium, as considered by Cachon and Swinney [22] where "the second order opportunity eliminates lost sales (...) [but] therefore, the penalty for ordering too little in the first order is that one may be required to purchase additional units in the second order at a higher cost." We refer to this kind of newsvendor problem involving out-of-pocket costs and penalty costs as the"penalty cost problem" whereas the classical newsvendor problem as considered by Schweitzer and Cachon [122] is referred to as the "opportunity cost problem". Since only the type of costs is different, the balancing problem remains mathematically identical and the decision maker is still facing the same underlying trade-off concerning ordering too little and ordering too much (see Cachon and Swinney [22]). Gavirneni and Isen [52] show that most people are able to compute the overage and underage costs accurately, but fail to determine the optimal inventory level. Therefore, a different behavior in the penalty cost and the opportunity cost problem implies that the assessment of costs changes for different cost types. Consequently, in order to investigate our research question we set up an experimental study where we differentiate between these two problems. Since previous research has shown that people anchor on the mean demand, we further distinguish between three cases with different margins for each of the two problems.

The main contribution of this research project is twofold. First, we systematically investigate how the assessment of costs influences a decision maker in a newsvendor situation. We propose a behavioral approach, including a higher weighting of penalty costs than of opportunity costs and order decisions which are biased towards the mean. Our model explains large portions of the observed behavior in our experimental study. A different weighting of costs can be seen as the main driver for higher order quantities in the penalty cost problem compared to the opportunity cost problem. Based on our findings, we identify situations in the newsvendor problem which are particularly unfavorable for the performance of a decision maker. Furthermore, our insights allow us to detect newsvendor situations where the behavioral effects partially compensate for each other and therefore lead to a better performance of decision makers. Second, our experimental study gives important insights into how people behave in newsvendor situations which are affected by penalty costs. For many business decisions, the underage case of the underlying newsvendor trade-off is influenced by penalty or reorder costs and not by opportunity costs. Typical areas where expensive reorders, contractual penalties or second production runs occur instead of lost profits are procurement problems if too little was ordered (e.g. Cachon and Swinney [22]), inventory problems if too little was stored (e.g. Eppen [41]), or production problems if too little was produced (e.g. Donohue [38]).[3,4] In order to apply the findings from behavioral studies of the newsvendor problem to a broad field of business situations, it is important to check validity and to identify limitations. The results of our study clarify that the behavior in a newsvendor situation which is affected by penalty costs is significantly different from the behavior in a situation which is influenced by opportunity costs.

Chapter 2 is organized as follows: Section 2.1 provides an introduction and a literature review before we define our hypotheses in Section 2.2. The experimental setup and design is described in Section 2.3, and we discuss the results in Section 2.4. Finally, in Section 2.5 we draw conclusions and discuss managerial implications.

[3] Identical to the opportunity cost problem, the overage case of the penalty cost problem involves out-of-pocket costs like production costs, holding costs, and purchasing costs.

[4] In a broader context, the penalty cost problem is also inherent in stochastic project management settings, such as the determination of feeding buffers (e.g. Trietsch [133]) or due dates (e.g. Zhu et al. [139]) assuming costs for starting activities earlier and tardiness penalties. Furthermore, a typical application in health care management is the reservation of operating room capacity under uncertainty considering costs for operating room time and overtime costs (e.g. Olivares et al. [108], Wachtel and Dexter [135]).

2.2 Definitions and Research Hypotheses

To investigate the influence of different cost types on decision making, we differentiate between two from a mathematical point of view identical newsvendor problems with the only difference that the type of costs in the underage case is different. We consider one situation where penalty costs occur, and one situation where opportunity costs occur. We expect a different behavior in the penalty cost and the opportunity cost problem.

The opportunity cost problem is identical to the classical newsvendor problem as described, e.g. in the paper of Schweitzer and Cachon [122] and in most follow-up newsvendor studies. A vendor orders goods for the next period where he faces an uncertain demand d. The cumulated demand distribution $F(D)$ is known. Purchasing costs per item are c and the selling price is p. Consistent with the newsvendor literature, we define the purchasing costs as out-of-pocket costs. If the demand exceeds the order quantity q, the foregone opportunity to make more profit by selling more products leads to lost sales and thus to lost profits which are also referred to as opportunity costs. The opportunity costs per item which cannot be delivered, termed as "underage costs", is $c_u = p - c$. If demand is less than the order quantity, assuming a salvage value of 0, the costs for each unit ordered too much, called "overage costs", are $c_o = c$.

Analogous to the opportunity cost problem, in the penalty cost problem a vendor orders q units for the next period where he faces an uncertain demand d with a known cumulated demand distribution $F(D)$. For each unit he orders before demand takes place, he has purchase costs of c (out-of-pocket costs). If the demand exceeds the order quantity, he has to reorder units for higher reorder costs of $s > c$ to satisfy the excess demand.[5,6] The costs for each unit ordered too little ("underage costs") are the additional "penalty" costs of the reorder, i.e. $c_u = s - c$. If demand is less than the order quantity, the costs for each item ordered in excess of the realized demand ("overage

[5] In contrast to Cachon and Swinney [22], we consider a reorder obligation instead of a reorder possibility. It is obvious that the second order should equal the unfulfilled demand. In order to avoid additional behavioral biases, we prefer to maintain a situation involving only one decision.

[6] An obligation to reorder may be interpreted as a commitment for a service level of 100%.

costs") are equal to the purchase costs, i.e. $c_o = c$.[7]

In both newsvendor situations, the expected costs of overestimating and underestimating demand have to be minimized. The only difference between both situations is the different type of costs in the underage case (see Table 2.1). The underage costs correspond to penalty costs in the penalty cost problem, while they correspond to opportunity costs in the opportunity cost problem. Since only the type of costs is different, we can determine the

	Penalty cost problem	Opportunity cost problem
Underage costs	Penalty costs	Opportunity costs
Overage costs	Out-of-pocket costs	Out-of-pocket costs

Table 2.1: Summary of cost types

"optimal order quantity" q^* for both problems with the classical newsvendor formula

$$q^* = F^{-1}\left(\frac{c_u}{c_u + c_o}\right) \tag{1}$$

with a problem specific definition of the underage costs as given above. By simple algebraic reformulation, we obtain

$$c_o \cdot F\left(q^*\right) = c_u \cdot \left(1 - F\left(q^*\right)\right) \tag{2}$$

which shows the trade-off a decision maker faces: The optimal order quantity q^* can be derived from balancing the probability of being over and under stocked weighted with the overage and underage costs, respectively. In order to depict a biased assessment of costs, we include the underage cost weight $\beta > 0$ which specifies how much the underage costs, relative to the overage costs, influence a decision. Since the overage costs are equal to the purchase costs in both problems, we scale these out-of-pocket costs with a weight of 1. An underage cost weight of $\beta > 1$ indicates that a decision maker has a

[7] Considering a reorder possibility instead of a reorder obligation does not change the overage and underage costs and therefore the optimal order quantity, given that the selling price is above the costs (see Eeckhoudt et al. [39]).

stronger weighting of underage costs relative to overage costs. An underage cost weight of $\beta < 1$ indicates that the decision maker weights the underage costs lower than the overage costs. We denote the consequences of the cost weight on the order quantity as the "assessment of costs effect" (ACE). To integrate the biased assessment of the different costs in the balancing problem, we extend Equation (2) by the overage cost weight 1 and the underage cost weight β and obtain

$$1 \cdot c_o \cdot F\left(q^{ACE}\right) = \beta \cdot c_u \cdot \left(1 - F\left(q^{ACE}\right)\right) \tag{3}$$

where q^{ACE} denotes the adapted optimal order quantity.[8] Reformulation of (3) leads to

$$q^{ACE} = F^{-1}\left(\frac{\beta \cdot c_u}{\beta \cdot c_u + c_o}\right). \tag{4}$$

We assume that the weight of the underage costs β depends on the type of costs only and not on absolute values. In the penalty cost problem, the underage costs correspond to penalty costs that occur because an expensive reorder has to be placed. Since individuals are trying to avoid penalties, we derive our first hypothesis with β_{pen} as underage cost weight in the penalty cost problem:

> *H1: In the penalty cost newsvendor problem, people have a higher weighting of penalty costs compared to out-of-pocket costs, that is* $\beta_{pen} > 1$.

In the opportunity cost problem, the underage costs have the character of opportunity costs. As decision makers tend to underweight opportunity costs, we derive our second hypothesis with β_{opp} as underage cost weight in the opportunity cost problem:

> *H2: In the opportunity cost newsvendor problem, people have a lower weighting of opportunity costs compared to out-of-pocket costs, that is* $\beta_{opp} < 1$.

[8] The incorporation of an underweight factor in the newsvendor model is similar to Chen et al. [26]. They show that the payment timing affects ordering behavior, and they can explain this behavior by the effect that decision-makers underweight order-time payments.

Our central research question is whether decision makers behave differently in the penalty cost and in the opportunity cost problem. On the one hand, we expect opportunity costs to be lower weighted than out-of-pocket costs and, on the other hand, we expect penalty costs to be higher weighted than out-of-pocket costs. This leads to our third hypothesis:

> *H3: The weighting of opportunity costs in the opportunity cost newsvendor problem is lower than the weighting of penalty costs in the penalty cost newsvendor problem, that is* $\beta_{opp} < \beta_{pen}$.

As many newsvendor studies have shown that human behavior depends on the margin, we consider several cases. In the opportunity cost problem, the margin is defined as $\frac{p-c}{p}$, while in the penalty cost problem the margin is defined as $\frac{s-c}{s}$. Therefore, the margins are equal to the critical ratios. We differentiate between a "high margin case" where the critical ratio exceeds 0.5, a "medium margin case" where the critical ratio equals 0.5, and a "low margin case" where the critical ratio is less than 0.5. Assuming symmetric demand distributions, this leads to optimal order quantities above, equal to, and below the mean demand. As we consider a medium margin case, we can discuss a situation where deviations from the optimal order quantity may not be solely explained by the mean anchor effect. Benzion et al. [11] proposed the following formula to consider the mean anchor effect (MAE) where the order quantity q^{MAE} is determined by a linear combination of the mean demand μ and the optimal order quantity q^* with mean anchor weight α:

$$q^{MAE} = \alpha \cdot \mu + (1 - \alpha) \cdot q^*. \tag{5}$$

For $0 < \alpha < 1$, the resulting order quantity is consistent with the mean anchor effect. We assume that the mean anchor effect is symmetric, so the strength of the shift towards the mean neither depends on the order quantity being above or below the mean, nor its distance from the mean. We further assume that the mean anchor weight is the same for the opportunity cost and the penalty cost problem. As the mean anchor effect has been documented in many previous studies, it can be assumed to have a significant effect on the order decision. This leads to our fourth hypothesis:

H4: The mean anchor effect exists, that is $0 < \alpha < 1$.

For an integrated model of human behavior, a combined consideration of both the assessment of costs effect and the mean anchor effect is needed. To model the human order decision, we combine both effects in a straightforward way. We denote the resulting effect as "combined effect" (CE). The logic of the combined effect is as follows: The assessment of costs effect leads to the adapted optimal order quantity considering the cost weights of the decision maker. This adapted optimal order quantity is biased by the mean anchor effect towards the mean, resulting in the order quantity q^{CE}. The formula for the combined effect is then

$$q^{CE} = \alpha \cdot \mu + (1 - \alpha) \cdot F^{-1} \left(\frac{\beta \cdot c_u}{\beta \cdot c_u + c_o} \right) \tag{6}$$

where, depending on the problem, β stands for β_{opp} in the opportunity cost problem and β_{pen} in the penalty cost problem, respectively.

We employ a 2×3 design where we combine two problems (opportunity cost problem, penalty cost problem) with three margin cases (high margin, medium margin, and low margin) and thus obtain six different combinations. To compare the opportunity cost and the penalty cost problem, we set the selling price p equal to the reorder costs s. As we consider the same purchase costs c in both problems, the critical ratios are equal. By assuming an identical demand distribution, we achieve the same optimal order quantities. This enables a clear comparison of human behavior in the opportunity cost problem and the penalty cost problem, as the identical optimal order quantity can be used as a reference point. To achieve the different margin cases, we vary the costs only.

Based on our hypotheses, we consider the consequences of the combined effect, including the assessment of costs effect and the mean anchor effect on the order decision. The higher weighting of penalty costs leads to an increase in the order quantity in the penalty cost problem, while in the opportunity cost problem the order quantity is reduced by a lower weighting of opportunity costs. Furthermore, the mean anchor effect leads to a shift towards the mean demand. The comparison given in Table 2.2 clarifies that based on our hypotheses, the order quantities of the penalty cost problem should exceed the ones of the corresponding opportunity cost problem in all margin cases.[9] We note that the assessment of costs effect and the mean

	Penalty cost problem			Opportunity cost problem	
	ACE	MAE		ACE	MAE
High margin case	$\Uparrow$	$\Downarrow$	$>$	$\Downarrow$	$\Downarrow$
Medium margin case	$\Uparrow$	$\Downarrow$	$>$	$\Downarrow$	$\Uparrow$
Low margin case	$\Uparrow$	$\Uparrow$	$>$	$\Downarrow$	$\Uparrow$

Table 2.2: Expected consequences for order quantities

anchor effect lead in the same direction in both the high margin case of the opportunity cost problem and the low margin case of the penalty cost problem, respectively. We therefore expect results in these situations that are especially far away from the optimal order quantity. On the other hand, in the remaining situations the two effects work in opposite directions; therefore, they should partially compensate for each other and thus the deviations from the optimal order quantity should not be as big.

2.3 Experimental Setup

To test our hypotheses, we set up a laboratory study using a 2×3 between-subjects design where we distinguish between six combinations of problem and case, as given in Table 2.2. In all six experiments, we examine a discrete uniform demand distribution with the boundaries 0 and 100. The realization of the demand was randomly drawn in advance and is used for all six experiments.

Furthermore, we consider the same critical ratio for the opportunity cost and the penalty cost problem in each case. The parameters are set to $s = 12$ in the penalty cost problem and to $p = 12$ in the opportunity cost problem.

[9] Extreme examples could lead to a situation where the assessment of costs effect leads to quantities above the mean in the low margin case or below the mean in the high margin case. In these cases the mean anchor effect will change direction as depicted in Table 2.2. The overall order quantities will still be greater in the penalty cost problem than in the corresponding opportunity cost problem.

The costs are set to $c = 3$ in the high margin case, to $c = 6$ in the medium margin case, and to $c = 9$ in the low margin case. The obtained optimal order quantities of $q^* = 75$, $q^* = 50$, and $q^* = 25$, are above, equal to, and below the mean demand of $\mu = 50$.

All experiments were conducted at the "Munich Experimental Laboratory for Economic and Social Sciences" (MELESSA). For every experiment, 25 separated PC terminals were ready to use. Participants were recruited from the subject pool of the MELESSA with the help of a recruitment-software. All participants were students without profound knowledge of the newsvendor problem, and they came from different fields of study. Each student participated in one experimental study only, and altogether 148 students participated in the six different experiments. We ran four experiments with 25 participants and two experiments with 24 participants. Despite an overbooking of 3, only 24 students participated in the high margin case of the opportunity cost problem and in the high margin case of the penalty cost problem. The experiment was programmed and conducted with the software z-Tree (see Fischbacher [46]). Before the experiments, the instructions were read aloud (see Appendix A.1). Every period started with a decision screen where the participants had to make their order decision. After every decision, they received information about the realization of the demand, their order quantity, and the resulting profit or costs of this period on the information screen. The profits or costs were displayed in "experimental currency units" (ECU). In all six experiments, the purchase decision was repeated for thirty periods. The duration of one experiment was about 45 minutes. Using control questions, we ensured that all subjects understood their job within the experiment. After completing the session, the accumulated earnings were paid privately and in cash. In all six experiments, we chose the factor and the fixed amount such that an income of €14 could be obtained if the optimal order quantity was placed in each period. The performance oriented compensation was explained in the instructions and therefore known in advance. Across all six experiments, the subjects earned on average €10.47 including a show-up fee of €4. The standard deviation was €1.72.

2.4 Results

2.4.1 General Results

As in previous studies, we observe average order quantities per period and over all periods which are significantly higher in the high margin case than in the medium margin case (one-tailed Wilcoxon, $p < 0.005$), and significantly lower in the low margin case than in the medium margin case (one-tailed Wilcoxon, $p < 0.005$). This holds true for the penalty cost problem (PCP) and for the opportunity cost problem (OCP). For each case, the average order quantities of the subjects are shown in the Figures 2.1, 2.2, and 2.3. For

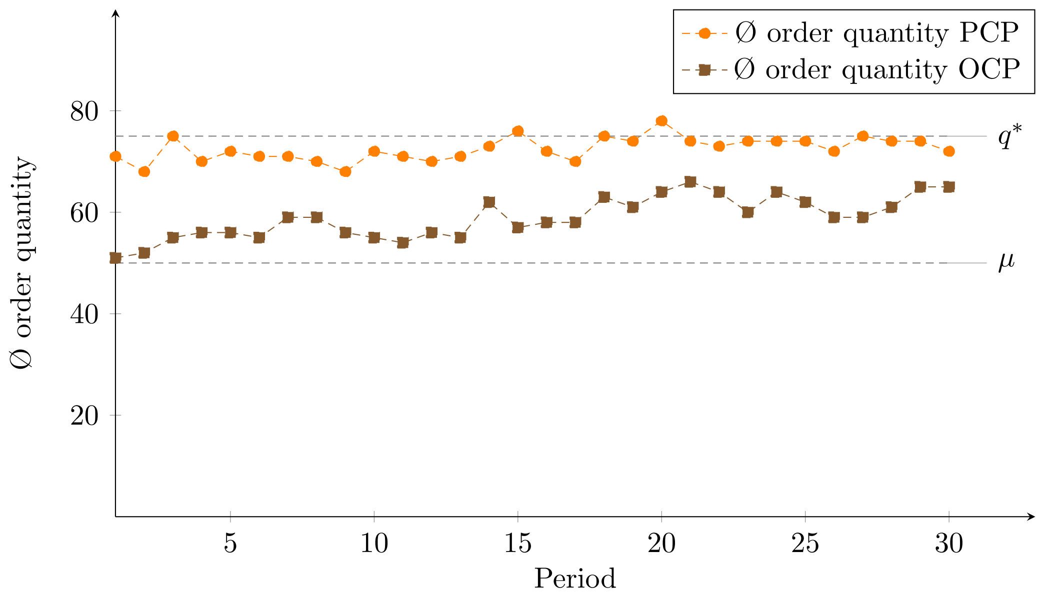


Figure 2.1: Average order quantities in the high margin case

both the penalty cost and the opportunity cost problem, Figures 2.1 to 2.3 illustrate that the average order quantities differ from the mean demand as well as from the optimal order quantities. As provided in Table 2.3, the difference is not significant only for the medium margin case of the opportunity cost problem. Furthermore, the average order quantities in the penalty cost problem significantly exceed the average order quantities in the opportunity cost problem for each margin case (see Table 2.3). Our results show that the order quantities are especially far away from the optimal order quantity

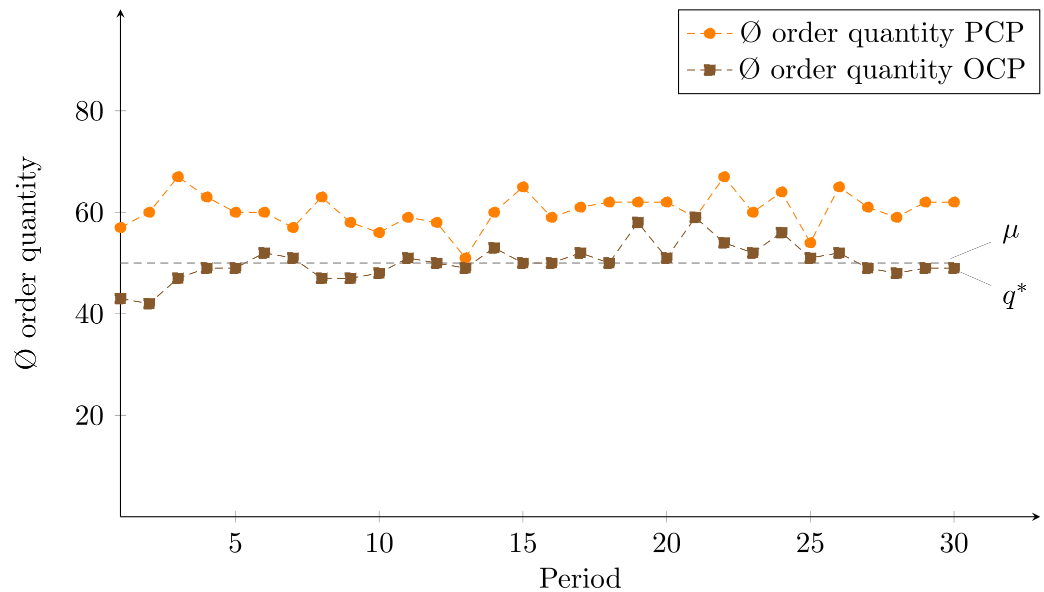


Figure 2.2: Average order quantities in the medium margin case

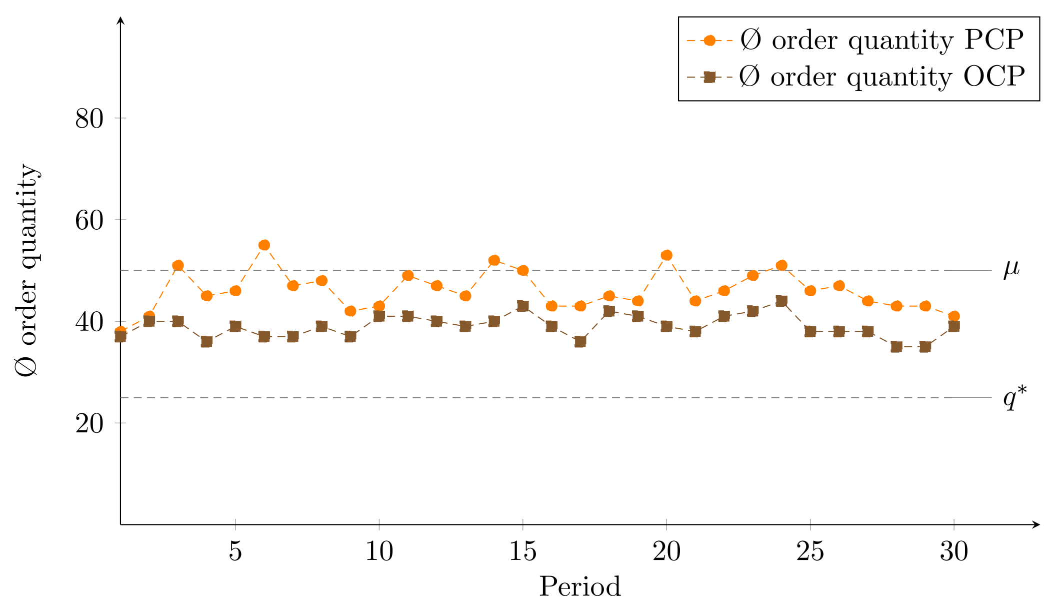


Figure 2.3: Average order quantities in the low margin case

in both the high margin case of the opportunity cost problem and the low margin case of the penalty cost problem. This is in line with our expectations outlined in Section 2.2.[10] To investigate learning effects, we conducted

	Problem	Optimal order quantity q^*	Mean average order quantity	Difference from μ	Difference from q^*	PCP is significantly higher than OCP
High margin case	PCP	75.0	72.6	$p < 0.005$	$p < 0.005$	$p < 0.005$
	OCP		58.9	$p < 0.005$	$p < 0.005$	
Medium margin case	PCP	50.0	60.3	$p < 0.005$		$p < 0.005$
	OCP		50.3	$p = 0.829$		
Low margin case	PCP	25.0	46.0	$p < 0.005$	$p < 0.005$	$p < 0.005$
	OCP		38.9	$p < 0.005$	$p < 0.005$	

Table 2.3: Wilcoxon test for average order quantities

a regression analysis on the average order quantities for each of the six experiments, where we define learning as a trend towards the optimal order quantity. As illustrated in Table 2.4, we observe significant learning in the high margin case, but there is no significant learning in the medium margin case and the low margin case. Over all six experiments no consistent learning or trend pattern can be approved.

We further analyzed how the average order quantities over the 30 periods of the decision makers are distributed. Figure 2.4 provides a box plot diagram

[10] Considering the high and the low margin case, we observe a too low/too high pattern for the penalty cost problem as well as for the opportunity cost problem.

	Penalty cost problem	Opportunity cost problem
High margin case	0.141 ($p < 0.005$)	0.367 ($p < 0.005$)
Medium margin case	0.061 ($p = 0.425$)	0.192 ($p = 0.009$)
Low margin case	-0.023 ($p = 0.783$)	0.000 ($p = 0.996$)

Table 2.4: Trend values of the regression analysis on the average order quantities

for each of the six experiments with the lower quartile, the median, and the upper quartile. The end of the "whiskers" show the lowest and the highest datum within the 1.5 interquartile range. Outliers are illustrated as well. Looking at the box plots, it is obvious that the average order quantities of the individuals (shown on the y-axis) are significantly higher in the penalty cost problem than in the opportunity cost problem for all three margin cases (one-tailed Mann-Whitney, $p < 0.005$ for all three cases). The box plots also clarify the systematic difference of the average order quantities in the three margin cases for both problems. Our results provide a good prediction of

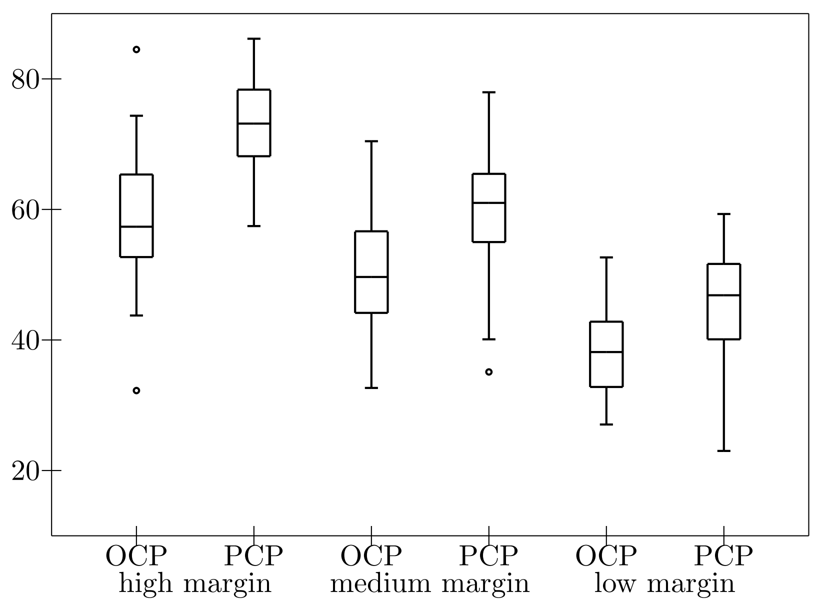


Figure 2.4: Box plot diagram of the average order quantities of the participants

the behavior of an "average" decision maker, since the box plots illustrate that most of the observed average orders of individuals are distributed closely around the median.

2.4.2 Testing of the Hypotheses

Our general results confirm that humans do not behave optimally in the newsvendor setting. However, their decisions are not random. For both problems, the level of the average order quantities in the different margin cases correspond to the level of the optimal order quantities. Furthermore, systematic differences between the penalty cost and opportunity cost problems exist. Consequently, we investigate whether a combined effect, including a higher weighting of penalty costs than of opportunity costs and order decisions which are biased towards the mean, is consistent with the observed behavior. Therefore, we adapt Formula (6) describing the combined effect to test our hypotheses and to evaluate our explanatory approach. Based on the six experiments, we estimate the three relevant parameters for Formula (7), wherein q_t is the average order quantity in period t. For the estimation, we integrated an error term ϵ_t.

$$q_t = \alpha \cdot \mu + (1 - \alpha) \cdot F^{-1} \left(\frac{\beta \cdot c_u}{\beta \cdot c_u + c_o} \right) + \epsilon_t \tag{7}$$

For the six experiments with 30 periods each, we estimate one common mean anchor weight α according to our assumptions in Section 2.2. For the three experiments in the penalty cost problem, we estimate a common penalty cost weight β_{pen} and for the three experiments in the opportunity cost problem, we estimate a common opportunity cost weight β_{opp}. The parameters are estimated using a least square estimation (R^2=0.88) where the test statistic follows asymptotically the standard normal distribution. We obtain $\beta_{pen} = 2.42$, and we can show that $\beta_{pen} > 1$ is significant (one-tailed z-test, $p < 0.005$). Consequently, the first hypothesis is verified: Penalty costs are higher weighted than out-of-pocket costs. As $\beta_{opp} = 0.95$, the opportunity cost weight is lower than one. However, the difference is quite small. Since $\beta_{opp} < 1$ (one-tailed z-test, $p = 0.073$), we find support for the second hypothesis: Opportunity costs are lower weighted than out-of-pocket costs. Furthermore, we can verify the third hypothesis as β_{opp} is significantly lower than β_{pen} (one-tailed Welch's t-test, $p < 0.005$): Penalty costs are higher weighted than opportunity costs. For the mean anchor weight, we obtain $\alpha = 0.49$.[11] The fourth hypothesis is also confirmed since we can show that $0 < \alpha < 1$ is significant (two-tailed z-test, $p < 0.005$) and thus a mean anchor effect exists.

Our results clarify that the mean anchor effect can be seen as the strongest driver for the non-optimal order quantities since the mean demand is weighted by almost 50%. However, the mean anchor effect cannot explain the large differences of the order quantities between the opportunity cost problem and the penalty cost problem. Hence, our results show that the different weighting of costs can be seen as the main driver for higher average order quantities in the penalty cost problem compared to the opportunity cost problem. Our approach leads to a mean absolute error of 3.2 (standard deviation of 2.4) concerning the order quantity.[12] This is very low compared to a mean absolute error of 11.1 when using the optimal order quantity as an estimator, of 10.0 when using the mean demand as an estimator, and of 6.2 when using the mean anchor effect (see Formula 5) as an estimator. The comparison demonstrates the high explanatory quality of our approach.[13]

2.5 Conclusion

Our research highlights differences in human decision making in situations involving different types of costs. Motivated by the literature, we expect that opportunity costs are underweighted compared to out-of-pocket costs while penalty costs are overweighted. In order to investigate the research question,

[11]This is in line with the results from previous studies in Western countries. Considering a high margin case and a low margin case of the opportunity cost problem, e.g. Bostian et al. [18] obtain a mean anchor weight of 0.47 and the data from Bolton and Katok [14] correspond to a mean anchor weight of 0.54.

[12]The remaining error can be partially explained by demand chasing, see Schweitzer and Cachon [122]. Since the underlying demand vector is identical in all experiments, demand chasing systematically influences the average order quantities per round.

[13]Our approach reduces the error by 68% compared to the mean demand as an estimator and by 71% compared to the optimal order quantity as an estimator. The isolated mean anchor effect reduces the error by 38% compared to the mean demand and by 44% compared to the optimal order quantity. This clarifies that the mean anchor effect explains only a part of the behavioral deviations.

we set up an experimental study in a newsvendor setting which provides a simple yet realistic environment to investigate the assessment of penalty costs and opportunity costs in two mathematically identical situations. To the best of our knowledge, we are the first to compare the assessment of these two cost types in an operations management setting. We observe that individuals order significantly more in a newsvendor setting with penalty costs than in a newsvendor setting with opportunity costs. We propose a behavioral approach which incorporates decision biases in the newsvendor model to explain the observed behavior. Besides the assessment of costs effect, we also include the mean anchor effect. Based on our approach, we tested our hypotheses and could confirm the mean anchor effect as well as a different weighting of different cost types. We found that penalty costs are higher weighted than opportunity costs. Our approach is valuable to predict actual ordering behavior and, furthermore, it allows us to quantify the extent of the psychological biases. Based on our findings, we conclude that the performance of a newsvendor depends clearly on the underlying situation. Situations where the assessment of costs effect and the mean anchor effect lead in the same direction result in particularly bad performance while situations where the two effects partially compensate each other result in a better performance of human decision makers. Since in many business decisions, the underlying newsvendor trade-off is influenced by penalty costs instead of opportunity costs, our study gives important insights in order to apply behavioral findings of newsvendor studies in a broader field. Typical examples where contractual penalties occur are inventory problems or purchase situations. It can be misleading to relate the insights from the opportunity cost newsvendor problem to situations incurring penalty costs instead of lost sales. This already happens, e.g. in research concerning the bullwhip effect (see Niranjan et al. [106]) or operating room planning (see Wachtel and Dexter [135]). The main finding of our research is that a biased perception of opportunity costs as well as a biased perception of penalty costs can explain the observed behavior. We show that decision makers are more sensitive to penalty costs than to opportunity costs. Consequently, we conclude that people have a different assessment of different cost types.

Our work has several limitations. We assume that the mean anchor effect is symmetric, even if the effect is stronger in the low margin context than in the high margin context in many studies. A first promising research considering the asymmetry in ordering behavior is done by Moritz [103]. He finds support that cognitive dissonance explains a portion of this behavior. As there are considerable differences in the asymmetry and since the asymmetry is even reverse in several studies (e.g. Ho et al. [66], Rudi and Drake [116], Lurie and Swaminathan [95]), further investigation is needed. If an asymmetry of the mean anchor effect could be validated and measured in terms of different mean anchor weights for different margin cases, it could be easily included in our approach. We agree with Bostian et al. [18] that the exploration of the asymmetry is one of the most promising directions for further research. Another interesting research area is cross-cultural differences between Western and Eastern countries. Even though the mean anchor effect is a predominant cross-cultural effect, Feng et al. [45] have shown that this effect is stronger in Eastern cultures. The investigation of differences and similarities in the assessment of costs between Eastern and Western cultures is a promising area for further research. Furthermore, the explanatory power of our approach could be increased by the integration of additional behavioral factors. Rudi and Drake [116] state that, besides the “level behavior”, the “adjustment behavior” can be seen as the main driver of the mismatch costs. Therefore, consideration of demand chasing could be worthwhile. Another example is the integration of a learning factor which would be especially useful for long-term consideration, e.g. 100 periods as investigated by Bolton and Katok [14]. These extensions would lead to a more complex approach but they could also enable an even more realistic description and prediction of the behavior.

Based on our approach and our findings, we conclude several managerial implications: From an internal company point of view, our insights could be used in a control process to detect situations which lead to systematic deviations from the optimal order quantity that are particularly unfavorable. Identifying these situations may allow corrective actions. Another internal aspect related to the planning process is that one could create situations where the deviations of the order decisions are relatively small, and therefore the decision maker performs better. From a supplier’s perspective, it may be possible to create situations where the behavior of the decision maker leads to deviations which are favorable for the supplier. Another important aspect for a supplier is to identify situations where the customer systematically orders too little. Consulting the customer may help to improve the situation for both, e.g. by a modified contract.[14]

[14] The performance of different contracting mechanisms in a two-echelon supply chain in which the retailer faces the opportunity cost newsvendor problem is investigated by Katok and Wu [74]. Based on our results, further research concerning the performance of contracting mechanisms in the penalty cost newsvendor problem as well as further research concerning the use of different cost types to increase contract performance in general would be interesting.

Chapter 3

Over- and Under-Utilization of Operating Rooms: Insights from Behavioral Health Care Operations Management

3.1 Introduction

Numerous recent studies encourage researchers to take into account human behavior in operations management (e.g. Loch and Wu [93], Gino and Pisano [58], Bendoly et al. [6]). Health care operations management has a particularly strong behavioral influence (see Brailsford and Schmidt [19]), since health care services are provided by people who may be influenced by cognitive biases, social preferences, and cultural norms (see Loch and Wu [93]). Even though people issues are vital for the processes in healthcare, very little research investigates the effects of human behavior on process performance in this industry. A promising opportunity to come up with more realistic health care operations management theories and to develop models which take into account human behavior is provided by experimental research. While behavioral experiments are a well-established research methodology for studying human issues in many disciplines including several business disciplines as well as medical research, combining findings from behavioral operations management with health care applications is a virtually untouched area.

In this study, we approach the field of behavioral health care operations management by investigating surgeons' behavior in the operating room (OR), one of the most important resources in hospitals. Guerriero and Guido [62] cite more than 100 studies on operating room management and Cardoen et al. [23] write "in the last 60 years, a large body of literature on the management of operating theaters has evolved". This comes as no surprise as around 40% of hospital expenses arise in the operating theater (see Denton et al. [35]) and more than 60% of hospital admissions are for surgical operations (see Peltokorpi [110]). Although the high importance of optimizing the usage of this scarce resource is evident, there is still much room for improvement. Rhodes and Barker [115] report poor utilization of ORs and Pandit and Carey [109] argue that 10-40% of all scheduled elective surgeries are canceled or rescheduled at least once. Sometimes OR managers make bad decisions resulting in staff working overtime (e.g. Wachtel and Dexter [135]).

Low OR utilization, rescheduling of surgeries and staff overtime are consequences of poor planning of surgery durations. Obviously, centralized planning cannot account for the specific patient knowledge of the responsible surgeon. Therefore, it is common practice in most hospitals that each surgery duration is planned independently by the surgeon in charge of the patient. In the literature surgeons' behavior is mainly discussed in the context of medical decision making. There are a few studies indicating non-optimal behavior of surgeons considering operating room management. Yule et al. [138] conduct a literature review on non-technical skills of doctors in the OR and they conclude that non-technical skills such as planning skills, resource management, and communication are often neglected, despite being vital for efficient OR management. Carter [24] presents an example where doctors only considered fairness when planning the ORs but ignored negative consequences for other units and Abouleish et al. [1] state that OR management is often based on convenience and tradition rather than on efficiency optimization. A systematic underestimation of surgery durations is found by Dexter et al. [36]. Wachtel and Dexter [135] discuss that the newsvendor model could be used to determine the time period, where staff is required in the OR. They also provide a literature review on behavioral newsvendor studies as they suspect biases known from the newsvendor model are present in the OR staffing problem as well. However, they do not account for differences between the operating room staffing problem and the inventory newsvendor problem. Furthermore, they do not carry out an experimental investigation. Hence, how the newsvendor model can be used in an OR setting is still an open issue.

Our research was motivated by a project with a medium sized hospital that asked for help with one of their major concerns - low operating room utilization and a high amount of overtime. The surgeons in charge of planning surgery durations received no guidelines or any information about the associated consequences of poor planning. Planned and realized durations were not monitored. In our study, we demonstrate the complexity of planning surgery durations based on empirical data from OR planning. We verify that variability in surgery durations exists and we analyze the negative consequences of planning too long and too short surgery durations. Thus we conclude that a newsvendor equivalent minimal cost model fits the problem of planning surgery durations. This is in line with several studies where the trade-off in surgery planning is modeled according to the newsvendor framework - even though the framework is rather stylized (e.g. Strum et al. [130], Olivares et al. [108], Wachtel and Dexter [135]). To test the behavioral effects of planning surgery durations we undertake an experimental study with senior surgeons. We chose doctors with experience in OR management since previous studies observed that, even if the direction of behavioral effects is the same, the magnitude of effects for students and experienced professionals as subjects may differ (e.g. Bolton et al. [15]). As no consistent definition of the consequences of planning too long and too short surgery durations exists in the literature, we employ an experimental study with two scenarios. Within our experimental study, we demonstrate significant non-optimal planning of surgery durations by experienced surgeons and we show that cost improvements of about 3.3% could be achieved in our setting. In a hospital with an annual budget of 100 million Euro and annual OR costs for staff and fixed capacities of about 30 million Euro, the savings potential is thus about 1 million Euro a year. Our research provides a better understanding of both the underlying problem and surgeons' behavior when planning surgery durations. Creating transparency of the problem and awareness of behavioral deviations is the first step in managing decision situations, such as communicating transfer prices for OR utilization or in developing debiasing methods to improve planning results.

The remainder of Chapter 3 is organized as follows: In Section 3.2, we present the problem framework of planning surgery durations and derive our hypotheses. In Section 3.3, the experimental setup is explained and the results are discussed. We draw conclusions and analyze managerial implications in the final Section 3.4.

3.2 Planning of Surgery Durations

Planning of surgery durations is a challenging task for surgeons since every patient is different, surgery durations are uncertain, and bad planning leads to undesirable consequences (e.g. May et al. [98]). To obtain first insight into planning behavior in real life, we analyzed 6 months (12/2011 - 05/2012) of surgery data from a German university hospital. The durations of elective surgeries are usually planned a few days before the surgery. We compare the planned and the realized durations of three exemplary operations from different specialties: Varicose veins crossectomy and stripping, cholecystectomy, i.e. the surgical removal of the gallbladder, and a specific joint fracture surgery. We compare the planned and the realized durations of these three surgeries in Figures 3.1 - 3.3 and present some additional information in Table 3.1.

	Crossectomy	Cholecystectomy	Joint fracture
Surgeries planned too long	83%	52%	29%
Surgeries planned too short	14%	47%	67%
Mean (st. dev) planned durations	90.0 (0.0)	63.3 (6.5)	66.0 (21.5)
Mean (st. dev) realized durations	68.1 (24.0)	65.9 (26.8)	79.9 (45.2)
Average plan deviation	+ 31%	- 4%	- 17%

Table 3.1: Comparison of planned and realized durations of three different surgeries

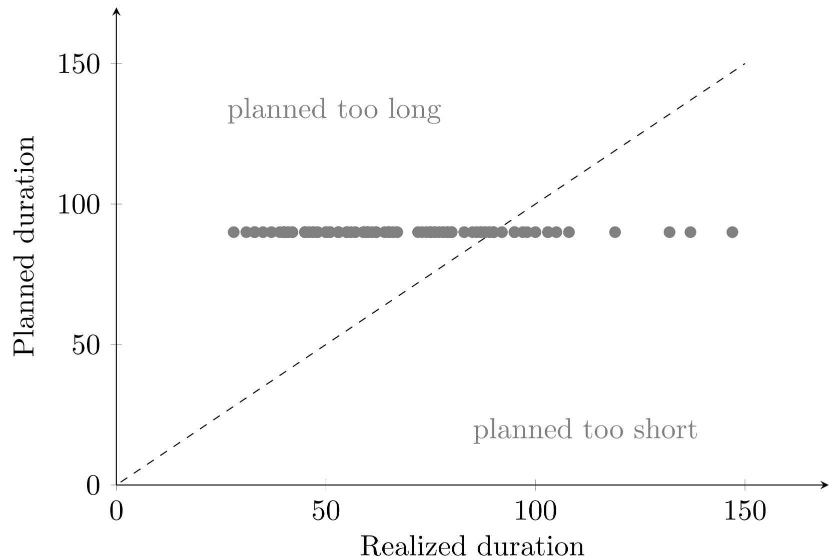


Figure 3.1: Comparison of planned and realized durations - Crossectomy

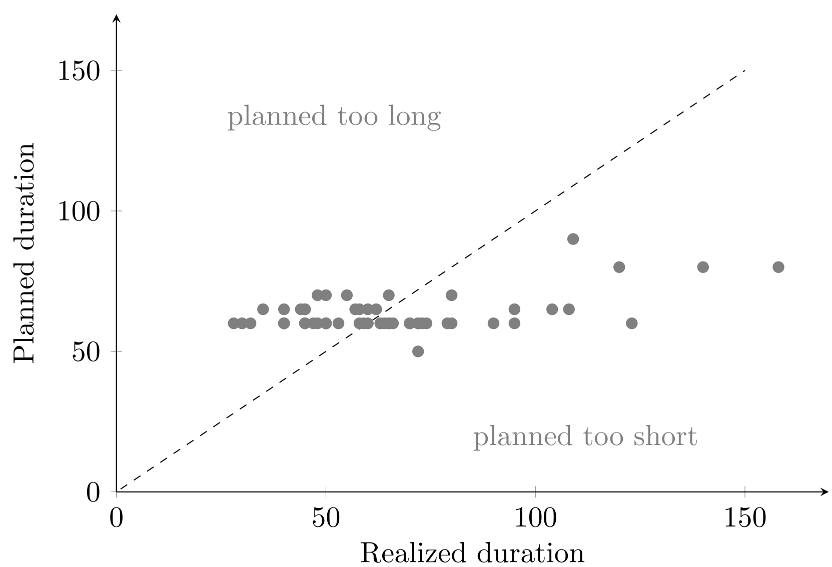


Figure 3.2: Comparison of planned and realized durations - Cholecystectomy

Figure 3.3: Comparison of planned and realized durations - Joint fracture

Crossectomy and stripping was significantly planned too long (one-tailed Wilcoxon, $p < 0.005$), cholecystectomy surgeries were on average planned close to the expected duration (two-tailed Wilcoxon, $p = 0.978$), and joint fracture surgeries were systematically planned too short (one-tailed Wilcoxon, $p = 0.030$). All surgeries have in common that the planned durations showed less variation than the realized ones. In fact, crossectomy and stripping surgeries were always planned with 90 minutes. We derive three main findings from these data. First, it is obviously not possible to always plan the exact surgery duration, as surgery times are stochastic. Second, different specialties seem to plan their surgeries in a different way, which may be a consequence of different cost structures. Third, some surgeries are systematically planned too long, while others are systematically planned too short.

Planning of surgery durations is a complex task due to two main characteristics of the problem. First, variability in surgery durations exists. Second, both planning too long and too short durations results in different negative consequences. As a result, a trade-off decision minimizing these consequences has to be made.

3.2.1 Variability of Surgery Durations

There are two reasons for variability in surgery durations: Uncertainty and "diversity of situation". Uncertainty in surgery durations is caused by many factors that cannot be predetermined. A typical example is unexpected bleeding that extends the duration. With diversity of situation we take into account a priori known factors, such as patient age or OR-team experience. Estimating the distribution of surgery durations is discussed widely in the literature. Strum et al. [129] and May et al. [99] use lognormal distributions to model surgery times, while Silber et al. [124] estimate surgical and anesthesia procedure times using data obtained from the US Medicare system. All these studies show that there is significant uncertainty in surgery durations. Furthermore, there are several empirical studies showing that surgeons' estimates do not meet the realized durations. Wright et al. [137] compared time estimates of software scheduling systems to those made by surgeons. Even though the software systems could not outperform the surgeons, modeling could help the surgeons to improve their time estimates. Eijkemans et al. [40] demonstrated that, in addition to the surgeons' estimates, diversity of situation factors such as surgery and team characteristics and, to a lesser extent, patient characteristics like age and body mass index proved to be relevant for surgery times.

3.2.2 Consequences of Planning too Long or too Short

The second driver of the complexity of planning surgery durations is that both planning with too long and too short time estimates for surgeries lead to undesirable consequences. If the realized surgery duration falls below the planned duration, OR idle time will be the consequence. In line with Strum et al. [130], we define this as underutilization. If the realized surgery duration is above the planned surgery duration multiple consequences might occur. Following surgeries may have to be rescheduled which involves a considerable organizational effort, reduces patient and staff satisfaction as well as medical quality. Furthermore, the scheduled surgery or following surgeries might end after regular working hours, i.e. staff works overtime. Overtime caused by planning too short surgery durations is defined as overutilization. To obtain insight into the consequences of inaccurate planning, we analyzed data from the hospital mentioned above. We performed regression analysis to determine the effects of planning too long and too short on OR under- and overutilization, respectively. In Figure 3.4 we relate for each OR and each day the number of minutes surgeries were planned too long with total operating time. We observe that the more minutes surgeries were planned

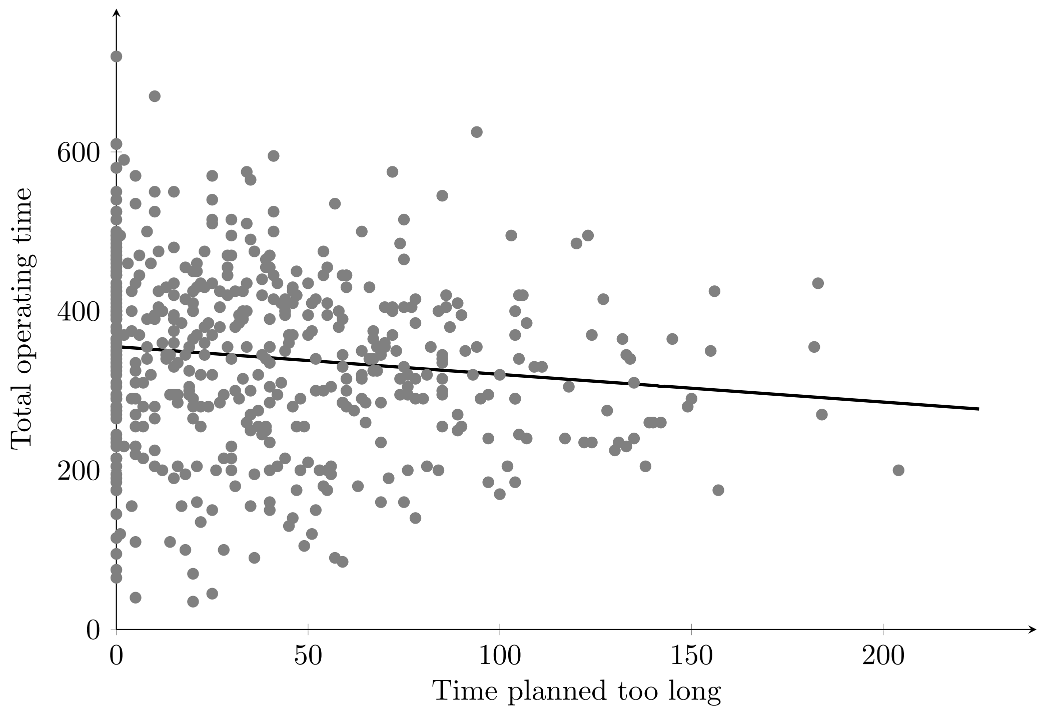


Figure 3.4: Consequences of planning too long (in minutes)

too long, the less was the total operating time (0.348 minutes of operating time per minute planned too long, $p = 0.007$) and thus the more idle time occurred. We further compared the number of minutes planned too short with the minutes of overtime (between 4pm and 10pm). As presented in Figure 3.5, the more minutes surgeries were planned too short, the more overtime occurred (0.483 minutes of OR overtime time per minute of planned too short, $p < 0.005$). Both underutilization and overutilization of ORs are associated with additional costs. Typically, costs for underutilization are created by idle OR and staff capacities, while costs for overutilization represent the additional overtime payments and costs for reorganizing the schedule. These costs can also include further negative effects on employee satisfaction (for working unplanned overtime or for being rescheduled) and patient satisfaction (for rescheduling their surgeries and for increased waiting times). Olivares et al. [108] state that "the costs of OR idle time were perceived, on average, as approximately 60% higher than the costs of schedule overrun," while Wachtel and Dexter [135] assume that the costs of OR overutilization are twice as high as the costs of OR underutilization. Thus, there is no clear ratio of these costs in the literature, which might be caused by different assessments of under- and overutilization in different hospitals.

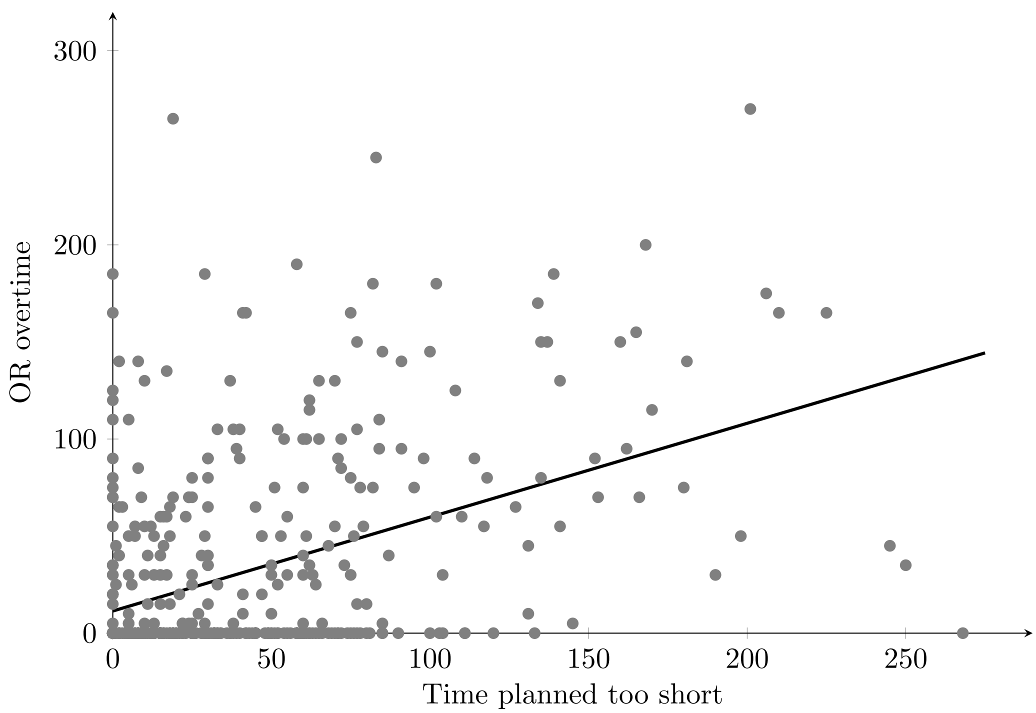


Figure 3.5: Consequences of planning too short (in minutes)

3.2.3 Minimal Cost Model

To minimize the expected costs of under- and overutilization, Strum et al. [130] propose a minimal cost analysis model. Dexter and Traub [37] define the sum of cost-weighted under- and overutilization as OR inefficiency. Although surgeons often perform a series of surgeries, each surgery duration is usually planned individually. In all three hospitals that cooperated with us for this study, it is common practice that surgery durations are planned independently by the surgeon in charge of the patient. Thus, we define a variation of the minimal cost analysis model concentrating on the costs for one surgery. c^u are the costs for each minute of underutilization, c^o for each minute of overutilization, and c for each minute of used OR capacity. Depending on the planned duration p and the realized duration D the OR inefficiency for one surgery is:

$$C(p, D) = c^u \cdot \max\{p - D, 0\} + c^o \cdot \max\{D - p, 0\} + c \cdot D. \tag{8}$$

The minimal cost analysis model is mathematically equivalent to the well-known newsvendor problem, which is also used for example by Olivares et al. [108] to conduct a structural estimation of the costs for OR under- and overutilization. As in the newsvendor problem the planned duration p^* that minimizes the expected costs $E[C(p)]$ is:

$$p^* = F^{-1}\left(\frac{c^o}{c^o + c^u}\right), \tag{9}$$

where F^{-1} denotes the inverse of the cumulative distribution function of the realized duration D. For the sake of brevity, we denote p^* as "optimal duration" and $\frac{c^o}{c^o+c^u}$ as "critical ratio" in the following. In both the minimal cost analysis model and the newsvendor problem, individuals face a decision under uncertainty with known distribution and a trade-off between planning (ordering) too long (too many) or too short (too little) durations (products) has to be made. The optimal solution can be derived analytically. On the other hand, the two problems are obviously not the same. Important differences are the main task - planning time versus quantities; the different context - operating room planning versus inventory settings; and the different decision makers - surgeons with no management training versus inventory managers. Furthermore, the consequences of not reserving enough time and not ordering enough quantities vary as well: Overtime with additional (penalty) costs in the OR case since surgeries have to be completed versus opportunity costs for lost sales in the inventory situation.

3.2.4 Hypotheses

The complexity of planning a surgery's duration is considerable, even if optimal durations can theoretically be derived with the newsvendor model. In everyday life surgeons lacking training in capacity management plan surgery durations. All studies using the minimal cost model have in common that a rational decision maker is assumed but they do not take into account that a human decision maker may not act rationally. As several studies show that people do not behave optimally in the related inventory situation (e.g. Schweitzer and Cachon [122], de Véricourt et al. [34], Moritz et al. [104]), and since some studies have observed biased surgeon behavior in general, we expect that surgeons do not plan optimally. Due to the similarities, we expect that some behavioral effects in the inventory problem can be found in the OR planning problem as well. One bias that is consistently found in all newsvendor studies is the mean anchor effect, where orders are too high when the optimum lies below mean demand and too low when the optimum lies above the mean demand. Schweitzer and Cachon [122] are the first to describe this pattern and they also discuss possible explanations for the observed behavior. An explanation they found support for is that decision makers use the mean as an anchor and only insufficiently adjust towards the optimal solution. This bias was replicated in numerous follow-up studies (e.g. Bolton et al. [15], Kremer et al. [82]). Combining these findings with our empirical observations for surgeons planning behavior for different surgeries and the OR literature (e.g. Abouleish et al. [1]) we derive the first hypothesis of our experimental study:

> *H1: Surgeons consistently plan too long (too short) in cases where the optimal duration* p^* *is below (above) the average duration* μ *of a surgery.*

Furthermore, we expect additional effects to those of classical newsvendor studies. In contrast to the classical newsvendor problem, where ordering too little results in lost profits, the consequences of planning too short in our context differ. Too short planning of surgeries is associated with additional costs, since operations have to be finished. Therefore, the planning of surgeries has similarities to a situation where penalties occur when ordering too little and demand has to be fulfilled. Schiffels et al. [120] analyze the impact of "penalty" costs instead of opportunity costs in a trade-off situation and find strong support that people are more sensitive to additional costs than to opportunity costs. In their experiments involving two situations with identical optimal solutions, decision makers made more of an effort to avoid underestimating when there are penalties associated than in a situation where opportunity costs occur. Considering this effect we derive our next hypothesis:

> *H2: Surgeons avoid overutilization rather than underutilization. As a consequence, planned durations are biased upwards.*

Decision behavior is often sensitive to task and contextual factors. Kremer et al. [82] compared a classical newsvendor situation (operations setting) with a context-free but mathematically equivalent neutral setting. They discovered that the bias towards the mean demand was much stronger in the operations setting than in the neutral setting. Furthermore, Bolton et al. [15] observed that even if the direction of behavioral effects is the same the magnitude of effects may differ for students and experienced managers. Therefore, context as well as professional background matters when considering behavioral biases. As doctors lack non-technical skills (see Yule et al. [138]) and focus rather on medical quality than on management decision making, we expect doctors to show stronger biases than inventory managers. Thus, even though newsvendor biases are likely to be relevant for our OR planning problem, biases in the OR context might be stronger than those in the inventory context. We derive our last hypothesis:

> *H3: Surgeons confronted with planning surgery durations perform worse than decision makers in comparable inventory newsvendor studies, i.e. the shift to the mean is stronger.*

We expect that the answers to these hypotheses will provide valuable insights into surgeons' behavior when planning surgery durations. Identifying and understanding behavioral biases in surgery planning could be an important step to improve the trade-off between overutilization and underutilization in hospitals. Increasing the OR utilization, or decreasing overtime and rescheduling, should not be driven by behavioral biases since both can affect the financial performance, patient satisfaction and medical quality.

3.3 Experimental Study

To investigate surgeons' behavior when planning surgeries we set up an experimental study. In Section 3.3.1 we describe the experimental setup and we discuss the results in Section 3.3.2.

3.3.1 Experimental Setup

The empirical findings from Section 3.2 show that based on the variability of surgery durations and the negative consequences of deviations from the realized duration the planning of a surgery can be described as a decision under uncertainty minimizing the expected costs for underutilization and overutilization. The variability of surgeries includes both uncertainty and diversity of situation. In the experiment we provide information about the stochastic distribution (i.e. uncertainty) of the surgery duration for a specific situation (i.e. surgery team, specific type of surgery, and patient characteristics). To avoid different assessments of the situation we neither communicate details nor change these diversity aspects. We apply a uniform distribution for simplicity since surgeons may lack in expertise in probability theory. This also increases the comparability to the inventory literature, where the uniform distribution is used in most studies even though normal or lognormal distributions better fit real life distributions. Benzion et al. [11] have shown that for an inventory setting the same behavioral effects are observed for different demand distributions. As discussed in the previous section, there is conflicting literature about the specific costs for underutilization and overutilization and these may differ between hospitals. To account for different situations on the one hand, and to be comparable to the literature on the other hand, we differentiate between two exemplary cases. The "low quantile case" with relatively high underutilization costs c^u indicates a hospital where idle capacities are of greater concern, while the "high quantile case" with relatively high overutilization costs c^o indicates a hospital where overtime is of greater concern. As our research question focuses on the behavior of surgeons when planning surgeries, we chose only doctors with relevant experience in scheduling surgeries as subjects. The experiment was carried out with 40 doctors from three German university hospitals, 20 in the low quantile case and 20 in the high quantile case. They were all senior physicians or chief physicians with an average age of 43 years. None of them had previous knowledge of the minimal cost analysis model. We used a between subject design with 20 participants in each treatment. All experiments were conducted in hospitals. We set up the experiments in a separated office room with a computer and we ensured that the physicians had no time pressure and that they were not disturbed or interrupted during the experiment. At the beginning of the experiment we provided the instructions (see Appendix A.2). The subjects were asked to schedule the duration of one surgery at a time. We provided information on the surgery duration, the OR costs c per reserved minute (underutilization costs $c^u = c$) and the increased costs per minute overtime s (overutilization costs $c^o = s - c$). For simplification, each minute scheduled too long results in a minute of underutilization, and each minute scheduled too short results in a minute of overutilization. Further details are depicted in Table 3.2. The experiment was programmed and conducted with the software z-Tree (see Fischbacher [46]). Either the low quantile case or the high quantile case was tested for each subject. After an initial screen, where the subjects had to enter a planned duration, feedback about the realized durations and the occurred costs was provided. The subjects performed 20 decision periods. The duration for each round was randomly drawn in advance and the same for all subjects. After planning the 20 surgery durations the subjects answered a questionnaire. The average duration of each experiment was 25 minutes. Money was the only incentive used. Payments were based on total costs and ranged between €19 and €39 with a mean of €33.2. Thus, the average payment matched the income of experienced doctors.

	Low Quantile Case	High Quantile Case
Distribution of surgery duration	$\mathrm{U}(100, 200)$	$\mathrm{U}(100, 200)$
Costs for planned time c	90	30
Costs for overtime s	120	120
Underutilization costs $c^u = c$	90	30
Overutilization costs $c^o = s - c$	$30(= 120 - 90)$	$90(= 120 - 30)$
Critical ratio	0.25	0.75
Optimal planned time	125	175

Table 3.2: Costs and optimal planning times for low and high quantile case

3.3.2 Results

To validate that the newsvendor model fits the reality in planning surgery durations, we conducted an ex-post assessment of our experiment by a questionnaire. The surgeons reported that cost pressure has a great impact on their decisions (average: 4.6 out of 7, 1 being no impact, 7 being huge impact) and they reported as well that they do encounter uncertainty when planning surgeries (average: 4.4 out of 7, 1 being no uncertainty, 7 being huge uncertainty). As the newsvendor problem is a cost-minimization model based on a trade-off decision under uncertainty, we conclude that the newsvendor framework is rather stylized, but appropriate for modeling planning surgery durations. This is also in line with the informal feedback we received from the surgeons after the experiment who stated that the experiment matched the trade-off situation they face daily.

As expected, we observed average planned durations for all subjects that are significantly higher in the high quantile case (HQC) (162.2) than in the low quantile case (LQC) (149.5) (one-tailed Wilcoxon, $p < 0.005$). In neither case did doctors plan the optimal duration. The box plots of the average planned durations per subject are presented in Figure 3.6. The average planned duration of all subjects is marked with a bold circle for both cases. The average duration of 150 minutes and the optimal durations for both the low quantile case (125 minutes) and the high quantile case (175 minutes) are represented by dotted lines. In Figure 3.6, it is apparent that the average planned durations per subject are closely distributed around the average planned duration of all subjects in both cases. In fact, they are approximately normally distributed (K-S test for normal distribution, low quantile case: $p = 0.714$, high quantile case: $p = 0.993$). As stated previously, the planned durations differ from the optimal duration in both cases. On average, in the low quantile case the planned durations are significantly above the optimal duration of 125 minutes (one-tailed Wilcoxon, $p < 0.005$) and close to the the mean duration of 150 minutes (two-tailed Wilcoxon, $p = 0.493$). In the high quantile case the planned durations are below the optimal duration of 175 minutes (one-tailed Wilcoxon, $p < 0.005$) and above the mean duration (one-tailed Wilcoxon, $p < 0.005$). Thus we can confirm Hypothesis 1: Surgeons consistently plan too long (too short) in cases where the optimal duration p^* is below (above) the average duration μ of a surgery. To measure the degree of non-optimality, we define in Equation (10) the relative cost increase as the

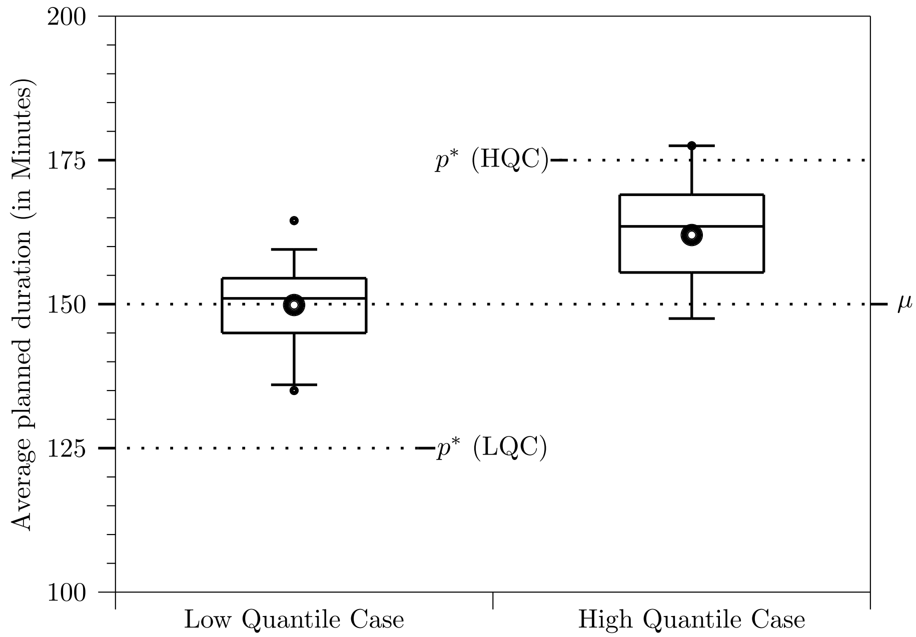


Figure 3.6: Average planned durations

percentage of avoidable costs due to non-optimal planning. For each decision of the subjects in our experimental study, we calculate the difference between the expected costs of the planned duration and of the optimal decision and divide this difference by the expected costs of the planned duration.

$$I(p) = \frac{E[C(p)] - E[C(p^*)]}{E[C(p)]} \tag{10}$$

We calculated the average relative cost increase according to Equation (10) for all participants and all rounds. We found an average relative cost increase of 3.3% in the low quantile and 3.4% in the high quantile case. The relative cost increase is slightly higher in the high quantile case, as the total costs are lower in this case. Consistent with inventory management studies, we further gained some insight into learning behavior (trend in the planned durations towards the optimum) and adjustment behavior (the tendency to adjust period-to-period the planned duration in the direction of the previous realized duration). Using linear regression for both the low and the high quantile case, no significant learning could be observed. This comes as no surprise, as Bolton and Katok [14] show significant learning effects in the long run only. In line with Kremer et al. [82], we found that subjects are more likely to adjust their planned duration in the direction of the previous realized duration than away from it in both cases. To test Hypothesis 2, we asked all subjects in the questionnaire administered after the experiment whether they had sought to avoid overutilization or underutilization. The results depicted in Table 3.3 show that in both cases the subjects tended to avoid overutilization. We expect this to result in planned durations that are

	Low Quantile Case	High Quantile Case
Avoid overutilization costs	65%	65%
Indifferent	15%	10%
Avoid underutilization costs	20%	25%

Table 3.3: Motivation when planning surgery durations

biased upwards. In the low quantile case, this should lead to planned quantities that are further away from the optimal duration, as the effect adds up to the bias towards the mean, than in the high quantile case, where both biases partially compensate for each other. In the low quantile case, the planned duration is on average 24.8 minutes above the optimal duration, while in the high quantile case, the planned duration is on average 12.8 minutes below the optimal duration. Therefore, the bias away from the optimal duration is significantly stronger in the low quantile case (one-tailed Mann-Whitney, $p < 0.005$). We confirm Hypothesis 2: Surgeons avoid overutilization rather than underutilization. As a consequence, planned durations are biased upwards. To test Hypothesis 3, we compared our data with the corresponding data from Schiffels et al. [120]. There, subjects were asked to order newspapers in a penalty cost based scenario with critical ratios of 0.25 and 0.75 and a uniform demand distribution between 0 and 100. In order to compare the results we shifted the data of Schiffels et al. [120] by 100. Therefore, differences in the order quantities/planned durations can be referred back to the different tasks, contexts, and professional backgrounds. As illustrated in Figure 3.7, the values were significantly lower in Schiffels et al. [120] (average: 146.0) compared to our study (average: 149.5) (one-tailed Mann-Whitney, $p < 0.005$) in the low quantile case. In the high quantile case, the values were significantly higher in Schiffels et al. [120] (average: 172.6) compared to our study (average: 162.2) (one-tailed Mann-Whitney, $p < 0.005$). In both cases planned durations were more strongly biased towards the mean in our

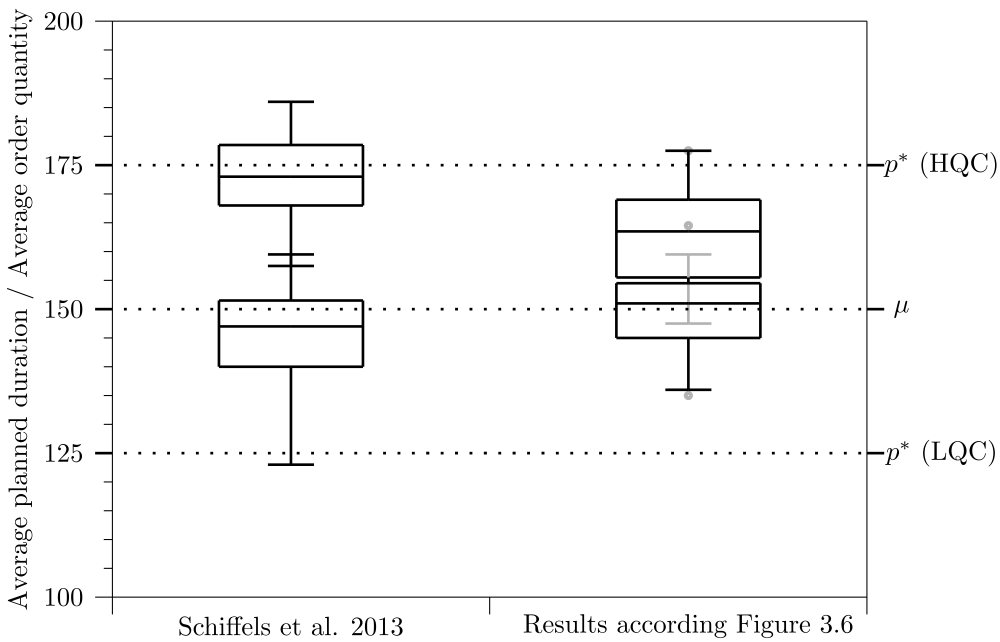


Figure 3.7: Comparison to the study of Schiffels et al. 2014

study. We conclude that different contexts, in our example reserving OR time versus ordering newspapers, and professional background do influence the behavior and confirm Hypothesis 3: Surgeons confronted with planning surgery durations perform worse than decision makers in comparable inventory newsvendor studies, i.e. the shift to the mean is stronger.

3.4 Conclusion

Many studies have shown that human behavior has a great impact on operations management decisions. Although the operating theater is the most expensive resource in hospitals, and its efficient usage is crucial, the behavior of health care decision makers in hospitals is generally ignored in research. It was challenging and took a tremendous effort to run experiments with experienced surgeons. However, we believe that this effort was necessary to gain acceptance with health care professionals and to close the gap between inventory management problems and the health care context. We could replicate biases known from previous newsvendor experiments in our study. Furthermore, we demonstrate that different tasks, contexts, decision makers, and the penalty cost situation lead to different results. Planned durations showed greater biases in situations where idle capacities are expensive compared to situations where overtime is expensive. Our study demonstrates that even in a simplified environment the planning behavior of surgeons is not efficient, systematic biases can be observed, and avoidable costs accrue.

Our work has several limitations and thus provides a staring point for future research. The newsvendor approach, and especially the simplified experimental framework, are stylized models of reality to investigate the behavioral effects considering the trade-off between planning too long and too short surgery durations. In practice, there are many other factors that influence the planning of surgery durations, such as capacity limitations (e.g. the only slot available is shorter than the desired time), scheduling restrictions (e.g. durations are only planned in 15 minute intervals), and interpersonal effects (e.g. the doctor with the succeeding surgery is particularly unhappy if there are delays). Further research on these factors might lead to a better understanding of planning behavior. Furthermore, we assume a cost minimization model to define the negative effects of overutilization and underutilization. Although this model is repeatedly used in the literature (e.g. Strum et al. [130], Olivares et al. [108], Wachtel and Dexter [135]), different hospitals might employ different incentive schemes. Future research could provide insights into whether these lead to different planning behavior, and which schemes are suitable to minimize OR inefficiency. Since surgeons may lack in expertise in probability theory and to be consistent with the inventory literature, we chose a uniform distribution of surgery durations in our experimental setup. Typically, surgery durations tend to follow a lognormal distribution (see Strum et al. [129]). Benzion et al. [11] show that in an experimental newsvendor setting, different distributions yield the same behavioral biases. However, a possible next step could be to set up experiments using historical data of surgery durations instead of distributions. In our experiment, we did not find any significant learning behavior but for long run experiments, we would expect small learning effects as described by Bolton and Katok [14]. The strongest learning effects can be expected if the time intervals between planning surgeries in the same situation are not too long. The same situation, i.e. the same combination of surgery type, OR team, and patient characteristics like age and body mass index does not appear that often during short time intervals. Therefore, the investigation of cross-learning effects, i.e. learning over a sequence of different surgeries, is a promising field for further research. As we gave full disclosure of the distribution in our experiment, we do not investigate how surgeons account for diversity of situation and concentrate on uncertainty of the duration. An interesting empirical research project would be to analyze surgeons' behavior considering their assessment of different information on diversity aspects.

As planning of surgery durations is a task of high economic impact for all hospitals, and as we have shown significant and systematic non-optimal behavior of experienced surgeons, important managerial implications may be derived. From our findings one can infer that in hospitals where idle capacities are more expensive than overtime, surgeons planned too long, while they planned too short when overtime costs exceed costs for idle capacities. Hospital management could react to these findings and create incentives for planning optimal surgery durations, develop debiasing methods to obtain better planning results, or improve the planning skills of surgeons with training. In the introduction, we mentioned a project in a hospital that triggered this study. The research described in this chapter helped this hospital in several ways: First, the consequences of planning too long and too short durations were analyzed and evaluated. Profitability issues the hospital was facing could be partly traced back to low OR utilization. Second, a target critical ratio was defined by the hospital management to decrease planned surgery durations in most departments in order to increase OR utilization. Third, the hospital management, OR management, and the surgeons in charge were informed about behavioral biases when planning surgery duration. Based on this, guidelines for planning surgery durations were defined. In this hospital seven out of ten medical specialties sharing the same OR resources systematically planned between 5% and 25% too long, while three specialties systematically planned around 5% too short. Each specialty was than provided with feedback whether they should plan more or less time than previously to meet the target critical ratio according to the new guidelines. Besides the recommendations the project helped the hospital management to gain a better understanding of the complexity as well as the behavioral biases of their employees. It also provided the surgeons with a better understanding of managerial targets.

Since the health care sector is the largest industry in industrialized countries in terms of number of employees, and as human decision making plays an important role, more research should be conducted in this field. We hope to encourage future research since we are convinced that many biases in the field of behavioral health care operations management are still to be discovered, and managing these biases could greatly impact the health care sector.

Chapter 4

Behavioral Portfolio Decision Making: Insights from an Experimental Study

4.1 Introduction

Selecting a subset from a discrete set of alternatives subject to various constraints is a ubiquitous problem in socio-economic decision making (see Kleinmuntz [78]). Research in the area of Portfolio Decision Analysis (see Salo et al. [118]) has brought forth a wide range of quantitative approaches to provide guidance for such problems. Frequently, decision problems in supplier selection (see Ho et al. [67]), new product development (see Cooper et al. [31]) and project portfolio selection (see Heidenberger and Stummer [65]) are considered. In contrast to numerous scientific publications on the subject, quantitative decision support approaches have only seen limited practical application (see Booker and Bryson [16], Cooper et al. [30], Loch [91]). Unique decision making environments, difficulties in evaluating alternatives and decision maker preferences, as well as the strategic orientation of decision problems, cause practitioners to rely on management expertise rather than utilizing elaborate quantitative decision support approaches (see Kester et al. [76], Martinsuo [97]). Thus the sole responsibility for portfolio decisions with grave impact often lies with human decision makers, who have been shown to behave irrationally in various decision environments (Bendoly et al. [6]).

Several studies from the past three decades have investigated human decision making in resource allocation problems (see Fasolo et al. [44]) as well as general project management settings (see Bendoly et al. [7]). To the best of our knowledge our study is the first to consider behavioral issues in the most basic setting of portfolio decision making, the knapsack problem (see Martello and Toth [96]). Here decision makers face a set of alternatives each characterized by their value and resource requirement, from which they must choose the subset with highest portfolio value and portfolio resource requirements not exceeding a given resource capacity. We propose a novel, generic experimental routine to investigate human portfolio decision making in an environment where subjects may dynamically select and deselect from a list of items to build their desired portfolio. Our experimental framework allows us to study both subjects' decision quality as well as their decision making process. The experimental results demonstrate that the complexity of even small instances of the knapsack problem is too high for people to solve them optimally, even when accounting for learning behavior. We propose that human decision making first focuses on selecting alternatives to construct an initial portfolio, which serves as a baseline solution for further improvement. Motivated by portfolio selection practice we investigate subjects' adherence to simple constructive heuristics considering either the value or resource requirement of the alternatives, as well as combinations of both. Subjects' behavior in our experiment is partially explained by adherence to two heuristics selecting items according to the maximum ratio of value divided by resource requirement and maximum difference between value and resource requirement. Furthermore, decision quality is shown to be affected by limitations on the amount of information that humans are able to receive, process, or retain. During portfolio construction subjects select items in close proximity to the previously selected item within the item list. The explanatory power of heuristics is increased by accounting for selection behavior which focuses only on a subset of alternatives.

Our goal is to create awareness of possible caveats of human decision making in portfolio decision environments by investigating decision biases. We strive to obtain greater understanding of decision maker heuristics in order to aid the development of debiasing strategies and effective decision support with greater compatibility to human decision processes (see Gigerenzer and Selten [55]). Even if the knapsack problem is stylized in comparison with most practical situations, insights into decision maker behavior in this baseline setting are fundamental. Building on our results, ample opportunities exist to study decision processes in more complex environments with e.g. limited availability of information, project dependencies, or group decision making to mention just a few.

The remainder of Chapter 4 is structured as follows. Section 4.2 reviews previous research on human behavior in portfolio selection and resource allocation problems, motivating our problem setting, which we introduce in Section 4.3. After describing our experimental framework in Section 4.4, we set up two experimental studies and explain the results in Section 4.5. We conclude our research with potential extensions and managerial implications in Section 4.6.

4.2 Related Literature

Fasolo et al. [44] review experimental and empirical studies dealing with behavioral issues in problem settings related to portfolio decision analysis. In particular they consider resource allocation problems, where variable amounts of resources have to be assigned to a set of alternatives. In the context of resource-allocation behavior Gingrich and Soli [56] conducted the earliest experimental study on human decision making in a problem setting solvable by linear programming. They investigate suboptimization when subjects solve a simple production planning problem. Busemeyer et al. [21] consider a similar problem setting and study subjects learning behavior as well as the impact of conveying information about the performance to subjects. Langholtz et al. [89] summarize a series of publications extending the work of Gingrich and Soli and Busemeyer et al.. Langholtz et al. [85] examine a multi-period resource-allocation problem under certainty, risk and uncertainty. Langholtz et al. [86, 87] study how subjects cope with possible resource breakdowns and abundance of resources. Langholtz et al. [88] consider a three-dimensional resource allocation problem which is solvable by integer programming while Ball et al. [2] apply a verbal protocol analysis technique to examine cognitive strategies used by participants. Gonzalez et al. [60] examine resource-allocation problems where the goal is to achieve a fixed objective while minimizing resource consumption. A different related field of research studies human behavior in financial decision making. Funk et al. [49] as well as Rapoport [114] consider a financial portfolio selection problem, where funds have to be allocated to a risky and risk-less asset in a multistage betting game. Kroll et al. [84] test the application of a mean-variance model for portfolio selection and report on an experiment, where subjects assign investment capital to two independent assets with stochastic returns. Their findings show considerable suboptimal results by subjects due to cognitive biases as well as only limited learning effects of the subjects. Another line of research considers portfolio selection problems in the context of decision support systems. Gettinger et al. [53] and Stummer et al. [131], for example, focus on the effect of different visualization techniques of multi-criteria evaluations of alternatives.

We conclude that previous research efforts have focused on practical decision making settings and have mainly considered continuous resource allocation problems. None of these previous studies directly addresses behavioral heuristics and biases when solving portfolio selection problems, where a subset from a discrete set of alternatives has to be selected. In line with other studies from the domain of behavioral operations management (see Bendoly et al. [6]) we want to address a controllable and basic problem setting, whose investigation provides fundamental insights into a wide range of concrete problem settings.

4.3 Decision Maker Behavior in the Knapsack Problem

To investigate decision making behavior in a basic portfolio setting we consider the knapsack problem (see Martello and Toth [96]). A set of N items is given, with vector $v \in \mathbb{R}_+^N$ indicating values and vector $k \in \mathbb{R}_+^N$ indicating required resources of the items. The objective is to choose a subset of items with maximum sum of values, i.e. portfolio value, while the sum of required resources, i.e. portfolio resource requirement, must not exceed resource capacity $c \in \mathbb{R}$. This problem can be formulated as a binary optimization problem

$$\max_{x \in \{0,1\}^N} \left\{ v^T x | k^T x \leq c \right\} \tag{11}$$

with decision variables $x \in \{0,1\}^N$ indicating the selection ($x_j = 1$) or exclusion ($x_j = 0$) of item $j \in \{1, \ldots, N\}$. While the knapsack problem is NP-hard, dynamic programming approaches exist to solve it in pseudo-polynomial time (see Martello and Toth [96]).

Human decision makers will generally not be able to solve a NP-hard optimization problem as the knapsack problem to optimality. In many settings it has been proven that decision makers employ simple heuristics instead of performing complete optimal searches due to their limited mental capacities in handling complex tasks (see Loch and Wu [93]). We expect suboptimization even in small problem instances with tens of alternatives, which is common for real-life project portfolio selection problems (see Golabi et al. [59], Loch et al. [92], Gurgur and Morley [63], Lindstedt et al. [90]). Based on the complexity of the problem, we further expect that decision makers will not overcome suboptimization through learning by repetition.

In practical portfolio selection problems, decision makers typically base their decisions on alternatives' value and resource requirement (see Salo et al. [118]). These project characteristics may be combined to derive further metrics as net present value or discounted cash flow, which consider the difference between value and resource requirement (see Heidenberger and Stummer [65]). Alternatively, the ratio of value and resource requirement is considered, following the "value for money" principle (see Skaf [126], Phillips and Bana e Costa [112], Phillips [111]). It is intuitive to apply simple decision rules with "evaluation criteria" based on value and resource requirement in the knapsack problem, and obvious heuristics that decision makers might use to construct a portfolio are to select items with maximum value (MaxV), minimum resource requirement (MinK), maximum ratio of value divided by resource requirement (MaxR) or maximum difference between value and resource requirement (MaxD). We define a "constructive heuristic" as a decision making approach which selects items based on the ranks of all available items according to an evaluation criterion. In iterative steps $s \in \{1, \ldots, N\}$ the highest ranked item according to a specific evaluation criterion $h(\mathcal{A}_s)$ is selected from set $\mathcal{A}_s$ of unselected items, whose selection does not cause portfolio resource requirement to exceed resource capacity. With $x^s \in \{0,1\}^N$ denoting the partial portfolio containing selected items $h(\mathcal{A}_1)$ through $h(\mathcal{A}_s)$ and the initial empty portfolio $x^0 = \vec{0}$ the following algorithm gives the selection process for constructive heuristics.

Algorithm 1: Constructive Heuristic

Initialization $x^0 = \vec{0}$
repeat $s = 1, 2, \ldots$

$$x_j^s = \begin{cases} 1, & \text{if} \quad j = h(\mathcal{A}_s) \\ x_j^{s-1} & \text{else} \end{cases} \qquad \forall j \in \{1, \ldots, N\}$$

where $\mathcal{A}_s = \left\{ j \in \{1, \ldots, N\} | x_j^{s-1} = 0, k^T x^{s-1} + k_j \leq c \right\}$

until $\mathcal{A}_s = \emptyset$

The evaluation criteria for constructive heuristics MaxV, MinK, MaxR and MaxD are given by

$$h^{\text{MaxV}}(\mathcal{A}_s) = \arg\max_{j \in \mathcal{A}_s} \{v_j\} \tag{12}$$

$$h^{\text{MinK}}(\mathcal{A}_s) = \arg\min_{j \in \mathcal{A}_s} \{k_j\} \tag{13}$$

$$h^{\text{MaxR}}(\mathcal{A}_s) = \arg\max_{j \in \mathcal{A}_s} \{v_j / k_j\} \tag{14}$$

$$h^{\text{MaxD}}(\mathcal{A}_s) = \arg\max_{j \in \mathcal{A}_s} \{v_j - k_j\}. \tag{15}$$

We refer to a portfolio as a "complete portfolio" if the portfolio value cannot be increased by additionally selecting any unselected and resource-feasible item. This holds for all portfolios resulting from one of the constructive heuristics. Furthermore, we define a human's selection process until the first complete portfolio is achieved as the "construction phase". In contrast to the constructive heuristics, the human decision making process might involve deselection steps, through which people might further adjust a complete portfolio achieved in the construction phase. We define the phase of decision making after establishing a first complete portfolio as the "improvement phase".

It is well-known that people apply decision rules in their decision making process (see Gino and Pisano [58], Gans and Croson [50], Bendoly et al. [6]), in many problem settings, e.g. the secretary problem (see Seale and Rapoport [123]), the newsvendor problem (see Schweitzer and Cachon [122]), or revenue management problems (see Bearden et al. [3]). We therefore assume that decision makers apply heuristics in the knapsack problem as well. The requirements of the decision maker's cognitive system for all four heuristics are low. To investigate whether people rely on constructive heuristics in portfolio selection problems, we formulate hypotheses a) to d) for the MaxV, MinK, MaxR and MaxD heuristics.

> *H1: During the construction phase the selection process of decision makers is based on*
> *a) the MaxV heuristic,*
> *b) the MinK heuristic,*
> *c) the MaxR heuristic,*
> *d) the MaxD heuristic.*

It remains to be investigated which heuristic is used.

Since constructive heuristics require ranking all available items in every step, with a theoretical worst-case complexity of $O(N^2)$ (see Knuth [79]), this task becomes increasingly difficult for decision makers with increasing instance size. Ericcson et al. [42] emphasize the limited capacity of human short-term memory, placing constraints on the human ability to process information for problem solving. While Miller [102] claims that the limit on the capacity for processing information is about 7 elements, Cowan [32], reviewing a wide range of studies, proposes a limit of 4. Although quantifying human mental capacity is a matter of debate, there is no doubt that only a limited amount of information can be bound into one functional context (see Jonides et al. [72]). We assume for the knapsack problem that decision makers' ability to keep track of all available items is limited as well. We distinguish between ideal "global selection behavior", which considers all available items, and "localized selection behavior", which due to cognitive limitations considers only a subset of all available items. For small problem sizes, this subset might be equal to the complete set of items but we expect that the impact of localized selection behavior increases with increasing problem size.

> *H2: Decision makers apply localized selection behavior.*

In order to investigate our research hypotheses, we set up two experimental studies. The first study provides general insights into the performance of decision makers, as well as how the performance changes with the problem size. We furthermore derive initial results regarding the use of heuristics.

Based on the findings of the first study, we determine an adequate problem size to study the selection process in more detail in a second study, focusing on the most promising heuristics.

4.4 Experimental Framework

4.4.1 Experimental Routine

Subjects are presented with the experimental task of solving instances of the knapsack problem in subsequent rounds. Money is the only incentive offered and participants receive a variable monetary payout, linearly dependent on the final portfolio value that they achieve in the knapsack problem. In each round participants are given a list of items, their values and resource requirements as well as the available capacity. This information is instance-dependent and may differ in each round.

Within a given time limit, subjects may freely select and deselect items from the list of items. The portfolio value and remaining resource capacity is updated after each decision. Subjects are informed if they try to select an item whose resource requirement exceeds the remaining capacity and the attempted selection is denied. Subjects are provided with a calculator as well as pen and paper. The maximum duration of each round is five minutes but participants are free to (irrevocably) proceed to the following round at any time. Results of preliminary studies have shown that 5 minutes are enough time for subjects to not perceive any time pressure for the problem instances that we consider. Participants are presented with new rounds until the experiment ends after exactly 35 minutes and therefore the numbers of rounds that subjects play is dependent on the time taken for each round. At the end of the experiment, for each subject one completed round is randomly drawn to determine the payout. Portfolio values are converted to Euro using round-dependent conversion factors, communicated to participants in each round. A subject's payout is obtained by reducing the converted portfolio value of the drawn round by a fixed charge of €100. The conversion factor and fixed charge are chosen so that participants can achieve a maximum payout of €20, when solving the knapsack problem to optimality. We hereby ensure that differences in performance result in clear differences in payout.

Subjects receive at least a show-up fee of €3.

Instructions, read by subjects before the experimental rounds begin, are given in Appendix A.3 and contain a screenshot illustrating an exemplary problem instance. Subjects are furthermore presented with 3 practice rounds consisting of 25 items and a fixed duration of 5 minutes each. Practice rounds are not considered in the incentive scheme. At the end of the session participants are asked to fill out a short questionnaire and are informed of their performance in each round as well as the resulting payout.

4.4.2 Problem Instances

Pisinger [113] as well as Smith-Miles and Lopes [127] discuss parameters to gauge the difficulty of knapsack problems. We consider these metrics when designing the knapsack instances for our experiments.

Pisinger [113] evaluates different solution procedures for knapsack instances of varying difficulty. In a computational study, the author considers knapsack problems with between 50 and 10,000 items. Knapsack instances, whose values and resource requirements are independently drawn from the same range $[1, R]$, are termed "uncorrelated" instances. "Weakly correlated" instances are obtained by randomly drawing resource requirements from the range $[1, R]$ and sampling the value of each item $j \in \{1, \ldots, N\}$ from the reduced range $[max(1, k_j - R/10), k_j + R/10]$. For "strongly correlated" instances the value is fixed to $v_j = k_j + R/10$. Solution times of algorithms are shown to increase with higher correlation of item values and resource requirements, as well as increasing number of items. Solution times also increase when increasing the data range from $[1, 1{,}000]$ to $[1, 10{,}000]$.

Smith-Miles and Lopes [127] discuss features impacting algorithmic performance in several combinatorial optimization problems. In addition to the correlation of item values and resource requirements discussed by Pisinger [113], they characterize knapsack problems by their "constraint slackness", the ratio of available budget to the sum of the resource requirements of all items. For slackness levels close to 0, only a few items may be selected without violating the budget restriction, while for slackness levels close to 1, almost all items may be selected, reducing the problem to which items not to select. Such problem instances are assumed to be easier to solve, while Chvátal [29] proposes slackness ratios of around 0.5 for difficult instances.

In our experimental study we want to present subjects with challenging knapsack instances, ensuring that we neither overwhelm their mental capacity nor present trivial problems. Studies investigating knapsack algorithms commonly consider instances consisting of hundreds or thousands of items. Such dimensions are not adequate for investigating human decision making, due to limited cognitive capacity as well as limited time in experimental settings. In order to give subjects enough time to solve several different knapsack instances, we examine considerably smaller instances consisting of 5, 10, 15 and 25 items. We opt to use weakly correlated instances in our experiments as Pisinger [113] argues that they represent real-world knapsack problems in the most realistic way. Weakly correlated instances typically do not contain "trivially" selectable items with very high values and small resource requirements, which frequently occur in uncorrelated instances. All values and resource requirements fall into the range $[1, 1{,}000]$ and values are higher than resource requirements for all items. In line with Chvátal [29], slackness levels are set to between 0.4 and 0.6. At these levels, knapsack instances have the highest number of complete portfolios, limiting subjects' possibilities to achieve an optimal solution by random selection. In order to distinguish the human decision behavior clearly, we ensure that each instance has a unique optimal solution and that the four constructive heuristics considered lead to different, unique and non-optimal solutions. In order to ensure unique selections according to heuristics, within an instance no two items may have the same value, resource requirement, difference between value and resource requirement, or ratio of value divided by resource requirement. The order in the list of items is random and equal for all subjects.

4.5 Experimental Studies

The experimental studies were performed at the "experimenTUM" laboratory of TUM School of Management. The experiments were programmed and conducted with the software z-tree (see Fischbacher [46]) and were administered using the software ORSEE (see Greiner [61]). We follow the experimental routine explained in Section 4.4.1 and employ knapsack instances based on the specifications given in Section 4.4.2. We conducted two multi-round experiments in order to test our hypotheses and to obtain insights into human portfolio decision making through quantitative and statistical analysis. The first experimental study (Study 1) focuses on the degree of suboptimization with regard to the instance size, learning behavior as well as the comparison of subjects' selected complete portfolios to heuristic portfolios. The second experimental study (Study 2), employing test instances with different specifications, focuses on subjects' selection processes in the construction phase. We investigate adherence of selection steps to constructive heuristics as well as localized selection behavior.

4.5.1 Exposition

A preliminary one-round study, independent from Study 1 and Study 2, was conducted to provide initial visual insights into human portfolio decision making. A subject's decision making process, i.e. the attempt to solve a knapsack problem through selection and deselection of items, can be visualized by plotting the value and resource requirement of (partial) portfolios resulting from each selection and deselection step. While opting for weakly correlated problem instances in Section 4.4.2 due to their practical relevance, such instances are unsuitable for proper visual inspection since differences in value and resource requirement between portfolios typically are too small. Furthermore, instances with more than ten items usually have 100,000 or more different feasible portfolios, when considering slackness levels of 0.5 as proposed in Section 4.4.2, preventing visualization as well. In this one-round pretest, 16 undergraduate business students were asked to solve a knapsack instance consisting of 10 items with uncorrelated values and resource requirements. The budget level was set to half the sum of resource requirements of all items. Subjects with the highest aggregate value received a fixed payout of €10. The modest number of feasible portfolios for this instance allows a clear graphical representation of the resource requirement and value solution space, while subjects can still select 53 different complete portfolios. We provide an illustration of subjects' decision making behavior in Figure 4.1. Despite the small instance size, differences in selection behavior can be seen. The decision making process of each subject, containing all selection and deselection steps, is unique. Participants typically terminate the decision making process with the selection of a complete portfolio. Although

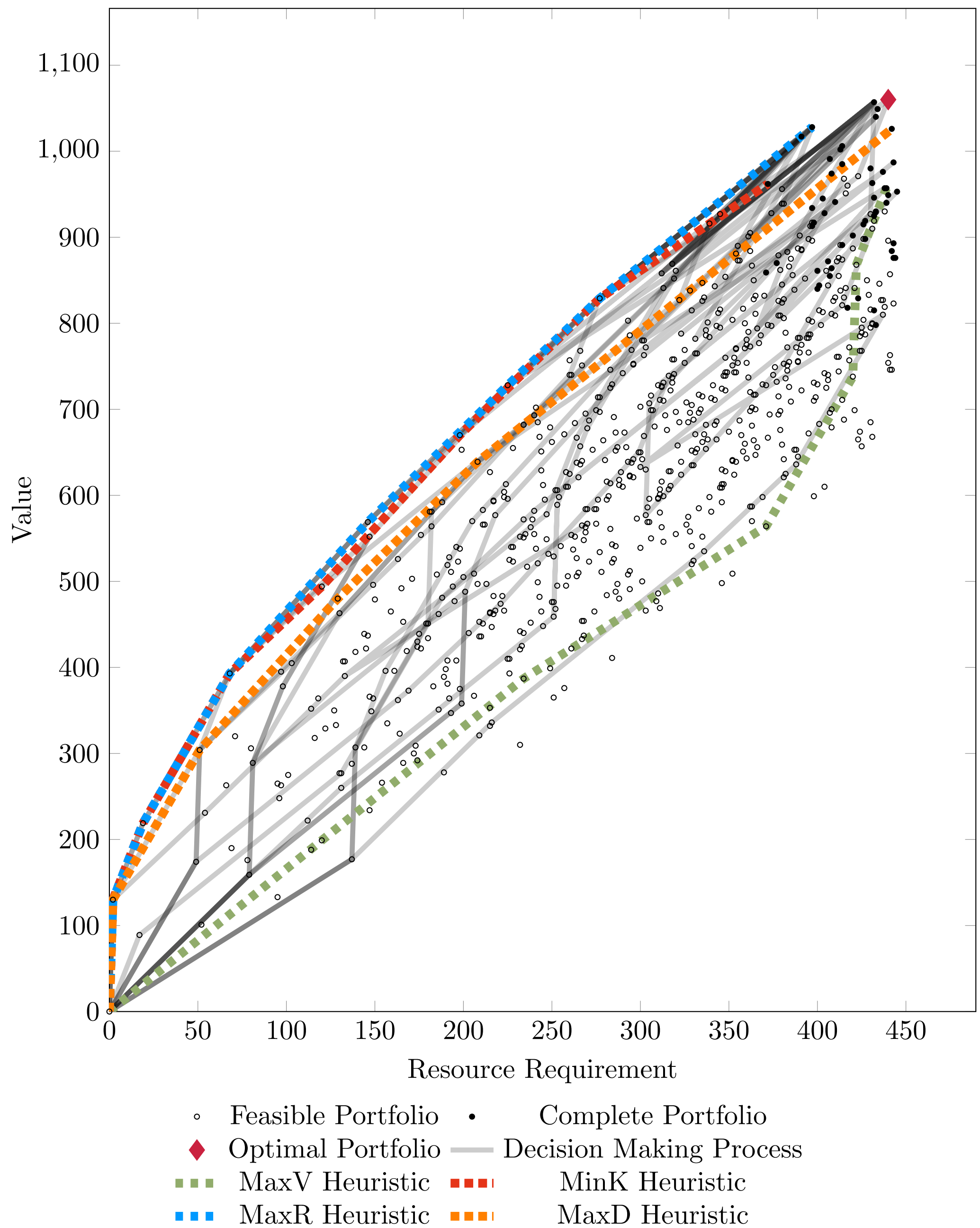


Figure 4.1: Feasible portfolios, subjects decision making processes, and the heuristic selection processes

these portfolios are of high value, Figure 4.1 indicates suboptimization for most subjects. While subjects' behavior partially overlaps with the selection processes of the four constructive heuristics, introduced in Section 4.3, participants do not strictly follow them. It remains to be investigated whether parts of subjects' selection processes can be explained by the constructive heuristics.

4.5.2 Experimental Protocol of Study 1

Study 1 has been conducted with 29 undergraduate business students in 2 separate experimental sessions. The rounds are based on knapsack instances with 5, 10, 15 or 25 items. In order to investigate learning behavior, rounds 3 through 8 are based on the same two instances, one instance with 15 items and one instance with 25 items, which are repeated alternately. To prevent subjects from noticing similarities between these rounds, all item values, resource requirements and budgets are multiplied by an integer "scaling factor" from the range $[1, 3]$ and furthermore, the sequence in which items are presented to the subjects is "randomly permuted". Table 4.1 summarizes the specification for the different rounds of Study 1, where distinct rounds with similar specifications like round 9 are shown until the experiment ends after 35 minutes. Including time to read instructions, time for three prac-

	Round									
	1	2	3	4	5	6	7	8	9	...
Instance	1	2	3	4	3	4	3	4	5	...
N	5	10	15	25	15	25	15	25	25	...
Scaling factor	1	1	3	2	1	1	2	3	1	...
Random permutation	-	-	-	-	✓	✓	✓	✓	-	...

Table 4.1: Specifications for the consecutive rounds of Study 1

tice rounds, as well as time to fill out the questionnaire, each session took approximately 60 minutes. At the end of the sessions subjects were paid in private, earning on average €10.97 including a show-up fee of €3.00 with a standard deviation of €3.28.

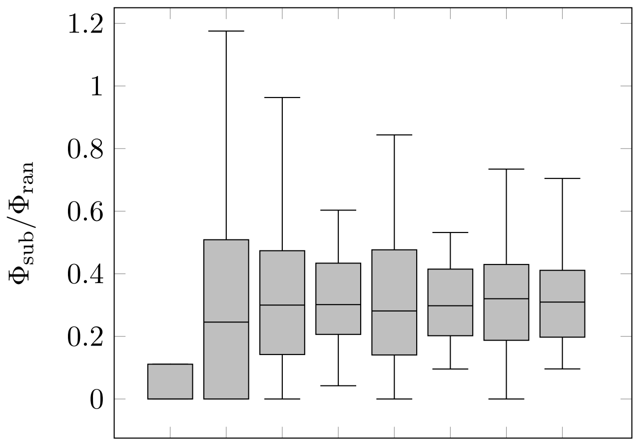
4.5.3 Results of Study 1

Our analysis focuses on rounds 1 to 8, which were completed by all participants. Out of 232 decision making processes for all 29 subjects and the 8 rounds, 5 processes are excluded as participants in these rounds did not end with the selection of a complete portfolio.

We evaluate the quality of subjects' final portfolio choice using the percentage deviation of the obtained from the optimal portfolio value, where p_{opt} and p_{sub} denote the optimum portfolio value and the portfolio value generated by a subject, respectively.

$$\Phi_{\text{sub}} = \frac{p_{\text{opt}} - p_{\text{sub}}}{p_{\text{opt}}} \tag{16}$$

As we wish to assess the quality of subjects' decision making, we compare Φ_{sub} to Φ_{ran}, the expected optimality gap if items are randomly selected until a complete portfolio is obtained. We calculate Φ_{ran} by sampling 10,000 complete portfolios with Monte Carlo simulation. For each sample, we apply algorithm 1 selecting items with equal probability in each step. The relation of Φ_{sub} to Φ_{ran} measures the performance of subjects compared to a random selection process which serves as a benchmark. Figure 4.2 reports on the distribution of $\Phi_{\text{sub}}/\Phi_{\text{ran}}$ values for all subjects. For reasons of clarity, outliers are not illustrated in the box plots. In line with our expectations, subjects show suboptimization as indicated by $\Phi_{\text{sub}}/\Phi_{\text{ran}}$ values greater than 0. Suboptimization can already be observed for very small instances consisting of 5 items, with a percentage of optimally solved instances ($p_{\text{sub}} = p_{\text{opt}}$) of 59%, and becomes more prominent with increasing instance size. In the following we focus on the instances with 15 and 25 items where subjects predominantly show suboptimal behavior. While few subjects succeed in finding the optimal solution for the instances with 15 items, problems consisting of 25 items are not solved to optimality by any subject. We consider both groups of instances separately when assessing learning behavior across all subjects through linear regression of $\Phi_{\text{sub}}/\Phi_{\text{ran}}$. The slopes of the trend lines, $b_1 = -0.02$ for $N = 15$ ($R^2 = 0.01$) and $b_1 = 0.00$ for $N = 25$ ($R^2 = 0.00$), show that suboptimality is maintained on a similar level throughout the experiment. Low trend values and R^2-values demonstrate that there are no learning effects in the short run. Decisions of subjects do not improve merely by repeating a similar instance



Round	1	2	3	4	5	6	7	8
Instance	1	2	3	4	3	4	3	4
N	5	10	15	25	15	25	15	25
$p_{\text{sub}} = p_{\text{opt}}$	59%	31%	7%	0%	4%	0%	4%	0%

Figure 4.2: Boxplot charts of $\Phi_{\text{sub}}/\Phi_{\text{ran}}$ for all subjects and rounds 1 - 8

several times.

In order to analyze subjects' decision making behavior in more detail, we consider the number of selection and deselection steps. As a matter of course, the number of selection steps increases with increasing instance size as shown in Figure 4.3. In our experimental framework, decision makers may deselect items freely, giving them the opportunity to correct previously made decisions. We analyze selection and deselection steps separately for the construction phase, i.e. the decision making process leading to a subject's first complete portfolio, and the subsequent improvement phase, i.e. all steps undertaken after a complete portfolio has been obtained. Considering instances with $N = 15$ (round 3, 5, and 7), the construction phase has more than twice as many steps (mean 8.91) as the improvement phase (mean 4.07). While during the construction phase more selection steps (mean 8.45) are performed than deselection steps (mean 0.46), this relationship is almost equalized (means 2.06 and 2.01) in the improvement phase, indicating that after a complete portfolio has been obtained, on average, one item is removed from the portfolio in order to add a new one. Considering larger instances with $N = 25$ (round 4, 6, and 8), the average number of selection

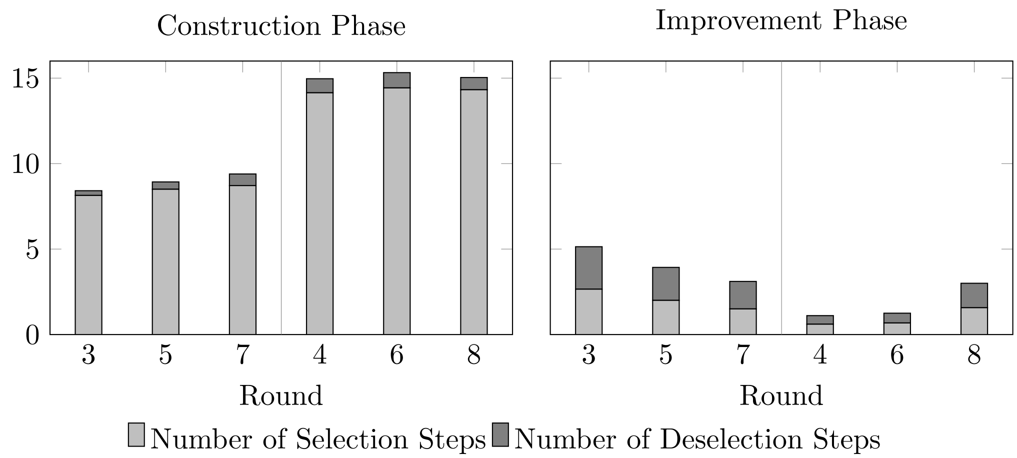


Figure 4.3: Mean number of selection and deselection steps in the construction phase and the improvement phase

and deselection steps in the construction phase (mean 15.11) is higher, while there are fewer steps in the improvement phase (mean 1.79). For both the construction phase and the improvement phase, the relationships between selection and deselection steps are similar to the instances with $N = 15$. Analyzing decision making patterns of selection and deselection steps, we find that the first selection step following one or more deselection steps often considers an item previously deselected, i.e. decision makers annul their previous deselection. Likewise, we find that the last selection step before one or more deselection steps often considers an item deselected in the subsequent deselection steps, i.e. decision makers annul their previous selection. Both behavioral patterns are caused by decision makers annulling their previous decision and we define them as "annulment patterns". The two patterns, annulling the previous selection or annulling the previous deselection, can also coincide if, for example, an item is selected, deselected and immediately reselected. Table 4.2 gives the percentage of selection and deselection steps which are associated with annulment patterns for the construction phase and the improvement phase, separated for $N = 15$ and $N = 25$. For all rounds under consideration, less than 18% of the selection steps and more than 76% of the deselection steps of the construction phase can be explained by annulling. This is in line with our expectation that decision makers follow a constructive heuristic during the construction phase, where deselection steps

Round	N	Construction Phase		Improvement Phase	
		Selection	Deselection	Selection	Deselection
3		15.52	100.00	59.09	65.15
5	15	13.57	90.00	45.10	67.35
7		17.04	87.50	52.94	69.23
4		11.18	88.89	66.67	76.92
6	25	8.21	76.19	61.11	75.00
8		9.44	91.67	54.55	66.67

Table 4.2: Percentage of steps associated with annulment patterns within the construction phase and the improvement phase

are only undertaken to correct erroneously selected items. In contrast, in the improvement phase, between 45% and 77% of both selection and deselection steps are associated with annulment patterns. Decision makers optimize their existing complete portfolio primarily by iteratively deselecting and selecting items, in many cases annulling their previous selection or deselection.

Additional information on subjects' decision making processes are obtained by analyzing the structure of complete portfolios at the end of the construction phase and the improvement phase. In line with algorithm 1, constructive heuristics iteratively select items ranked highest according to an evaluation criterion. We measure the relative frequency of the ith highest ranked item according to the evaluation criteria (12) - (15) included in subjects' first complete portfolios as well as their final complete portfolios. Since rounds within the two groups $N = 15$ and $N = 25$ are rescaled and resorted versions of the same instance, the ranking of the items is identical for the rounds in each group, and we aggregate the frequencies of the ranked items included in subjects' portfolios for all rounds in each group. We present the relative frequencies for both groups $N = 15$ and $N = 25$ and all evaluation criteria (12) - (15) in Figure 4.4. Furthermore, we illustrate the relative frequencies of items being included in all possible complete portfolios as comparison. For the h^{MaxV} and h^{MinK} evaluation criteria, no rank dependent differences between all complete portfolios and subjects' portfolio choices can be observed. In contrast, we find that subjects choose highly ranked items and omit low-ranked items for the h^{MaxR} and h^{MaxD} evaluation

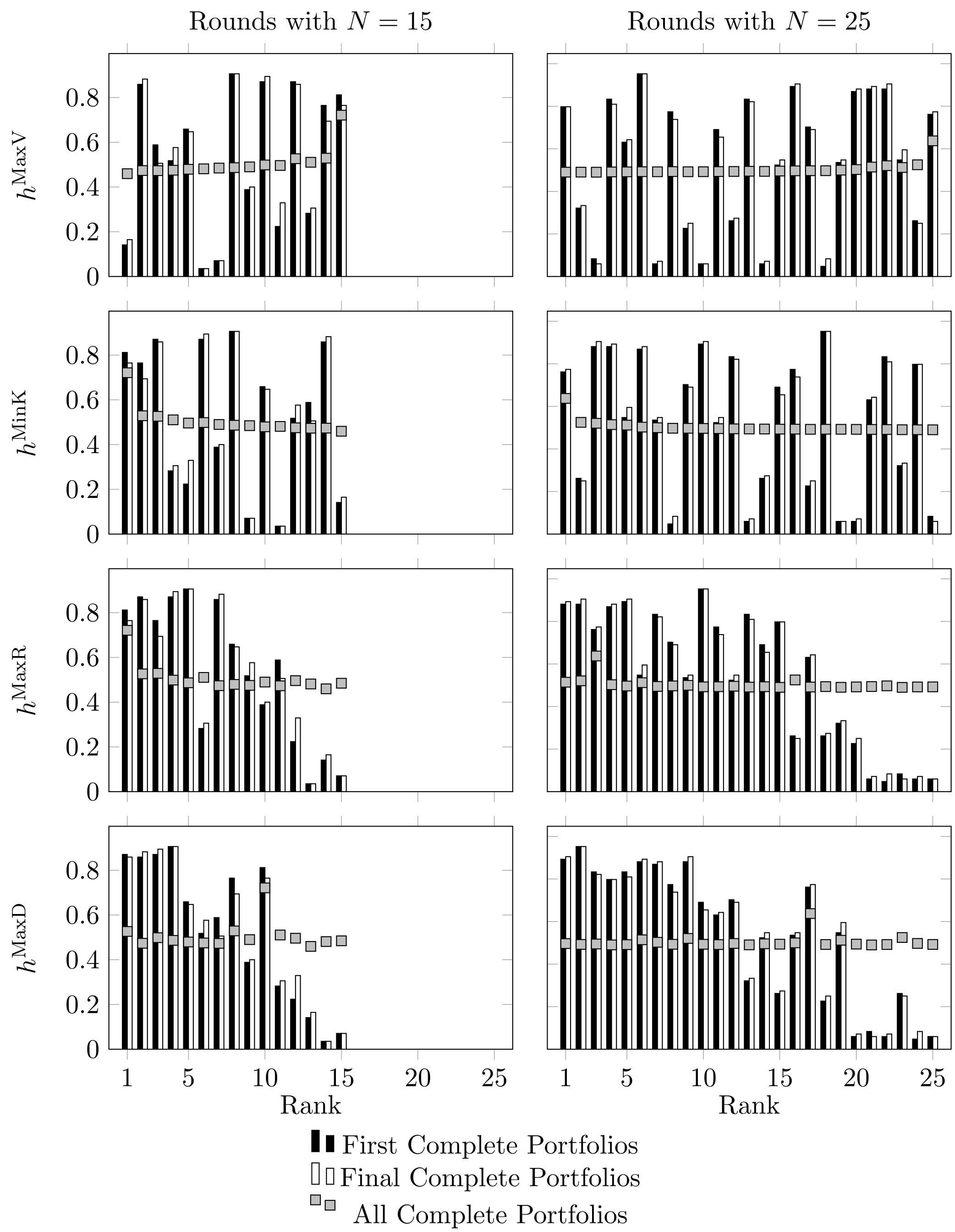


Figure 4.4: Selection frequency of the ith highest ranked item for all complete portfolios as well as subjects' first and final complete portfolios

criteria, and therefore subjects' behavior leads to systematically different portfolios as expected for random behavior for both criteria. Considering the first complete portfolios, the regression analysis provided in Table 4.3 verifies that subjects put stronger emphasis on items with high ratio of value divided by resource requirement and high difference between value and resource requirement in the construction phase as there is a significant negative trend b_1 for both evaluation criteria. Furthermore, Figure 4.4 indicates that the relative frequencies change only slightly between the construction phase and the improvement phase. The final portfolios of both phases differ on average by only 0.76 items for $N = 15$ and 0.37 items for $N = 25$.

Evaluation Criterion	Rounds with $N = 15$			Rounds with $N = 25$		
	b_1	R^2	p	b_1	R^2	p
h^{MaxV}	0.02	0.06	$= 0.398$	0.01	0.02	$= 0.475$
h^{MinK}	-0.02	0.09	$= 0.283$	-0.01	0.06	$= 0.231$
h^{MaxR}	-0.06	0.72	< 0.001	-0.04	0.74	< 0.001
h^{MaxD}	-0.06	0.79	< 0.001	-0.04	0.79	< 0.001

Table 4.3: Regression statistics for the frequencies of selected items, ranked corresponding to the four evaluation criteria

We conclude that the results from Study 1 demonstrate suboptimization even in small instances as well as no learning behavior in the short run. Based on the number of selection and deselection steps, we can verify that subjects behave differently in the proposed construction phase and improvement phase. While the construction phase is primarily based on a straight forward selection of items until a complete portfolio is obtained, it seems natural that the improvement phase is characterized by a similar number of deselection and selection steps. Analyzing subjects' selected complete portfolios, we find items with high ratio of value divided by resource requirement and high difference between value and resource requirement to be over-represented. Observing only a few deselections, this indicates that decision makers selection process is in line with the MaxR and the MaxD heuristic. While we focus on the complete portfolios in the first study, we concentrate on the selection process in the construction phase in the second study.

4.5.4 Experimental Protocol of Study 2

For Study 2, we consider a new set of instances to specifically investigate subjects' selection processes and the use of heuristics in the construction phase. We only consider instances with $N = 25$ in order to extend the number of selection steps in the construction phase. Study 2 has been performed with 53 undergraduate business students in 3 separate sessions. Participants from Study 1 were excluded from taking part in Study 2. Since we also wish to investigate whether the sequence in which items are presented to the subjects has an effect on their selection behavior, we repeat the instances from rounds 1 to 4 in rounds 5 to 8, while reversing the order in which items are presented to subjects. In order to prevent subjects from noticing the repetition, item values, resource requirements, and budgets are scaled by an integer from the range $[1, 3]$ for one out of the two related rounds, respectively. Furthermore, to separate subjects' use of the MaxR and MaxD heuristic, we ensure that the complete portfolios of both heuristics have as few joint items as possible ("heuristic distinction") for some rounds. For the instance of round 1, out of 16 items in the portfolio according the MaxR heuristic 7 items are also in the portfolio obtained by the MaxD heuristic, and, for round 3, out of 15 items in the portfolio derived from the MaxR heuristic 8 items are in the portfolio according the MaxD heuristic, as well. Table 4.4 summarizes the specification for the rounds of Study 2. After round 9 different instances

	Round									
	1	2	3	4	5	6	7	8	9	...
Instance	1	2	3	4	2	1	3	4	5	...
N	25	25	25	25	25	25	25	25	25	...
Scaling factor	2	3	1	1	1	1	3	2	1	...
Reversed order	-	-	-	-	✓	✓	✓	✓	-	...
Heuristic distinction	✓	-	✓	-	-	✓	✓	-	-	...

Table 4.4: Specifications for the consecutive rounds of Study 2

with the same specifications as in round 9 are presented until the experiment ends after 35 minutes. In total, sessions lasted approximately 60 minutes and subjects earned on average €10.48, including a show-up fee of €3.00, with a standard deviation of €3.39.

4.5.5 Results of Study 2

We focus on rounds 1 to 8. Out of 424 selection processes (53 subjects, 8 rounds), 23 are excluded, as no complete portfolio is achieved during the construction phase. As our evaluations do not consider the improvement phase, we do not require subjects to end their selection process with a complete portfolio as in the previous study. Study 1 has shown that most deselection steps in the construction phase can be explained by annulment patterns, which reflect reconsidered decisions, not constructive decision making. In order to focus on systematic portfolio development during the construction phase and to allow a comparison with the construction heuristics, we exclude all selection steps associated with items which are later deselected. Out of 5,924 selection and deselection steps for all considered rounds and subjects, we exclude 620 steps in the construction phase, roughly half of which are associated with annulment behavior.

In order to measure the degree to which subjects adhere to the heuristics, we consider the relative frequency of a subject selecting items in line with heuristics during the construction phase. For each subject and each round let $j_s \in \{1, \ldots, N\}$ indicate the item considered in step $s \in \{1, \ldots, S\}$ during the construction phase. All selection and deselection steps are given by $\mathcal{S}^{\mathrm{s}} \subseteq \{1, \ldots, S\}$ and $\mathcal{S}^{\mathrm{d}} = \{1, \ldots, S\} \setminus \mathcal{S}^s$. $y^s \in \{0,1\}^N$, $s \in \{1, \ldots, S\}$ denotes a subject's partial portfolio resulting from steps 1 through s where

$$y_j^s = \begin{cases} 1, & \text{if} \quad j = j_s, s \in \mathcal{S}^s \\ 0, & \text{if} \quad j = j_s, s \in \mathcal{S}^d \\ y_j^{s-1} & \text{else} \end{cases}$$

and $y^0 = \vec{0}$. In order to indicate whether a subject's selection is in line with one of the heuristics (12) - (15), we define function

$$\alpha(s) = \begin{cases} 1, & \text{if} \quad j_s = h(\mathcal{A}_s) \\ 0 & \text{else} \end{cases} \tag{17}$$

where $\mathcal{A}_s = \left\{j \in \{1, \ldots, N\} | y_j^{s-1} = 0, k^T y^{s-1} + k_j \leq c\right\}$. Absolute and relative heuristic adherence is given by (18) and (19), respectively.

$$A_{abs} = \sum_{s \in \mathcal{S}^s} \alpha(s) \tag{18}$$

$$A_{rel} = \frac{A_{\text{abs}}}{|\mathcal{S}^s|} \tag{19}$$

If a subject selects items exactly in line with a heuristic, $A_{rel} = 1$. $A_{rel} = 0$ holds, if a subject ignores items chosen by a heuristic in every single step. Figure 4.5 reports on the A_{rel} values considering heuristics (12) - (15) for all considered rounds and subjects. We compare the distribution of A_{rel} for subjects' decision processes with the distribution resulting from randomly selecting items. As in our first experiment, random selection behavior is approximated by Monte Carlo simulation with a sample size of 10,000. Subjects' A_{rel} values are strictly smaller than 1 for all four heuristics, indicating that no subject completely adheres to one heuristic during the construction phase. Comparing subjects' distribution of heuristic adherence with the distribution for random selection behavior in Figure 4.5, there is no obvious difference for the MaxV or the MinK heuristic. In contrast, the degree of subjects' heuristic adherence to the MaxR as well as the MaxD heuristic is higher than for random selection behavior.

In order to investigate the heuristic adherence across all rounds, we further consider the absolute adherence A_{abs}. As there is no significant difference between the absolute adherence for all subjects and the absolute adherence for random selections with regard to MaxV (one-tailed Mann-Whitney, $p = 0.121$), we find no support for Hypotheses *H1a*. Considering the MinK heuristic, subjects' absolute adherence across all rounds is significantly higher than for random selections (one-tailed Mann-Whitney, $p < 0.001$). However, considering heuristic adherence in each round, as presented in Table 4.5, this finding only holds for two out of the eight rounds under consideration. We thus only find weak support for Hypothesis *H1b*. Investigating the MaxR heuristic, subjects' behavior is significantly more often in line with the heuristic than would be expected for random behavior across all rounds and in each round separately (one-tailed Mann-Whitney, $p < 0.001$). Hypotheses *H1c* is therefore confirmed. Across all rounds the absolute adherence to the MaxD heuristic is also significantly higher than for random selections (one-tailed Mann-Whitney, $p < 0.001$). While the effect is only mildly significant in

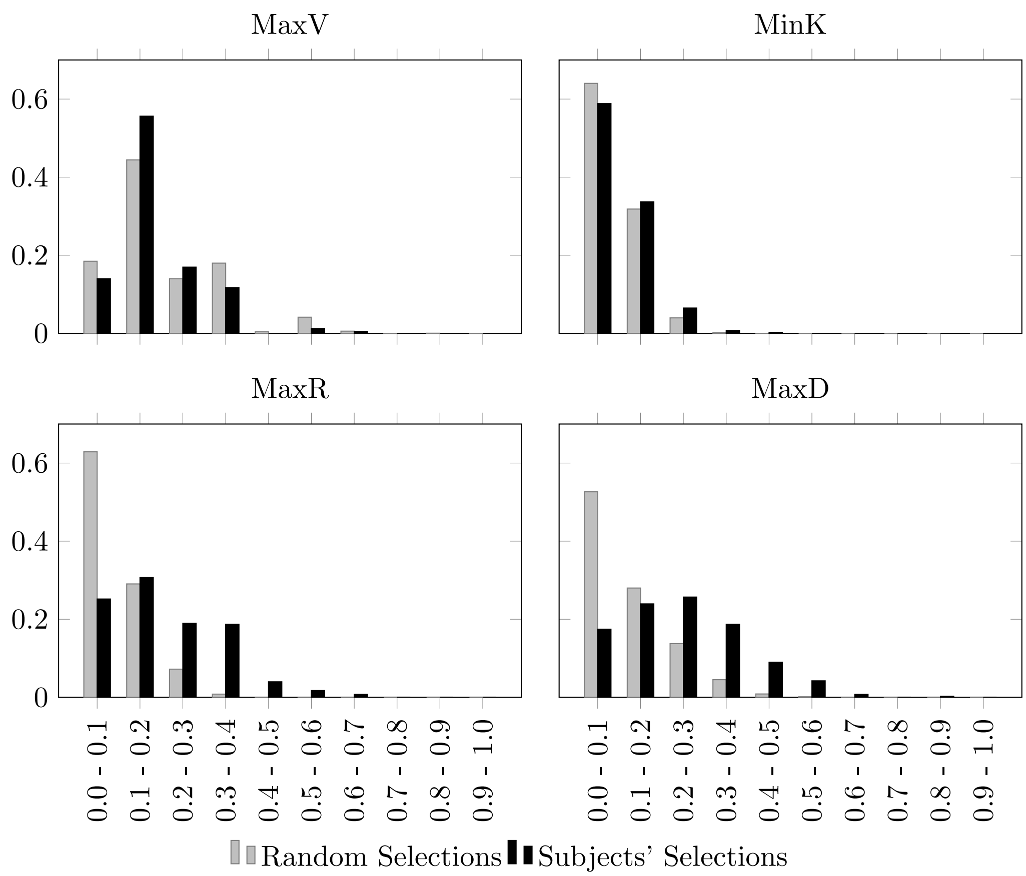


Figure 4.5: Relative adherence to heuristic selections by subjects and in case of random selections in deciles

round 3 (one-tailed Mann-Whitney, $p = 0.003$), there is strong significance in the remaining 7 rounds (one-tailed Mann-Whitney, $p < 0.001$). We conclude that people partially act in line with the MaxD heuristic and find support

Round							
1	2	3	4	5	6	7	8
p <0.001	p=0.496	p=0.197	p=0.016	p=0.464	p <0.001	p=0.081	p=0.319

Table 4.5: One-tailed p-values of the Mann-Whitney test for the MinK heuristic

for Hypotheses *H1d*. Overall, the experimental results show that the human selection process is partially in line with the MaxR and the MaxD heuristic. Focusing on the two instances for which we ensured that the portfolios resulting from the MaxR heuristic and MaxD heuristic have only few items in common, we cannot identify any significant differences between adherence to both heuristics (two-tailed Mann-Whitney, $p = 0.294$).

We find significant support that people prefer to select items with the highest ratio or the highest difference but nevertheless many selections are not according to the MaxR or the MaxD heuristic. Across all rounds the highest heuristic adherence of a subject to any heuristic is $A_{rel} = 0.41$, demonstrating that less than 50% of subjects' selection processes are in line with one of the four heuristics. To obtain more insight into these deviations, we further investigate subjects' selection processes. Assigning an item number to each item, beginning with number 1 for the item at the top of the list and ending with number 25 for the last item, for two consecutive selection steps we define the "selection span" as the difference in the list position between two selected items. A positive selection span results from selecting an item with a higher number than the previously selected item and a negative selection span results from selecting an item with a lower number. The selection span does not depend on whether items within the span are already selected or not. Again, we compare the relative frequencies for subjects' selection spans with the frequencies expected for random selection behavior. We determine the selection spans for 10,000 randomly generated selection processes. Figure 4.6 illustrates that about 50% of all selection spans between two selection steps lie in the range $[-3, 3]$, which is significantly more (more than twice as much) as expected for random selection behavior (Binomial test, $p < 0.001$). Subjects prefer to select items in close proximity to the previously selected item, while large spans are underrepresented in subjects' selection behavior. In absolute values, subjects' selection span is significantly smaller than for a random selection process in each round (one-tailed Mann-Whitney, $p < 0.001$). Therefore, Hypotheses *H2* can be approved: people apply localized selection behavior.

As illustrated in Figure 4.6, there are more positive than negative selection spans indicating that subjects move through the item list from top to bottom. Therefore, we consider the sequence in which items are selected,

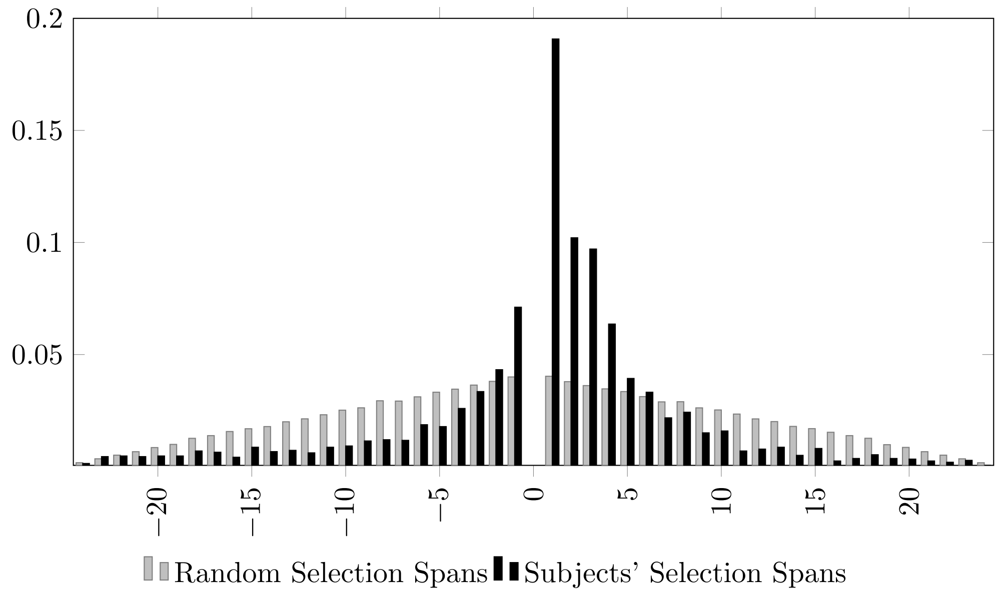


Figure 4.6: Histogram of selection spans for consecutive selection steps

based on their position in the item list. Figure 4.7 illustrates the distribution of selected item numbers during the first 10 selection steps. Item 1 is shown to subjects at the top of the item list, while item 25 is presented last. Figure 4.7 demonstrates that during the first 5 selection steps more than half of all selected items lie in the upper half of the item list (item number 1 - 13). Furthermore, the item number selected by subjects has a significantly positive trend during the first 5 selection steps (Jonckheere trend test, $p < 0.001$). This trend disappears for the following 5 selection steps (Jonckheere trend test, $p = 0.547$), which is not surprising, as people are likely to have traversed the complete item list. We conclude that people initially are biased towards items at the top of the item list and only later consider items from the bottom. Identifying this behavior, we consider whether the order of the item list also has an influence on the complete portfolios. Comparing the items in the complete portfolios of the construction phase for rounds 1 to 4 and the following identical rounds with reversed order of items verifies that there is no significant difference between the selected items for each pair of rounds, i.e. round 1 and 6 (two-tailed Mann-Whitney, $p = 0.889$), round 2 and 5 (two-tailed Mann-Whitney, $p = 0.936$), round 3 and 7 (two-tailed Mann-Whitney, $p = 0.271$), as well as round 4 and 8 (two-tailed Mann-Whitney, $p = 0.725$).

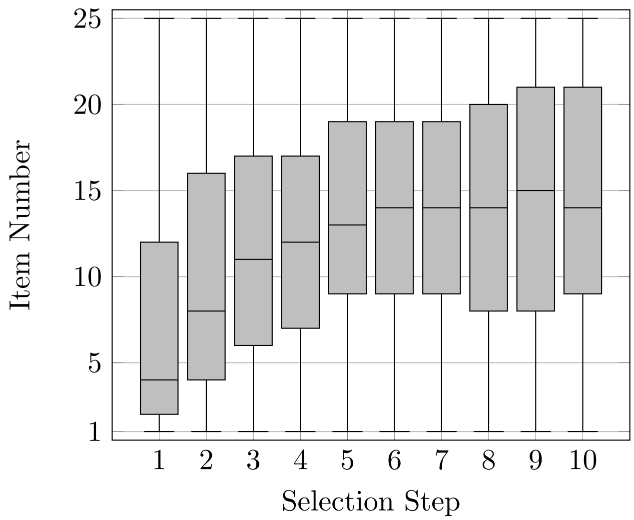


Figure 4.7: Box plot diagram representing the item numbers of the selected items in the 1st to 10th selection step

We investigate whether subjects' selection behavior is better explained by the MaxR or MaxD heuristic, when accounting for localized selection behavior. Function

$$\alpha^{b,f}(s) = \begin{cases} 1, & \text{if} \quad j_s = h(\mathcal{A}_s^{b,f}) \\ 0 & \text{else} \end{cases} \tag{20}$$

where $\mathcal{A}_s^{b,f} = \mathcal{A}_s \cap \{j_{s-1} - b, \ldots, j_{s-1} + f\}$, $s \in \mathcal{S}^s \setminus \{1\}$ indicates whether a subject's selection is in line with the selection according to a heuristic, when only considering unselected items whose indexes fall in the range $[j_{s-1} - b, j_{s-1} + f]$, $b, f \in \mathbb{N}_0^+$ of the previously selected item j_{s-1}. Relative heuristic adherence considering localized selection behavior is given by

$$A_{rel}^{b,f} = \frac{\sum_{s \in \mathcal{S}^s} \alpha_s^{b,f}}{|\mathcal{S}^s|}. \tag{21}$$

Figure 4.8 reports average $A_{rel}^{b,f}$ values considering the MaxR and MaxD heuristic for different b and f values, and across all rounds and subjects. The figure illustrates that subjects' selection processes coincide more with a

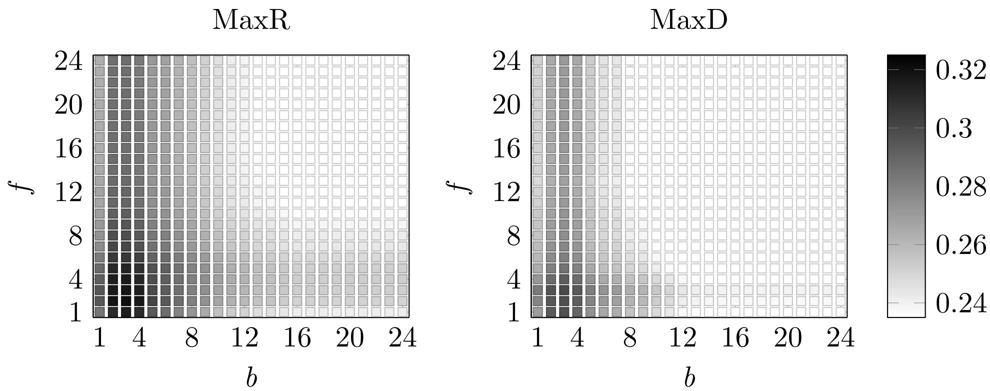


Figure 4.8: Average $A_{\text{rel}}^{b,f}$ values considering the MaxR and MaxD heuristic for different b and f values

heuristic when small b and f values are considered. While the average adherence value for MaxR and MaxD is 0.22 and 0.21, limiting subjects' backward search range to $b = 2$ and forward search range to $f = 3$ leads to increased average adherence to the MaxR heuristic at a level of 0.32 and 0.30 for the MaxD heuristic. An incomplete search pattern within a small range explains subjects' selection processes much better than a global selection behavior.

4.5.6 Discussion of the Results of the Experimental Studies

We investigate subjects' adherence to simple constructive heuristics during the construction phase. Based on our experimental studies, we conclude that subjects compare and select items "locally", based on a limited subset of the set of presented items in close vicinity to previously selected items. They start at the top of the presented item list and gradually move through it. In consequence, subjects' selection processes are only partially in line with a construction heuristic. A higher heuristic adherence would require considering the whole set of items. Further explanations are that people might try to use a heuristic but that they fail in choosing the best item or that they actually do not try to choose the best item according to an evaluation criterion and instead just select a "good" one. But even if the heuristic adherence in every single step is limited, the results of Study 1 demonstrate that the probability that an item is in the final portfolio significantly increases with a higher ratio or difference of value and resource requirement. Another finding from Study 1 is that the decision making behavior indicates that the improvement phase is dominated by simple annulment patterns and has only limited impact on decision quality.

4.6 Conclusion

Managers are responsible for portfolio decisions in strategic environments, where the application of quantitative decision support is limited. Greater understanding of human portfolio decision making behavior and associated decision biases will bring greater rationality to decision making settings. Our experimental findings that humans behave suboptimally even in the abstract settings of our experiments, serve as a reminder of the limits of human decision making. Salo et al. [118] emphasize the need for research on the occurrence, the impact, and the avoidance of decision biases in portfolio decision making settings. Companies spend millions of dollars on bad projects and these projects may be selected just because decision makers use decision rules which do not lead to good decisions in specific settings. Understanding fallacies of human decision making enables organizations to counteract such situations in the future by designing decision processes and support systems accordingly.

The scope of our work provides ample opportunities for further research. We investigate learning effects by considering three identical but rescaled and rearranged problem instances. We infer that the complexity of the problem hinders learning in the short run, but make no projections regarding long-term learning behavior. Overcoming the difficulty that people will realize when the same instance has been rescaled and rearranged too often, or focusing on cross learning effects over different instances, further investigations on learning effects as well as training would be worthwhile. As people are able to improve their memory span (see Ericcson et al. [42]) decision quality might also improve through training due to less localized selection behavior. Another promising research avenue would be to investigate whether instances in which the heuristics lead to bad performance coincide with a worse performance of decision makers compared to instances where the heuristics lead to good or even optimal solutions. If this is the case, particular critical situations could be predicted in advance. Predicting critical situations, training decision makers to handle them as well as general debiasing methods, and the investigation of differences on the individual level, seem to be the most important areas for further research in the fundamental knapsack problem setting that we examine. Considering many practical situations the baseline knapsack problem is rather stylized and therefore various extensions exist, taking into account, e.g. uncertainty, multiple objectives, group decisions, project dependencies and so on. Based on our framework, these topics could also be addressed and the results could be compared using our findings as a baseline. We believe that human behavior in a project selection decision with all its facets is a promising field for research.

Operational research aims to help people in problem solving and in order to come up with better results the problem owners should not be neglected (see Hämäläinen et al. [64]). Almost 30 years ago Booker and Bryson [16] stated that "the old 'black box' modeling techniques from utility theory and mathematical programming tend to restrict the user to rigid mathematical formulations, and thus lose accuracy". Since human decision makers are responsible for many project selection tasks, studying their behavior is crucial. We have demonstrated that decision makers behave in line with two simple heuristics in a simple project selection task and it is well-known that when decision makers use heuristics to solve complex problems this can result in systematic errors with serious implications (see Gino and Pisano [58]).

Chapter 5

Conclusion

Behavioral operations in general, and experimental studies in particular, provide ample opportunities to bridge the gap between analytical models and real life business situations. This dissertation considers the domains inventory, health care, and portfolio planning, focusing on two elementary problems of operations management and operations research, the newsvendor problem and the knapsack problem. While the former is a basis for a wide range of settings where uncertainty is involved, the latter is a building block for manifold operations research settings in a deterministic environment.

Human decision making in the newsvendor context has become one of the most investigated areas in behavioral operations. While many studies examine decision making in the opportunity-based newsvendor problem, as typical in inventory management, there are many newsvendor settings which are penalty-based. Furthermore, it is well known that the assessment of cost plays an important role in decision making. We experimentally test and compare decision maker behavior in both settings and demonstrate that a penalty-based newsvendor problem results in significantly different behavior, which can be referred to a different assessment of costs. Based on this theoretical foundation, we investigate surgery planning. Analyzing hospital data, we verify that the problem structure of surgery planning is identical with the cost-based newsvendor problem, and we test surgeons planning behavior conducting an experimental test with 40 experienced surgeons. Our results demonstrate a biased planning behavior of surgeons which is in line with the findings from the penalty-based newsvendor experiments but surgeons' decision making is even more strongly biased. While the newsvendor problem is a fundamental building block in stochastic settings, we consider the knapsack problem as an adequate problem to study behavior in deterministic settings. Based on an experimental study, we investigate how people behave during the selection process in portfolio planning. We show that their selection processes can be divided into a construction phase, in which they apply simple heuristics to construct a portfolio, and an improvement phase in which they try to further improve the portfolio. Our results demonstrate suboptimization even in very small problem instances with few alternatives. Furthermore, we show that due to cognitive limitations, decision makers apply heuristics only within a limited subset of all alternatives.

The common denominator of our findings is that biases and decision rules negatively impact operations efficiency. While we provide promising insights into decision maker behavior, the essentials are that there is still need for further investigation of human behavior in the newsvendor problem and the knapsack problem, as they are the backbones of a vast number of operations management and operations research settings with manifold extensions. Throughout this dissertation, we have made several cautious attempts to refer our findings to managerial decision making in real world situations. The fundamental question which has not been sufficiently answered by research thus far is how good behavioral findings demonstrated in a laboratory setting could be transferred to the business world. It remains open whether the insights from the simplified newsvendor and knapsack problem settings that we considered can hold up in more complex settings as typical for most real life situations. A promising approach to overcome this problem is to verify experimental results obtained in the laboratory with professionals and to adjust the simplified problem settings stepwise to more realistic problem descriptions. Conducting experiments with experienced surgeons in different hospitals for our second study involved a tremendous effort, but as a consequence the results received are of higher external validity. Laboratory experiments are an adequate starting point as conducting field experiments typically goes along with less control and an increase in effort; however, besides laboratory experiments field experiments are needed to obtain higher external validity. Both types of experiments are important tools to improve the descriptive accuracy of analytical models. Behavioral experiments and mathematical models can jointly advance operations management and operations research (see Bendoly et al. [7], Gans and Croson [50], Kremer and Minner [81], Katok [73], Hämäläinen et al. [64]).

The future of behavioral operations is both challenging and exciting and one could be confident that behavioral research will further contribute to bridging the gap between theory and practice.

Appendix A

Instructions for the Experiments

A.1 Instructions for the Experiments in Chapter 2

Instructions: The instructions are translated from German and shortened since the original instructions also contain examples and screenshots. Furthermore, they do not contain the price and cost values and the payment figures since they are different for the three margin cases. Differences in the penalty cost and the opportunity cost problem are set in italics. The instructions consist of four parts though Part 1 is identical for both problems.

1. General information: You are about to participate in an experiment in decision making. You will receive a fixed payment of €4 for your appearance. Furthermore, in the course of the experiment you can earn a considerable amount of money depending on your decisions. In the experiment, all monetary amounts are specified in Experimental Currency Units (ECU). They are converted according a fixed exchange rate into Euro (see payment determination) at the end of the experiment. The experiment is followed by a short questionnaire and, afterwards, you will be paid in cash. All your decisions and answers will be treated confidentially. Please read the following instructions carefully. If you have any questions, please raise your hand. An instructor will come to your place and answer your questions. During the

experiment you have to switch off your cell phone and communication with other participants is prohibited. If you fail to comply with these rules, we will exclude you from the experiment and you will receive no payment.

Opportunity cost problem (Part 2-4):

2. Experimental task: Your job is to determine the order quantity of a product before you know the demand. You know that the demand is equally probable for any value between 0 and 100. If your order quantity exceeds the demand, the remaining products are worthless. *If the demand exceeds your order quantity, the unsatisfied demand expires.* For each product you order, you pay a price of ECU ... to the wholesaler (the costs per product unsold correspond to ECU ...). *For each product sold, you will receive a price of ECU ... from your customers (the opportunity costs for each product ordered too little corresponds to ECU ...).*

- You cannot sell more products than are demanded.
- *You cannot sell more products than you have ordered.*

3. Experimental procedure: The experiment consists of 30 rounds and the demand in each round is independent of past demand. Every round consists of two screens. The first screen summarizes the information already given in the instructions. Furthermore, you have to enter the number of products you want to order in a red box and press the button “OK”. Please take sufficient time to make your decisions. Once all participants have confirmed their entry, the second screen appears. On the second screen, your order quantity is given again and you receive information about the realized demand. Furthermore, the resulting *gains/losses* are listed. When all participants have pressed “OK”, the next round starts.

4. Payment determination: You receive a fixed payment of €4 for your appearance. Furthermore, you can earn additional money dependent on your performance in the course of the experiment. At the end of the experiment, the *gains/losses* in ECU incurred in all rounds are added together. Your payoff is the resulting amount which is converted by a factor of ECU ... = €1 plus the €4 you receive for your appearance. In the event that you have generated a total loss, you still receive your show-up fee.

Penalty cost problem (Part 2-4):

2. Experimental task: Your job is to determine the order quantity of a product before you know the demand. You know that the demand is equally probable for any value between 0 and 100. *The demand of the customers has to be satisfied.* If your order quantity exceeds the demand, the remaining products are worthless. *If the demand exceeds your order quantity, you have to reorder products instantly at a higher price.* For each product you order, you pay a price of ECU ... to the wholesaler (the costs per product unsold correspond to ECU ...). *For each product ordered too little, you have to pay a price of ECU ... to the wholesaler (the additional costs for each product ordered too little correspond to ECU ...).*

- You cannot sell more products than are demanded.
- *You have to reorder products if demand exceeds the order quantity.*

3. Experimental procedure: The experiment consists of 30 rounds and the demand in each round is independent of past demand. Every round consists of two screens. The first screen summarizes the information already given in the instructions. Furthermore, you have to enter the number of products you want to order in a red box and press the button “OK”. Please take sufficient time to make your decisions. Once all participants have confirmed their entry, the second screen appears. On the second screen, your order quantity is given again and you receive information about the realized demand. Furthermore, the resulting *costs* are listed. When all participants have pressed “OK”, the next round starts.

4. Payment determination: You receive a fixed payment of €4 for your appearance. Furthermore, you can earn additional money dependent on your performance in the course of the experiment. At the end of the experiment, the *costs* in ECU incurred in all rounds are added together. *These costs will be deducted from a fixed budget of ECU ..., which is available to fulfill the task*. Your payoff is the resulting amount which is converted by a factor of ECU ... = €1 plus the €4 you receive for your appearance. In the event that you have generated a total loss, you still receive your show-up fee.

A.2 Instructions for the Experiments in Chapter 3

Instructions: The instructions are translated from German and shortened since the original instructions also contain screenshots. Furthermore, they do not contain the cost values and the payment figures since they are different for the instruction of the low quantile case and the high quantile case. The instructions consist of four parts.

1. General information: You are about to participate in an experiment in decision making. In the course of the experiment you can earn a considerable amount of money depending on your decisions. In the experiment, all monetary amounts are specified in Experimental Currency Units (ECU). They are converted according a fixed exchange rate into Euro (see payment determination) at the end of the experiment. The experiment is followed by a short questionnaire and, afterwards, you will be paid in cash. All your decisions and answers will be treated confidentially. Please read the following instructions carefully. If you have any questions, please ask.

2. Experimental task: Consider the following simplified decision situation about planning of surgery durations. Your job is to reserve time for a surgery in the operating room. You don't know how long the surgery will take but you know that the duration of that surgery (in minutes) is equally probable for any value between 100 and 200. Every reserved minute of the operating room is associated with costs. If your reserved time exceeds the duration, the remaining time can not be used otherwise. If the duration exceeds your reserved time, the additional time needed is associated with higher costs. The surgery can not be interrupted. For each minute you reserve the operating room, the costs are ECU ... (the costs per minute reserved too much correspond to ECU ...). For each minute the operating room is needed beyond the reserved time, the costs are ECU ... (the additional costs for each minute reserved too little correspond to ECU ...).

- The cost per minute reserved time even occur if the duration is shorter than the reserved time.
- The operation must be carried out until the end.

3. Experimental procedure: The experiment consists of 20 rounds and the surgery duration in each round is independent of past surgery durations. Every round consists of two screens. The first screen summarizes the information already given in the instructions. Furthermore, you have to enter the minutes you want to reserve the operating room (between 100 and 200 minutes) in the red box and press the button “OK”. Please take sufficient time to make your decisions. Afterwards, the second screen appears. On the second screen, your reserved time is given again and you receive information about the realized duration. Furthermore, the resulting costs are listed. After pressing “OK”, the next round starts. You have to plan 20 independent surgeries.

4. Payment determination: You can earn money dependent on your performance in the course of the experiment. At the end of the experiment, the costs in ECU incurred in all rounds are added together. These costs will be deducted from a fixed budget of ECU ..., which is available to fulfill the task. Your payoff is the resulting amount which is converted by a factor of ECU ... = €1. Depending on your performance, the payoff will be between €5 and €55.

A.3 Instructions for the Experiments in Chapter 4

Instructions: We present the instructions for the experiments, translated from German. The instructions for the first and second experiment are identical except the statement of the number of items considered in the knapsack problem. An illustrative screenshot from the original instructions supplemented with detailed descriptions is omitted.

1. General information: You are about to participate in an experiment in decision making. In the course of the experiment, you can earn a considerable amount of money depending on how good your decisions are. In the experiment, all monetary amounts are specified in Experimental Currency Units (ECU), which are converted according a fixed exchange rate into Euro at the end of the experiment (see experimental payout). All your decisions and answers will be treated confidentially. Please read the following instructions carefully. Should you have any questions, please ask. During the experiment you have to switch off your cell phone and communication with other participants is prohibited.

2. Experimental task and procedure: A set of items is given and each item generates a value but requires a capacity. Your task is to select a subset of items given that a higher aggregate value results in a higher payout while the aggregate resource requirement must not exceed the available capacity.

The experiment consists of several independent rounds with a different number of items (5,10,15, or 25), different item properties, and different capacities. Every round consists of a single screen displaying all items in a table containing information about the item properties (value and resource requirement) as illustrated in Figure A.1. For each item, you can decide to select it from the list and you are free to deselect already selected items at any time. Furthermore, the remaining capacity as well as the value of the portfolio is displayed on the screen. Please note that if the selected item results in an aggregate resource requirement exceeding the available capacity, an error message will appear. Please take sufficient time to make your decisions and once you have made your selection, press the continue button to go to the

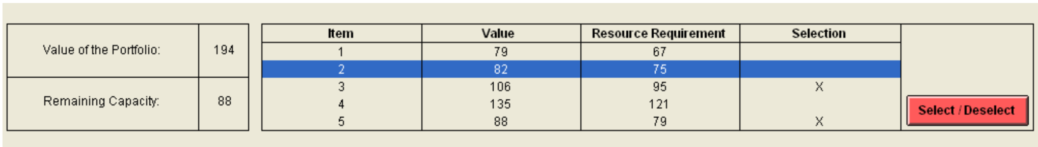


Figure A.1: Excerpt of the interface for portfolio selection presented to subjects

next round. At most you have 5 minutes to complete each round and when the time is over, you will be automatically taken to the next screen. The experiment ends after 35 minutes.

Before the main rounds start, there are three practice rounds and you have to fill out a short questionnaire at the end of the experiment. In total the experiment will take about 60 minutes.

3. Experimental payout: In each round the aggregate value of all selected items in ECU is converted by a linear exchange-rate to €. The exchange-rate is round-dependent and displayed on the screen. At the end of the experiment, one round out of all completed rounds is randomly chosen for payout. For the payout a fixed amount of €100 is subtracted from the aggregate Euro value in this round. In the unlikely case that the resulting payout is less than €3, you still receive a minimum of €3 as show-up fee.

Appendix B

Abbreviations, Notations, and Symbols

B.1 General Abbreviations

ACE	Assessment of costs effect
CE	Combined effect
ECU	Experimental currency units
HQC	High quantile case
K-S	Kolmogorow-Smirnow
LQC	Low quantile case
MAE	Mean anchor effect
MaxD	Maximum difference between value and resource requirement
MaxR	Maximum ratio of value divided by resource requirement
MaxV	Maximum value
MELESSA	Munich Experimental Laboratory for Economic and Social Sciences
MinK	Minimum resource requirement
OCP	Opportunity cost problem
OR	Operating room

PCP	Penalty cost problem
TUM	Technische Universität München

B.2 Notations and Symbols

Notations and Symbols in Chapter 2

α	mean anchor weight
β	underage cost weight
β_{opp}	underage cost weight in the opportunity cost problem
β_{pen}	underage cost weight in the penalty cost problem
c	purchasing costs per item
c_o	overage costs
c_u	underage costs
d	uncertain demand
D	realized demand
ϵ_t	error term in period t
$F(\cdot)$	cumulated demand distribution
$F^{-1}(\cdot)$	inverse of the cumulated demand distribution
μ	mean demand
p	selling price per item
q	order quantity
q^*	optimal order quantity
q^{ACE}	order quantity for the ACE
q^{CE}	order quantity for the CE
q^{MAE}	order quantity for the MAE
q_t	order quantity in period t
s	reorder costs
t	period

Notations and Symbols in Chapter 3

c	costs per minute of used OR capacity
c^o	costs per minute of overutilization
c^u	costs per minute of underutilization
$C(p, D)$	total costs as a function of p and D
D	realized duration
$E[C(\cdot)]$	expected costs for the planned duration
$F(\cdot)$	cumulated distribution function of the duration
$F^{-1}(\cdot)$	inverse of the cumulated distribution function of the duration
$I(\cdot)$	relative cost increase for the planned duration
μ	mean realized duration
p	planned duration
p^*	optimal planned duration
s	increased costs per minute of overtime
$U(\cdot)$	uniform demand distribution

Notations and Symbols in Chapter 4

$\alpha(s)$	consistency of step s to a heuristic
$\alpha^{b,f}(s)$	consistency of step s to a heuristic for b and f
A_{abs}	absolute heuristic adherence
A_{rel}	relative heuristic adherence
$A_{rel}^{b,f}$	relative heuristic adherence for b and f
$\mathcal{A}_s$	unselected items in step s which do not exceed capacity
$\mathcal{A}_s^{b,f}$	unselected items in step s for b and f which do not exceed capacity
b	backward search range
b_1	slope of the trend line
c	resource capacity
f	forward search range
Φ_{ran}	expected optimality gap for random portfolios
Φ_{sub}	optimality gap for subjects' portfolios
$h(\cdot)$	highest ranked item according to an evaluation criterion
j	item, $j \in \{1, \ldots, N\}$
j_s	selected item in step s
k	resource requirement, $k \in \mathbb{R}_+^N$
k_j	required resource of item j
N	number of items
p_{opt}	optimum portfolio value
p_{sub}	subjects' portfolio value
R	value range limit for instance generation
s	step, $s \in \{1, \ldots, S\}$
$\mathcal{S}^{\mathrm{d}}$	set of deselection steps
$\mathcal{S}^{\mathrm{s}}$	set of selection steps
v	value, $v \in \mathbb{R}_+^N$

v_j	value of item j
x^0	empty portfolio of a heuristic
x^s	partial portfolio of a heuristic containing selected items up to step s
x_j	binary variables indicating the selection or exclusion of item j for a heuristic
y^0	subjects' empty portfolio
y^s	subjects' partial portfolio containing selected items up to step s
y_j	binary variables indicating subjects' selection or exclusion of item j

Bibliography

[1] Abouleish, A. E., Dexter, F., Epstein, R. H., Lubarsky, D. A., Whitten, C. W., and Prough, D. S. (2003). Labor costs incurred by anesthesiology groups because of operating rooms not being allocated and cases not being scheduled to maximize operating room efficiency. *Anesthesia & Analgesia*, 96(4):1109–1113.

[2] Ball, C. T., Langholtz, H. J., Auble, J., and Sopchak, B. (1998). Resource-allocation strategies: A verbal protocol analysis. *Organizational Behavior and Human Decision Processes*, 76(1):70–88.

[3] Bearden, J. N., Murphy, R. O., and Rapoport, A. (2008). Decision biases in revenue management: Some behavioral evidence. *Manufacturing & Service Operations Management*, 10(4):625–636.

[4] Bearden, J. N. and Rapoport, A. (2005). Operations research in experimental psychology. In Smith, J. C., editor, *Tutorials in Operations Research: Emerging Theory, Methods, and Applications*, pages 213–236. Informs, Hanover, Maryland.

[5] Becker, S. W., Ronen, J., and Sorter, G. H. (1974). Opportunity costs – An experimental approach. *Journal of Accounting Research*, 12(2):317–329.

[6] Bendoly, E., Croson, R., Goncalves, P., and Schultz, K. L. (2010a). Bodies of knowledge for research in behavioral operations. *Production and Operations Management*, 19(4):434–452.

[7] Bendoly, E., Donohue, K., and Schultz, K. L. (2006). Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of Operations Management*, 24(6):737–752.

[8] Bendoly, E., Perry-Smith, J. E., and Bachrach, D. G. (2010b). The perception of difficulty in project-work planning and its impact on resource sharing. *Journal of Operations Management*, 28(5):385–397.

[9] Bendoly, E. and Swink, M. (2007). Moderating effects of information access on project management behavior, performance and perceptions. *Journal of Operations Management*, 25(3):604–622.

[10] Bendoly, E., Swink, M., and Simpson, W. P. (2013). Prioritizing and monitoring concurrent project work: Effects on switching behavior. *Production and Operations Management*. Article first published online, Doi 10.1111/poms.12083.

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

[12] Bland, J. M. and Altman, D. G. (1994). Statistics notes: Some examples of regression towards the mean. *British Medical Journal*, 309:780.

[13] Bleich, S. N., Özaltin, E., and Murray, C. J. (2009). How does satisfaction with the health-care system relate to patient experience? *Bulletin of the World Health Organization*, 87:271–278.

[14] Bolton, G. E. and 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.

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

[16] Booker, J. M. and Bryson, M. C. (1985). Decision analysis in project management: An overview. *IEEE Transactions on Engineering Management*, 32(1):3–9.

[17] Bornstein, B. H. and Emler, A. C. (2001). Rationality in medical decision making: A review of the literature on doctors' decision-making biases. *Journal of Evaluation in Clinical Practice*, 7(2):97–107.

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

[19] Brailsford, S. and Schmidt, B. (2003). Towards incorporating human behaviour in models of health care systems: An approach using discrete event simulation. *European Journal of Operational Research*, 150(1):19–31.

[20] Budde, M. and Minner, S. (2014). First-and second-price sealed-bid auctions applied to push and pull supply contracts. *European Journal of Operational Research*, 237(1):370–382.

[21] Busemeyer, J. R., Swenson, K. N., and Lazarte, A. (1986). An adaptive approach to resource allocation. *Organizational Behavior and Human Decision Processes*, 38(3):318–341.

[22] Cachon, G. P. and Swinney, R. (2009). The impact of strategic consumer behavior on the value of operational flexibility. In Tang, C. S. and Netessine, S., editors, *Consumer-Driven Demand and Operations Management Models*, pages 371–395. Springer, Berlin.

[23] Cardoen, B., Demeulemeester, E., and Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201(3):921–932.

[24] Carter, M. (2002). Diagnosis: Mismanagement of resources. *OR/MS Today*, 29(2):26–32.

[25] Chang, E. M., Hancock, K. M., Johnson, A., Daly, J., and Jackson, D. (2005). Role stress in nurses: Review of related factors and strategies for moving forward. *Nursing & Health Sciences*, 7(1):57–65.

[26] Chen, L., Kök, A. G., and Tong, J. D. (2013). The effect of payment schemes on inventory decisions: The role of mental accounting. *Management Science*, 59(2):436–451.

[27] Chod, J., Rudi, N., and Van Mieghem, J. A. (2010). Operational flexibility and financial hedging: Complements or substitutes? *Management Science*, 56(6):1030–1045.

[28] Chow, C. W. and Haddad, K. M. (1991). Relative performance evaluation and risk taking in delegated investment decisions*. *Decision Sciences*, 22(3):583–593.

[29] Chvátal, V. (1980). Hard knapsack problems. *Operations Research*, 28(6):1402–1411.

[30] Cooper, R. G., Edgett, S. J., and Kleinschmidt, E. J. (2001a). Portfolio management for new product development: Results of an industry practices study. *R & D Management*, 31(4):361–380.

[31] Cooper, R. G., Edgett, S. J., and Kleinschmidt, E. J. (2001b). *Portfolio management for new products*. Perseus Publishing, Cambridge, Massachusetts, 2 edition.

[32] Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1):87–114.

[33] Croson, R., Schultz, K., Siemsen, E., and Yeo, M. (2013). Behavioral operations: The state of the field. *Journal of Operations Management*, 31(1):1–5.

[34] de Véricourt, F., Jain, K., Bearden, J. N., and Filipowicz, A. (2013). Sex, risk and the newsvendor. *Journal of Operations Management*, 31(1):86–92.

[35] Denton, B., Viapiano, J., and Vogl, A. (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*, 10(1):13–24.

[36] Dexter, F., Macario, A., and Ledolter, J. (2007). Identification of systematic underestimation (bias) of case durations during case scheduling would not markedly reduce overutilized operating room time. *Journal of Clinical Anesthesia*, 19(3):198–203.

[37] Dexter, F. and Traub, R. D. (2002). How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. *Anesthesia & Analgesia*, 94(4):933–942.

[38] Donohue, K. L. (2000). Efficient supply contracts for fashion goods with forecast updating and two production modes. *Management Science*, 46(11):1397–1411.

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

[40] Eijkemans, M. J. C., van Houdenhoven, M., Nguyen, T., Boersma, E., Steyerberg, E. W., and Kazemier, G. (2010). Predicting the unpredictable: A new prediction model for operating room times using individual characteristics and the surgeon's estimate. *Anesthesiology*, 112(1):41–49.

[41] Eppen, G. D. (1979). Effects of centralization on expected costs in a multi-location newsboy problem. *Management Science*, 25(5):498–501.

[42] Ericcson, K. A., Chase, W. G., and Faloon, S. (1980). Acquisition of a memory skill. *Science*, 208(4448):1181–1182.

[43] Falk, A. and Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952):535–538.

[44] Fasolo, B., Morton, A., and von Winterfeldt, D. (2011). Behavioural issues in portfolio decision analysis. In Salo, A., Keisler, J., and Morton, A., editors, *Portfolio Decision Analysis*, pages 149–165. Springer, New York.

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

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

[47] Friedman, L. A. and Neumann, B. R. (1980). The effects of opportunity costs on project investment decisions: A replication and extension. *Journal of Accounting Research*, 18(2):407–419.

[48] Fügener, A., Schiffels, S., and Kolisch, R. (2014). Overutilization and underutilization of operating rooms: Insights from behavioral health care operations management. Working Paper, TUM School of Management, Technische Universität München.

[49] Funk, S. G., Rapoport, A., and Jones, L. V. (1979). Investing capital on safe and risky alternatives: An experimental study. *Journal of Experimental Psychology: General*, 108(4):415–440.

[50] Gans, N. and Croson, R. (2008). Introduction to the special issue on behavioral operations. *Manufacturing & Service Operations Management*, 10(4):563–565.

[51] Garland, H. (1990). Throwing good money after bad: The effect of sunk costs on the decision to esculate commitment to an ongoing project. *Journal of Applied Psychology*, 75(6):728–731.

[52] Gavirneni, S. and Isen, A. M. (2010). Anatomy of a newsvendor decision: Observations from a verbal protocol analysis. *Production and Operations Management*, 19(4):453–462.

[53] Gettinger, J., Kiesling, E., Stummer, C., and Vetschera, R. (2013). A comparison of representations for discrete multi-criteria decision problems. *Decision Support Systems*, 54(2):976–985.

[54] Ghasemzadeh, F. and Archer, N. P. (2000). Project portfolio selection through decision support. *Decision Support Systems*, 29(1):73–88.

[55] Gigerenzer, G. and Selten, R., editors (2002). *Bounded Rationality: The Adaptive Toolbox*. MIT Press, Cambridge, Massachusetts.

[56] Gingrich, G. and Soli, S. D. (1984). Subjective evaluation and allocation of resources in routine decision making. *Organizational Behavior and Human Performance*, 33(2):187–203.

[57] Gino, F. and Pisano, G. (2006). Do manager's heuristics affect R & D performance volatility? A simulation informed by the pharmaceutical industry. Working Paper 05-015, Harvard Business School.

[58] Gino, F. and Pisano, G. (2008). Toward a theory of behavioral operations. *Manufacturing & Service Operations Management*, 10(4):676–691.

[59] Golabi, K., Kirkwood, C. W., and Sicherman, A. (1981). Selecting a portfolio of solar energy projects using multiattribute preference theory. *Management Science*, 27(2):174–189.

[60] Gonzalez, R. M., Langholtz, H. J., and Sopchak, B. (2002). Minimizing cost in resource-allocation decisions. *Organizational Behavior and Human Decision Processes*, 89(2):1032–1057.

[61] Greiner, B. (2004). An online recruitment system for economic experiments. In Kremer, K. and Macho, V., editors, *Forschung und wissenschaftliches Rechnen 2003. GWDG-Bericht Nr. 63*, pages 79–93. Gesellschaft für Wissenschaftliche Datenverarbeitung, Göttingen.

[62] Guerriero, F. and Guido, R. (2011). Operational research in the management of the operating theatre: A survey. *Health Care Management Science*, 14(1):89–114.

[63] Gurgur, C. Z. and Morley, C. T. (2008). Lockheed martin space systems company optimizes infrastructure project-portfolio selection. *Interfaces*, 38(4):251–262.

[64] Hämäläinen, R. P., Luoma, J., and Saarinen, E. (2013). On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems. *European Journal of Operational Research*, 228(3):623–634.

[65] Heidenberger, K. and Stummer, C. (1999). Research and development project selection and resource allocation: A review of quantitative modelling approaches. *International Journal of Management Reviews*, 1(2):197–224.

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

[67] Ho, W., Xu, X., and Dey, P. K. (2010b). Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. *European Journal of Operational Research*, 202(1):16–24.

[68] Hoskin, R. E. (1983). Opportunity costs and behavior. *Journal of Accounting Research*, 21(1):78–95.

[69] Huang, X.-M. (1994). Patient attitude towards waiting in an outpatient clinic and its applications. *Health Services Management Research*, 7(1):2–8.

[70] Irvine, D. M. and Evans, M. G. (1995). Job satisfaction and turnover among nurses: Integrating research findings across studies. *Nursing Research*, 44(4):246–253.

[71] Jamal, M. and Baba, V. V. (1992). Shiftwork and department-type related to job stress, work attitudes and behavioral intentions: A study of nurses. *Journal of Organizational Behavior*, 13(5):449–464.

[72] Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., and Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, 59:193–224.

[73] Katok, E. (2011). Using laboratory experiments to build better operations management models. *Foundations and Trends in Technology, Information and Operations Management*, 5(1):1–86.

[74] Katok, E. and Wu, D. Y. (2009). Contracting in supply chains: A laboratory investigation. *Management Science*, 55(12):1953–1968.

[75] Katsikopoulos, K. V. and Gigerenzer, G. (2013). Behavioral operations management: A blind spot and a research program. *Journal of Supply Chain Management*, 49(1):3–7.

[76] Kester, L., Hultink, E. J., and Lauche, K. (2009). Portfolio decision-making genres: A case study. *Journal of Engineering and Technology Management*, 26(4):327–341.

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

[78] Kleinmuntz, D. (2007). Resource allocation decisions. In Edwards, W., Miles, R. F., and von Winterfeldt, D., editors, *Advances in decision analysis*. Cambridge University Press, Cambridge, Massachusetts.

[79] Knuth, D. E. (1968). *The art of computer programming*. Addison-Wesley series in computer science and information processing. Addison-Wesley, New York.

[80] Kremer, M. (2008). *Behavioral perspectives on risk sharing in supply chains*. PhD thesis, University of Mannheim.

[81] Kremer, M. and Minner, S. (2008). The human element in inventory decision making under uncertainty – a review of experimental evidence in the newsvendor model. *ZfB - Special Issue*, 4:83–97.

[82] Kremer, M., Minner, S., and Van Wassenhove, L. N. (2010). Do random errors explain newsvendor behavior? *Manufacturing & Service Operations Management*, 12(4):673–681.

[83] Kremer, M., Minner, S., and Van Wassenhove, L. N. (2013). On the preference to avoid ex post inventory errors. *Production and Operations Management*. Article first published online, Doi 10.1111/poms.12071.

[84] Kroll, Y., Levy, H., and Rapoport, A. (1988). Experimental tests of the mean-variance model for portfolio selection. *Organizational Behavior and Human Decision Processes*, 42(3):388–410.

[85] Langholtz, H., Gettys, C., and Foote, B. (1993). Resource-allocation behavior under certainty, risk, and uncertainty. *Organizational Behavior and Human Decision Processes*, 54(2):203–224.

[86] Langholtz, H., Gettys, C., and Foote, B. (1994). Allocating resources over time in benign and harsh environments. *Organizational Behavior and Human Decision Processes*, 58(1):28–50.

[87] Langholtz, H., Gettys, C., and Foote, B. (1995). Are resource fluctuations anticipated in resource allocation tasks? *Organizational Behavior and Human Decision Processes*, 64(3):274–282.

[88] Langholtz, H. J., Ball, C., Sopchak, B., and Auble, J. (1997). Resource-allocation behavior in complex but commonplace tasks. *Organizational Behavior and Human Decision Processes*, 70(3):249–266.

[89] Langholtz, H. J., Marty, A. T., Ball, C. T., and Nolan, E. C. (2003). *Resource-allocation behavior*. Kluwer Academic Publishers, Dordrecht, Netherlands.

[90] Lindstedt, M., Liesiö, J., and Salo, A. (2008). Participatory development of a strategic product portfolio in a telecommunication company. *International Journal of Technology Management*, 42(3):250–266.

[91] Loch, C. H. (2000). Tailoring product development to strategy: case of a european technology manufacturer. *European Management Journal*, 18(3):246–258.

[92] Loch, C. H., Pich, M. T., Terwiesch, C., and Urbschat, M. (2001). Selecting R&D projects at BMW: A case study of adopting mathematical programming models. *IEEE Transactions on Engineering Management*, 48(1):70–80.

[93] Loch, C. H. and Wu, Y. (2007). Behavioral operations management. *Foundations and Trends in Technology, Information and Operations Management*, 1(3):121–232.

[94] Luft, J. (1994). Bonus and penalty incentives – contract choice by employees. *Journal of Accounting and Economics*, 18(2):181–206.

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

[96] Martello, S. and Toth, P. (1990). *Knapsack problems: Algorithms and computer implementations*. J. Wiley & Sons, Chichester, England.

[97] Martinsuo, M. (2013). Project portfolio management in practice and in context. *International Journal of Project Management*, 31(6):794–803.

[98] May, J. H., Spangler, W. E., Strum, D. P., and Vargas, L. G. (2011). The surgical scheduling problem: Current research and future opportunities. *Production and Operations Management*, 20(3):392–405.

[99] May, J. H., Strum, D. P., and Vargas, L. G. (2000). Fitting the lognormal distribution to surgical procedure times. *Decision Sciences*, 31(1):129–148.

[100] McCaffery, E. J. and Baron, J. (2004). Framing and taxation: Evaluation of tax policies involving household composition. *Journal of Economic Psychology*, 25(6):679–705.

[101] McCaffery, E. J. and Baron, J. (2005). The political psychology of redistribution. *U.C.L.A. Law Review*, 52(6):1745–1792.

[102] Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81–97.

[103] Moritz, B. B. (2010). *Cognition and heterogeneity in supply chain planning: A study of inventory decision making*. PhD thesis, University of Minnesota.

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

[105] Moskowitz, A. J., Kuipers, B. J., and Kassirer, J. P. (1988). Dealing with uncertainty, risks, and tradeoffs in clinical decisions: A cognitive science approach. *Annals of Internal Medicine*, 108(3):435–449.

[106] Niranjan, T. T., Metri, B. A., and Aggarwal, V. (2009). Behavioral causes of the bullwhip effect: Breaking the mould. *International Journal of Services and Operations Management*, 5(3):350–374.

[107] Northcraft, G. B. and Neale, M. A. (1986). Opportunity costs and the framing of resource allocation decisions. *Organizational Behavior and human decision processes*, 37(3):348–356.

[108] Olivares, M., Terwiesch, C., and Cassorla, L. (2008). Structural estimation of the newsvendor model: An application to reserving operating room time. *Management Science*, 54(1):41–55.

[109] Pandit, J. J. and Carey, A. (2006). Estimating the duration of common elective operations: Implications for operating list management. *Anaesthesia*, 61(8):768–776.

[110] Peltokorpi, A. (2011). How do strategic decisions and operative practices affect operating room productivity? *Health Care Management Science*, 14(4):370–382.

[111] Phillips, L. D. (2011). The royal navy's type 45 story: A case study. In Salo, A., Keisler, J., and Morton, A., editors, *Portfolio Decision Analysis*, volume 162, pages 53–75. Springer, New York.

[112] Phillips, L. D. and Bana e Costa, C. A. (2007). Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. *Annals of Operations Research*, 154(1):51–68.

[113] Pisinger, D. (2005). Where are the hard knapsack problems? *Computers & Operations Research*, 32(9):2271–2284.

[114] Rapoport, A. (1984). Effects of wealth on portfolios under various investment conditions. *Acta Psychologica*, 55(1):31–51.

[115] Rhodes, M. and Barker, P. (2007). Operating room utilization. *Surgical Endoscopy*, 21(12):2339–2340.

[116] Rudi, N. and Drake, D. (2011). Observation bias: The impact of demand censoring on newsvendor level and adjustment behavior. Working Paper 12-042, Harvard Business School.

[117] Sachs, A.-L. and Minner, S. (2014). The data-driven newsvendor with censored demand observations. *International Journal of Production Economics*, 148:28–36.

[118] Salo, A., Keisler, J., and Morton, A. (2011). An invitation to portfolio decision analysis. In Salo, A., Keisler, J., and Morton, A., editors, *Portfolio Decision Analysis*, volume 162, pages 3–27. Springer, New York.

[119] Schiffels, S., Fliedner, T., and Kolisch, R. (2014a). Behavioral portfolio decision making: Insights from an experimental study. Working Paper, TUM School of Management, Technische Universität München.

[120] Schiffels, S., Fügener, A., Kolisch, R., and Brunner, J. O. (2014b). On the assessment of costs in a newsvendor environment: Insights from an experimental study. *Omega*, 43:1–8.

[121] Schultz, K. L., McClain, J. O., Robinson, L. W., and Thomas, L. J. (2007). The use of framing in inventory decisions. Johnson School Research Paper Series 02-07, Cornell University.

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

[123] Seale, D. A. and Rapoport, A. (1997). Sequential decision making with relative ranks: An experimental investigation of the "secretary problem". *Organizational Behavior and Human Decision Processes*, 69(3):221–236.

[124] Silber, J. H., Rosenbaum, P. R., Zhang, X., and Even-Shoshan, O. (2007). Estimating anesthesia and surgical procedure times from medicare anesthesia claims. *Anesthesiology*, 106(2):346–355.

[125] Silver, E. A., Pyke, D. F., and Peterson, R. (1998). *Inventory Management and Production Planning and Scheduling*. John Wiley & Sons, New York, 3 edition.

[126] Skaf, M. A. (1999). Portfolio management in an upstream oil and gas organization. *Interfaces*, 29(6):84–104.

[127] Smith-Miles, K. and Lopes, L. (2012). Measuring instance difficulty for combinatorial optimization problems. *Computers & Operations Research*, 39(5):875–889.

[128] Sodhi, M. S. and Tang, C. S. (2014). Guiding the next generation of doctoral students in operations management. *International Journal of Production Economics*, 150:28–36.

[129] Strum, D. P., May, J. H., and Vargas, L. G. (2000). Modeling the uncertainty of surgical procedure times: Comparison of log-normal and normal models. *Anesthesiology*, 92(4):1160–1167.

[130] Strum, D. P., Vargas, L. G., May, J. H., and Bashein, G. (1997). Surgical suite utilization and capacity planning: A minimal cost analysis model. *Journal of Medical Systems*, 21(5):309–322.

[131] Stummer, C., Kiesling, E., and Gutjahr, W. (2009). A multicriteria decision support system for competence-driven portfolio selection. *International Journal of Information Technology & Decision Making*, 8(2):379–401.

[132] Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, 1(1):39–60.

[133] Trietsch, D. (2006). Optimal feeding buffers for projects or batch supply chains by an exact generalization of the newsvendor model. *International Journal of Production Research*, 44(4):627–637.

[134] Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

[135] Wachtel, R. E. and Dexter, F. (2010). Review of behavioral operations experimental studies of newsvendor problems for operating room management. *Anesthesia & Analgesia*, 110(6):1698–1710.

[136] Whyte, G. (1991). Diffusion of responsibility: Effects on the escalation tendency. *Journal of Applied Psychology*, 76(3):408–415.

[137] Wright, I. H., Kooperberg, C., Bonar, B. A., and Bashein, G. (1996). Statistical modeling to predict elective surgery time: Comparison with a computer scheduling system and surgeon-provided estimates. *Anesthesiology*, 85(6):1235–1245.

[138] Yule, S., Flin, R., Paterson-Brown, S., and Maran, N. (2006). Non-technical skills for surgeons in the operating room: A review of the literature. *Surgery*, 139(2):140–149.

[139] Zhu, G., Bard, J. F., and Yu, G. (2007). A two-stage stochastic programming approach for project planning with uncertain activity durations. *Journal of Scheduling*, 10(3):167–180.