Department of Information and Service Economy

Behavioral Issues in Multiple Criteria Decision Making

Tommi Pajala

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 207/2017

Behavioral Issues in Multiple Criteria Decision Making

Tommi Pajala

Aalto University
School of Business
Department of Information and Service Economy

Main dissertation advisor
Professor Jyrki Wallenius, Aalto University School of Business

Preliminary examiners
Professor Konstantinos Katsikopoulos, University of Southampton, UK
Professor Gilberto Montibeller, Loughborough University, UK

Opponent
Professor Konstantinos Katsikopoulos, University of Southampton, UK

Aalto University publication series
DOCTORAL DISSERTATIONS 207/2017

ISBN 978-952-60-7682-9 (printed)
ISBN 978-952-60-7683-6 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-7683-6

Unigrafia Oy
Helsinki 2017

Finland

Supported by KAUTE, dr. h.c. Marcus Wallenberg foundation, HSE foundation, and LSR.

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Tommi Pajala

Name of the doctoral dissertation
Behavioral Issues in Multiple Criteria Decision Making

Publisher School of Business

Unit Department of Information and Service Economy

Series Aalto University publication series DOCTORAL DISSERTATIONS 207/2017

Field of research Management Science

Date of the defence 2 December 2017

Language English

☐ **Monograph** ☐ **Article dissertation** ☒ **Essay dissertation**

Abstract

Decision making is ubiquitous in the modern world. From one moment to the next, consumers are choosing between products, managers are choosing whom to hire, and politicians are choosing which legal project to support. Many decisions involve multiple criteria, necessitating the need to find a balanced alternative. Such decisions are modelled in a stream of research called Multiple Criteria Decision Making (MCDM).

In this dissertation, I connect MCDM methodologies with behavioral research from psychology. Despite the rise of psychological ideas in behavioral economics, behavioral operational research is still a relatively new field. This dissertation contributes to it by extending and applying psychological ideas in preference prediction and modeling.

Essay I empirically investigates the feasibility of approximating a decision maker's value function with a linear value function. We show that a linear value function is a suitable model for preference prediction, even when there is no linear value function that could theoretically explain all the decision maker's choices.

Essay II investigates how the context of the choice problem affects errors in decision making. While value theory treats decisions often without the context, we show that emotional attachment and product type affect the rate of choosing dominated alternatives. Furthermore, adding a third criterion does not necessarily lead to information overload, when the added information does not change the dominance relationships of the alternatives.

Essay III investigates the meaning of judgments of importance. We know decision makers often say that one criterion is more important than another. However, in the framework of multi-attribute value theory, interpreting such statements is difficult. In essay III, we take the concept of impact from previous literature, and come up with a suitable explanatory model by defining it as the product of AHP weight and the coefficient of variation for the criterion.

Essay IV analyzes whether psychological measures of decision style can explain errors in decision making. We find an effect for cognitive reflection, and the tendency to seek out alternatives. Intriguingly, decision makers who considered themselves analytically minded made the same number of mistakes as other decision makers, who did not consider themselves analytical.

Taken all together, these results show that there are important psychological and behavioral aspects in MCDM that highlight differences between properties of formal models and behavior of real decision makers. Ignoring the context in decision making may result in false conclusions, and utilizing psychological variables would improve our understanding of individual differences in decision making. Additionally, a simple linear value function may work just as well in prediction as more complex models.

Keywords Multiple Criteria Decision Making, behavioral OR

ISBN (printed) 978-952-60-7682-9	**ISBN (pdf)** 978-952-60-7683-6	
ISSN-L 1799-4934	**ISSN (printed)** 1799-4934	**ISSN (pdf)** 1799-4942
Location of publisher Helsinki	**Location of printing** Helsinki	**Year** 2017
Pages 123	**urn** http://urn.fi/URN:ISBN:978-952-60-7683-6	

Tekijä
Tommi Pajala

Väitöskirjan nimi
Behavioraaliset tekijät monikriteerisessä päätöksenteossa

Julkaisija Kauppakorkeakoulu

Yksikkö Tieto- ja palvelutalouden laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 207/2017

Tutkimusala Management Science

Väitöspäivä 02.12.2017

Kieli Englanti

☐ **Monografia** ☐ **Artikkeliväitöskirja** ☒ **Esseeväitöskirja**

Tiivistelmä

Päätöksenteko on suuressa roolissa taloudellisessa ja yhteiskunnallisessa toiminnassa. Joka päivä kuluttajat valitsevat tuotteiden välillä, ja johtajat valikoivat palkattavia työntekijöitä. Päätökset usein liittyvät moneen kriteeriin, mikä vaatii kykyä löytää eri kriteerien suhteen tasapainoinen vaihtoehto. Operaatiotutkimusessa näiden ongelmien tutkimuksen haaraa kutsutaan monikriteeriseksi päätöksenteoksi.

Tässä väitöskirjassa kytken monikriteerisen päätöksenteon metodologiaa psykologian tuloksiin ihmisten käyttäytymisestä. Huolimatta käyttäytymistaloustieteen noususta, käyttäytymisen rooli operaatiotutkimuksen piirissä on verraten uusi tutkimussuunta. Tämä väitöskirja kontribuoi alaan soveltamalla psykologian tutkimustuloksia preferenssien ennustamisessa ja mallinnuksessa.

Essee I tutkii empiirisesti sitä, voidaanko päätöksentekijän arvofunktiota approksimoida lineaarisella arvofunktiolla. Näytämme, että lineaarinen arvofunktio on toimiva malli preferenssien ennustamiseen, vaikka kaikille päätöksentekijöille ei edes teoriassa ole olemassa lineaarista arvofunktiota, joka selittäisi heidän tekemänsä päätökset.

Essee II tarkastelee kontekstin merkitystä päätöksenteon virheille. Vaikka teoria usein käsittelee päätöksentekoa ilman kontekstia, näytämme, että emotionaalinen kiinnittyminen tuotteeseen sekä tuotetyyppi vaikuttavat todennäköisyyteen valita dominoitu vaihtoehto. Lisäksi näytämme, että kolmannen kriteerin lisääminen ei välttämättä johda informaatioähkyyn, kun lisätty kriteeri ei muuta vaihtoehtojen dominanssisuhteita.

Essee III käsittelee kriteerien tärkeyden tulkintaa. Päätöksentekijät usein sanovat, että jokin kriteeri on jotain toista kriteeriä tärkeämpi. Kuitenkin monikriteerisen arvoteorian kehikossa kriteerien tärkeyden tulkinta on hankalaa. Esseessä III otamme aiemmasta kirjallisuudesta impaktin käsitteen, ja esitämme sille yhden toimivan määritelmän, jossa impakti määritellään AHP-tekniikalla estimoitujen painojen sekä kriteerin variaatiokertoimen tulona.

Essee IV analysoi sitä, voidaanko päätöksentekotyylin mittareita käyttää selittämään päätöksenteon virheitä. Näytämme, että kognitiivinen reflektiotesti, sekä tarve tutkia eri vaihtoehtoja liittyvät korkeampaan virhetodennäköisyyteen. Kiinnostavaa on myös se, että itsensä hyvin analyyttisiksi luokittelevat päätöksentekijät tekevät yhtä paljon virheitä kuin muutkin.

Kaiken kaikkiaan väitöskirjan tulokset näyttävät, että psykologiset ja behavioraaliset tekijät monikriteerisessä päätöksenteossa tulisi ottaa huomioon. Kontekstin huomiotta jättäminen voi johtaa vääriin johtopäätöksiin, ja psykologisten muuttujien tarkastelu lisää ymmärrystä yksilöiden välisistä eroista päätöksenteossa. Toisaalta lineaarinen arvofunktio voi toimia preferenssien ennustamisessa yhtä hyvin kuin monimutkaisemmat päätöksentekomallit.

Avainsanat monikriteerinen päätöksenteko, behavioraalinen OR

ISBN (painettu) 978-952-60-7682-9	**ISBN (pdf)** 978-952-60-7683-6	
ISSN-L 1799-4934	**ISSN (painettu)** 1799-4934	**ISSN (pdf)** 1799-4942
Julkaisupaikka Helsinki	**Painopaikka** Helsinki	**Vuosi** 2017
Sivumäärä 123	**urn** http://urn.fi/URN:ISBN:978-952-60-7683-6	

Preface

It takes a village to write a Dissertation, and this one is no exception.

Firstly, I am indebted to my supervisor and instructor, Professor Jyrki Wallenius, who has been an extraordinary supporter throughout the process. His expertise on multiple criteria decision making has made is feasible to complete this Dissertation with sufficient quality. What's more, his guidance in networking, publishing, and generally working in an academic environment ensured that I have enjoyed the process thoroughly. In my opinion, good advisor-student rapport is one important building block to a good experience in academia. It has truly been a privilege to work under your wing, and I have learned from you much more than this Dissertation can ever show.

In addition, the working environment at Management Science at the Aalto University School of Business has been a vital part of my daily work. My co-authors Pekka Korhonen, Pekka Malo, Niklas Ravaja, and Outi Somervuori generously provided me with avenues for scientific exploration and productivity. Jutta Heino and Merja Mäkinen ensured that administrative issues worked without a hitch, so that I could focus on my Dissertation.

Finally, I'm extremely thankful for the companionship provided by my fellow PhD student Ville Sillanpää. Academic work may be often a solitary effort, but you made sure our days were filled with not just work, but also fun. Sometimes what you need isn't a writing session, but fifty burpees. I think my legs still hurt.

My year at Mannheim University provided me with plenty of new ideas and perspectives. I'm very grateful to Prof. Dr. Martin Weber, Prof. Dr. Sebastian Müller, and all other colleagues in Germany for the opportunity

to learn from you.

As important as colleagues are, it's good for a PhD student to have a life outside the department. I'm lucky enough to have smart friends, who helped me look at the big picture. Without them, my days would have been much less exciting. Especially I'd like to thank Niko Perttilä, Juho Salmi, Joosef Valli, Maija & Tomi Alén, Anni Iso-Mustajärvi, Teemu Heiskanen, Anna Törrönen, Eevert Saukkokoski, Saara & Calle Sågbom, Rosa Leinonen, Satu Peltola, Kasperi Soininen, and the "Ajattelun Keidas" crew. You all played your own part in this dance, and for that I am very grateful.

I'd also like to thank my parents Eila Toiviainen and Tauno Pajala. Your support and encouragement through the years has been instrumental to the success of this project. My extended family including cousins, parents-in-law, Hanna & Mamba and Janne & Viivi also deserve a very special thank you.

One often hears that completing a Dissertation means being married to it. Thankfully, that marriage has not been the only one in my life. Dear Sonja Rajala, you have been ever accepting even when a journal rejected me, the beacon of light when servers went dark, and generally an amazingly supportive partner throughout this journey. You've helped me get through the worst, and made sure I remember to celebrate the best. This journey would not have been possible without you.

Helsinki, October 19, 2017,

Tommi Pajala

Contents

Preface **1**

Contents **3**

List of Publications **5**

Author's Contribution **7**

1. Introduction **9**

1.1 Objectives and Scope 10

1.2 Outline of the Dissertation 11

2. Theoretical Background **13**

2.1 Multiple Criteria Decision Making 13

2.2 Modeling Decisions in MCDM Research 13

2.2.1 Estimating the Value Function of the DM 16

2.2.2 Behavioral Aspects 16

2.3 Cognitive Aspects of Decision Making 17

2.4 Descriptive and Prescriptive Decision Analysis 19

3. Research Methods **23**

3.1 Experimental Research 23

3.2 The Epsilon Method for Estimating the DM's Value Function 24

3.3 Bayesian Hierarchical Models 25

3.4 Open Science and Preregistration 27

3.5 Psychological Measures 29

4. Overview of the Findings **31**

4.1 Essay I: Road to robust prediction of choices in deterministic MCDM 31
4.2 Essay II: Context matters: The impact of product type, emotional attachment and information overload on choice accuracy 32
4.3 Essay III: Judgments of importance revisited: what do they mean? 33
4.4 Essay IV: Explaining choice quality with decision style, cognitive reflection and decision environment 34

5. Summary and Conclusions **37**
5.1 Summary of Findings 37
5.2 Managerial Implications 38
5.3 Limitations 40
5.4 Further Research Suggestions 41

References **43**

Publications **47**

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Pajala, T., Korhonen, P., Wallenius, J.. Road to robust prediction of choices in deterministic MCDM. *European Journal of Operational Research*, 259, 1, 229-235, May 2017.

II Korhonen, P., Malo, P., Pajala, T., Ravaja, N., Somervuori, O., Wallenius, J.. Context matters: The impact of product type, emotional attachment and information overload on choice accuracy. *European Journal of Operational Research*, 264, 1, 270-279, January 2018.

III Pajala, T., Korhonen, P., Wallenius, J.. Judgments of importance revisited: What do they mean?. *Submitted manuscript*, March 2017.

IV Pajala, T.. Explaining choice quality with decision style, cognitive reflection and decision environment. *Submitted manuscript*, June 2017.

Author's Contribution

Publication I: "Road to robust prediction of choices in deterministic MCDM"

Pajala is the primary author. Korhonen and Wallenius proposed the research topic. Pajala gathered the data, did the analysis and wrote the paper under the supervision of Korhonen and Wallenius.

Publication II: "Context matters: The impact of product type, emotional attachment and information overload on choice accuracy"

Korhonen, Wallenius and Somervuori came up with the research topic and designed the study. Korhonen, Ravaja and Somervuori gathered the data. Malo and Pajala did the analysis. Somervuori and Pajala wrote the paper, everybody else contributed by revision commentary.

Publication III: "Judgments of importance revisited: What do they mean?"

Pajala is the primary author. Korhonen and Wallenius proposed the research topic. Pajala gathered the data, did the analysis and wrote the paper under the guidance of Korhonen and Wallenius.

Publication IV: "Explaining choice quality with decision style, cognitive reflection and decision environment"

Pajala is the sole author, and also generated the research problem.

1. Introduction

When modeling the preferences of people, a lot of previous work has focused on the mathematical properties of different kinds of value functions, and the boundary conditions of various models. Many papers propose new ways to model preferences, for example by accounting for imperfect information (Weber, 1987). In multicriteria decision making (MCDM), there are different schools of modeling, and these schools tend to have their own focus points. For a review of the field, see e.g. Wallenius et al. (2008) and Dyer et al. (1992).

A recent trend in the field of Operations Research (OR) has been the rise of behavioral OR (BOR). This stream of literature connects OR to the behavior of people, by studying the following two topics: 1) modeling human behavior in complex settings, 2) investigating the role of behavioral aspects in the use of OR models (Franco & Hämäläinen, 2016). This thesis contributes to the first of these research goals.

In this thesis, I connect results from psychology to the modeling in OR. More specifically, I use results from psychology to improve our understanding of the relevant variables in decision making, and how psychological aspects influence outcomes of decisions. I investigate the interpretation of judgments of importance, and examine the influence of context and decision style measures for choice quality.

The approach of this thesis builds upon the idea of idiosyncratic behavior in decision making. Psychologists have long assumed that each individual has a certain type of behavior. These types are usually not deterministic, but just mean that an individual has a tendency to behave in a certain fashion. For example, one such type of behavior is decision making style (Hamilton et al., 2016; Thunholm, 2004). By studying decision

styles and connecting them to behavior in decison making situations, we can explain some of the differences in differing decision making performance between individuals.

1.1 Objectives and Scope

This Dissertation develops methods, models, and interpretations for combining multiattribute value theory with psychological measures about the decision task and the decision maker (DM), in a context where we have one individual as the DM. The research done for this Dissertation shows how the emotional attachment to a product, the decision environment, or the decision style of the DM influence the DM's behavior. More specifically, the research questions posed in this Dissertation are the following:

RQ1: Is a linear value function a suitable, robust simplification for preference prediction?

RQ2: How do task characteristics such as the amount or the distribution of information influence choice quality?

RQ3: Can psychological measures (such as cognitive reflection or decision style) be used to predict errors in decision making?

RQ4: Can we interpret judgments of importance in an MCDM setting with multiattribute value theory?

Table 1.1 shows which Essays consider which research questions.

Table 1.1. Relationship of the research questions and essays

	Essay I	Essay II	Essay III	Essay IV
RQ1	x			
RQ2		x		x
RQ3		x		x
RQ4			x	

Essay I extends the result of Korhonen et al. (2012) by utilizing the estimated linear weights in a more complex, four-criteria setting. In the article, we show that the assumption of a linear value function is not too restrictive for preference prediction. Even though consistency with a linear

value function turns out to be a variable that does matter, the difference in prediction power is not large between individuals that are consistent with a linear value function, or those who are inconsistent with one.

Essay II investigates the effect of task characteristics for errors in decision making. We find that emotional attachment to the product type influences the DM's ability to choose nondominated alternatives in an MCDM task. Furthermore, we find that an increased amount of information can impair choice quality, depending on the nature of the extra information that is provided.

Essay III investigates the meaning of judgments of importance. These statements, for example "rent is more important than apartment size", have been traditionally hard to analyze in the framework of multiattribute value theory (Choo et al., 1999). Building on an idea of impact from previous literature (Goldstein, 1990), we try to come up with promising interpretations of such judgments. It turns out that there is no connection to linear value function weights, but by using AHP weights and coefficient of variation to define impact of the criterion, we can find a correlation between impact and the judgments of importance. This result supports the theoretical point of Goldstein (1990) that the impact of a criterion could be meaningful in explaining judgments of importance.

Essay IV uses psychological measures of decision making style and cognitive ability, together with measures of the information distribution, to examine which variables predict errors in an MCDM experiment. As expected, we find that most psychological measures have low explanatory power, with the Cognitive Reflection Test and Alternative Search scales having the highest explanatory power. Subjects with higher Cognitive Reflection Test and lower Alternative Search scores tend to make fewer mistakes.

1.2 Outline of the Dissertation

This Dissertation consists of this summary essay and the four original essays. The rest of the summary essay is structured as follows. Section 2 describes the theoretical background of the research in this Dissertation. Thereafter, in section 3 the used research methods are described and mo-

tivated. Section 4 presents a short overview of the findings of each of the four essays. Finally, section 5 provides a summary of the research done for this Dissertation, and ends with conclusions, managerial implications, limitations, and some future research suggestions. Part two of the Dissertation includes the original essays.

2. Theoretical Background

2.1 Multiple Criteria Decision Making

Multiple Criteria Decision Making (MCDM) is a field of economic research that looks at decision making in situations, which have multiple criteria that influence the desirability of an alternative (Ehrgott et al., 2010). This means that when making a decision, all these different criteria have to be taken into account, since translating them to a common unit is far from trivial. A good example of an MCDM case is buying an apartment. There are many attributes to consider: the size, the location, the condition, and the price of the apartment.

This Dissertation focuses solely on decision making by individuals. MCDM cases can involve organizations, groups, or individuals. Organizations tend to face complex decision situations most often. However, modeling decision making in groups and organizations would necessitate modeling group dynamics, and taking into account all kinds of additional influencing factors in the process. Hence, the focus on decision making by individuals enables the modeling of an individual's decision making process with greater fidelity.

2.2 Modeling Decisions in MCDM Research

In MCDM, a decision is typically a situation, in which we have a decision maker (DM) who is facing two or more alternatives, denoted as set $X = \{X_1, X_2, \dots\}$. The alternatives are defined in a space of criteria $\boldsymbol{C} = (c_1, c_2, c_3, \dots, c_k)$ with alternative-specific criterion values $\boldsymbol{X_i} =$

$[x_{11}x_{12}\dots x_{1k}]^T$, and the DM cares about at least two of these criteria in his decision. Therefore, the decision maker would like to choose the alternative, which is the best alternative for her, considering her objectives, and the performance of all the alternatives on these multiple criteria. Mathematically speaking, if we assume an additive value function, the decision maker's value from an alternative can be modeled as

$$u(X_i) = \sum_{j=1}^{k} v_j(x_{ij}) \in [0,1] \tag{2.1}$$

What this means is that, for each criterion k, we have a value function v_k, which maps the criterion value x_{ik} to the value gained from it. The total value of the alternative X_i is then the sum of these criterion-specific values.

The problem of finding the solution to the decision problem of a DM can be divided into a few subproblems:

1. How can the desires of the DM be written out mathematically?
2. How to model tradeoffs between different criteria?
3. Does the DM search for an optimal solution, or is "good enough" good enough?

The first of these problems considers the main problem in modeling: how can we make a mathematical model out of the decisions of the DM? When a DM is making choices between products, we have to make assumptions about which criteria the DM considers in their decision process, which alternatives are in the choice set, and how the DM differentiates between high and low values on the criteria at hand. For example, the DM could just classify criterion values simply as acceptable/unacceptable. On the other hand, the DM might employ a logarithmic function, so that a higher value is better, but with decreasing marginal value. Obviously, there are an infinite number of ways to model the value function or the decision process of the DM, which necessitates the usage of assumptions.

The second problem asks whether the model is compensatory or noncompensatory in nature, and how this is mathematically specified. A compensatory decision model allows a criterion to *compensate* for another criterion. For example, I can be willing to accept a car in bad condition, if it is

cheap enough. Here, the low price compensates for the low quality. What differentiates compensatory models from each other is the way compensation is modeled. In value function models, each criterion contributes a portion of the final value of the alternative, and these portions are then aggregated according to a value function. In other models, such as the tallying heuristic, we just count the number of criteria that reach some level of acceptance, and then select the alternative that has the highest tally. A noncompensatory model, in contrast, does not allow for compensation. In these models, if a criterion value is too low, the alternative will be excluded from consideration, regardless of the values of other criteria. A noncompensatory model would discard a car, which is too expensive, no matter how good the condition of the car.

The third question is especially relevant for experimental research of decision making. When faced with a decision problem, a DM can behave in a maximizing or a satisficing fashion. When maximizing, the DM tries to find the optimal solution or alternative, and hence choose as well as they possibly can. If the DM is satisficing, they just look for an alternative that is "good enough", and then choose the first alternative that reaches this level of acceptance. Obviously, when conducting experiments, it is important to pay attention to the behavior of subjects in order to determine, whether the subjects are trying to choose as well as they can, or whether they are trying to satisfice.

Different MCDM models diverge in their answers to the above questions. The theoretical base of this Dissertation, Multiattribute Value Theory (MAVT), assumes that the preferences of the DM can be written out as a value function, and that tradeoffs between criteria will be explicit in the weights of the value function. Finally, MAVT says that the DM should choose the option that gets the highest value score. In comparison, a lexicographic decision model (Baucells et al., 2008; Baron, 2000) would say that the DM's wishes can be written out as a set of aspiration levels, one for each criterion. Then, starting from the most important criterion, all alternatives that do not meet the aspiration level will be ignored. What will be left is only alternatives that meet all the aspiration levels of the DM, and then the decision between these can be made at random.

2.2.1 Estimating the Value Function of the DM

Traditionally, value function analysis in MCDM has built upon the approach used in decision analysis. In this stream of research, a lot of effort is spent on posing questions to the DM, and thereby eliciting the form of his or her value function (Morton & Fasolo, 2009). Unfortunately, when we have to elicit the value function for several subjects, such an approach would be prohibitively costly (Salo & Hämäläinen, 2010). Hence, in MCDM cases with many subjects, the idea is to try to find a simpler form of a value function that could be used. In our research, we have assumed that the DM's value function is linear. Even though this might not be true, Korhonen et al. (2012) provided evidence that such an assumption still might be suitable for preference prediction.

2.2.2 Behavioral Aspects

The approach of multiattribute value theory assumes that all that matters in the decision making situation are the values of the criteria for different alternatives. However, as decades of psychology and behavioral economics has shown, this might not be true (Montibeller & von Winterfeldt, 2015; Morton & Fasolo, 2009). In fact, people are susceptible to many kinds of effects, arising from reference points (Tversky & Simonson, 1993), framing of the alternatives (Tversky & Kahneman, 1981), and so on.

In this Dissertation, we have investigated several different factors that influence decision making, besides just the criterion values. These factors can be divided into *cognitive factors*, *decision task factors* and *decision environment factors*. Cognitive factors are effects arising from the individual characteristics of the DM. For example, the capability to suppress one's System 1 response and use System 2 instead is a crucial factor in situations that arouse an immediate yet incorrect System 1 response (Evans & Stanovich, 2013).

Decision task factors are the task's and the alternatives' properties. For example, the type of the product that is being chosen is a decision task factor. The type of the product can influence decisions, as some products, for example music, are more hedonic and generate an emotional response. Other products, for example scissors, have utility because they are useful.

Finally, decision environment factors are properties of the decision case in connection to the goal. For example, the distribution of information is a property of the decision environment. The distribution of information describes how different criteria contribute to the differences of the alternatives, and therefore can vary from "all criteria contribute equally" to "there are large differences in the contribution of criteria".

2.3 Cognitive Aspects of Decision Making

Decision making as a topic is not confined to the field of economics and business. Psychologists have also studied phenomena about decision making for several decades, and in fact the rise of behavioral economics as a field is largely due to the influence of psychologists on economics and business research (Katsikopoulos, 2014). This section will briefly summarize some main concepts and ideas of psychology, which form the basis for the research in this Dissertation.

One main concept, which has been crucial for most theories in behavioral economics, is the division between System 1 and System 2. They are meant to describe two different ways of processing and making decisions. System 1 is fast, associative, intuitive, implicit, and hard to control. System 2, on the other hand, is slow, purposeful, deliberate, and used with control. If you are asked whether the economic situation in your country is moving in a positive direction, a System 1 answer could be based on how many news articles you happen to remember about the topic, or how well your friends are doing economically. System 2, on the other hand, would recognize that media information is selective and hard to remember correctly, and therefore would probably reason that you need to look up actual statistics to be able to give a reliable answer. An important detail of the division between these systems is that they are not actually separate systems of circuits in the human brain. In fact, Keith Stanovich prefers the term "Type" in the place of "System" to combat this tendency (Evans & Stanovich, 2013).

The division between systems forms the basis for the idea that there are different ways of making decisions. Whereas microeconomics focuses on modeling decision making with the help of value functions, psychologists

have come up with different kinds of decision models, which are meant to describe *how* the DM makes their choices. For example, psychologists have investigated whether people in fact look at all the information that they would need to, if a linear value function was in fact a descriptively true model of the decision making. Other models, like the lexicographic heuristic, look at the alternatives one criterion at a time, and discard options that are inferior on a criterion. This leads to selecting ultimately only an alternative that is best in the first criteria that are looked at.

One important conceptual difference between the previously mentioned linear value function and lexicographic heuristic is that the first one takes into account values of all criteria, while the other one sequentially samples information. Hence, the decision models differ on the amount of information that they use. Theoretically speaking, the more information a model uses, the more finer distinctions it can draw between different alternatives.

Two important concepts that describe the decision maker on the individual psychological level are cognitive style and decision making style. Cognitive styles are *"stable attitutes, preferences, and habitual strategies which determine an individual's modes of perceiving, remembering, thinking and problem-solving"* (Dewberry et al., 2013). Decision making styles, on the other hand, are habitual patterns exhibited by an individual, when he or she is faced by a decision situation. Different decision making styles are defined in different psychological measures, but typically there is some kind of divide related to System 1 and System 2 processes. For example, Hamilton et al. (2016) define two decision styles: intuitive and rational. These are quite clearly linked to the two modes of processing commonly called System 1 and System 2.

Whether we should treat decision styles as a subset of cognitive styles tends to divide researchers into different schools of thought. Some (e.g. Thunholm 2004) believe that cognitive styles are a subset of decision styles. Others (e.g. Andersen, 2000) use the terms interchangeably. The third school, and the school this Dissertation belongs to - believes that decision styles are a subset of cognitive styles (Hamilton et al., 2016).

Psychological and economic approaches to decision making have differed previously based on how they treat the decision environment. Economic decision theory models have tended to assume that the subject tries to

maximize their utility, which often leads to using all, or at least most of the available information. The subject combines available information rationally, and bases his conclusions about utility of different alternatives based on these considerations. Being more interested in the decision process than economists, psychologists have looked at information usage, and questioned the assumption that subjects use every available nugget of information. For example, Gigerenzer has advocated strongly for simple heuristics, which use only a subset of the information, but still result in good decisions (Gigerenzer & Gaissmaier, 2011; Gigerenzer, 2008).

This divide regarding assumed use of information can be conceptualized through decision environment and satisficing. To take the latter concept first, satisficing means that the subject is happy with an alternative that is "good enough", and eschews the chance to optimize further (Katsikopoulos, 2014). This can even be rational in the utility-maximizing sense: if gathering information is costly, then ignoring some information may allow for faster decision making, and hence leave more resources for the future. This idea is not only prevalent in psychology, but information gathering costs have been featured also in economic models (Katsikopoulos & Fasolo, 2006; Simon, 1955).

The concept of a decision environment, on the other hand, has mostly been used in psychology. The idea is that the distribution of the information, and the goal of the decision drive human behavior. With more unequal information distributions (i.e., alternatives being more different from each other), it may make sense for the DM to ignore criteria that vary little between alternatives, because inspecting them is not usually going to result in better decisions. The importance of looking at the decision behavior in connection with the task and the environment has been heavily emphasized in recent decades, in comparison to older streams of research, which compared behavior just to mathematically optimal models of decision making (Gigerenzer, 2008; Hogarth & Karelaia, 2007).

2.4 Descriptive and Prescriptive Decision Analysis

A decision analyst, when modeling the behavior of a DM, must always ask the following question: are we modeling decisions as the DM *in fact makes*

them, or as the DM *should make them*. These positions are called descriptive and normative approaches to decision making (Edwards & Fasolo, 2001). Further, there is the question of how can decision makers improve their performance from their current one. This is called the prescriptive approach to decision making.

Depending on the pretheoretical stance on the relationships between the normative, descriptive, and prescriptive accounts, there are three alternatives to conceptualizing decision making called the Meliorist, Apologist, and the Panglossian position (Stanovich, 1999, pp.4-8). The Panglossian believes that human decision making is close to optimal: normative, descriptive, and descriptive accounts all coincide. The Apologist believes that human decisions are far from the normative, but there's not much that can be done about it. This means that for the Apologist, descriptive and prescriptive behavior coincide, but the normative position is far from them.

The position taken in this thesis most closely resembles the Meliorist position. To keep the explanation short, in this stance we assume that while the prescriptive model of the decision making is close to the normative one, there can be a large gap between the descriptive and the prescriptive models. In plain language, this means that humans can be close to optimal decision makers, when they perform according to the best of their ability. However, in many cases the actual performance level of decision makers falls short, as they are susceptible to a variety of biases and lapses of focus or attention.

However, as Stanovich (1999) notes, taking the Meliorist position too strongly is also inadvisable. It is not necessarily true that all errors in decision making are the cause of human irrationality. In some cases like Bayesian reasoning (e.g. Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995), the cause of many errors may be a way of representing information that is unsuitable for human cognitive machinery. Trying to meliorate errors in such cases by education (as the extreme Meliorist would like) is futile and inefficient, compared to just presenting the information in a better format.

This thesis is broadly situated in the aforementioned Meliorist stream of literature. As Essays II and IV look at errors in decision making, it is by definition already clear that we assume that errors in decision making

are possible. On the other hand, while the Apologist would think that improving human decision making is a somewhat futile project, we see the issue in a more positive light. The errors that are investigated in this Thesis are simple, and in our opinion could well be mitigated with better decision making behavior. We do not believe that the decision tasks in this Thesis are too complex for DMs to thoroughly process them. In fact, the large portion of correct choices in our tasks seems to imply that avoiding errors is well within the range of a decision maker's cognitive capabilities.

3. Research Methods

This chapter provides details about the used research methods in this Dissertation. Since all the research for this Dissertation is based on laboratory experiments, I will first start with a short introduction to experimental research. Then, I will describe some particular methods used in the Dissertation. First, I will explain the epsilon method, which is used to estimate the weights of the decision maker's value function. Secondly, I will show how the Dissertation uses Bayesian hierarchical models to estimate effects in a decision experiment with many subjects, all of whom make several choices. Finally, I will provide information about doing open science, since the research for Essay IV has been conducted with this scientific approach in mind.

3.1 Experimental Research

Experiments are a way to test hypotheses about the relationships of variables. For example, if I believe that drinking a cup of coffee will make people more attentive, I could test this in the lab by having people drink coffee, and then do a measurable task, in which performance depends on attention. Experiments can be divided into two categories: laboratory research and field research. In laboratory research, the task is carefully designed, and subjects either come into the lab to participate, or complete the task online with their own computers. In field research, the study is conducted outside the laboratory, and the task that subjects do is usually more realistic. All the research for this Dissertation was done in the laboratory.

The benefit of laboratory research is the tight control of the task. In the

lab, we can design the whole task from scratch, controlling the duration, order, structure, and the content of the task completely. In the case of decision making, we have full control over what information subjects see, and in which order they see that information. In a field experiment, for example when using real products, we could not manipulate the informational content such as brands. In such a case, if the brand is not an interesting variable for the research question, conducting a lab experiment allows us to leave that variable out.

3.2 The Epsilon Method for Estimating the DM's Value Function

With the epsilon method, we assume that the decision maker's preferences can be represented with a linear value function. When assuming the linear value function, we model the DM's preferences according to the function $v(X) = \sum_{j=1}^{p} \lambda_j x_j$, where

- p is the number of criteria,
- N is the number of alternatives,
- x_j refers to the value of the jth attribute,
- λ_j refers to the weight of the jth attribute, and
- ϵ is the difference in value between alternatives A and B.

Preferences are assumed to reflect the equation $v(X_A) - \epsilon > v(X_B)$, when alternative A is preferred to B. This gives us a preference set $P = \{(X_r, X_s) | X_r \succ X_s, r, s \in N\}$.

Using the DM's responses, constraints for the LP weight estimation problem are constructed. For each "I prefer A over B" response, the following inequality restriction is constructed:

$$\sum_{j=1}^{p} \lambda_j x_{Aj} - \epsilon \geq \sum_{j=1}^{p} \lambda_j x_{Bj}$$

A similar constraint is generated conversely for each "I prefer B over A" response.

Now, the ϵ can be solved by simple linear programming. If we find that $\epsilon > 0$, it means there exists a linear value function that the subject is consistent with. Conversely, if $\epsilon < 0$, then it isn't even theoretically possible

to find a linear value function that would explain all the answers of the DM. This follows from the above equation, since we know that alternative A was in fact preferred to B, and hence their value difference should be positive.

A difficult question arises when the DM answers that they cannot choose between A and B. There are two possible causes for this: 1) the alternatives A and B are truly equal, or 2) the DM doesn't know which they prefer. From a value function perspective, the first case implies that the values of A and B are the same. However, even a difference in the sixth decimal is theoretically enough to make the values differ. In practice, this creates a problem for the linear programming, because as even a tiny difference breaks the equivalence constraint, this places a heavy restriction on the λ_j. Therefore, we have assumed that in such cases, we look at the data, and try to empirically estimate how large the ϵ, the difference of A and B, can be, so that the indifference is still modeled correctly.

3.3 Bayesian Hierarchical Models

Bayesian statistics offers the opportunity to combine a prior probability and likelihood to form a posterior probability. For example, if a subject is examined for a disease, we can combine the prior probability of having the disease (the rate of the sickness in the base group) with the likelihood (a result of a diagnostic test run on the subject) to form the posterior (the probability that the subject has the particular disease). Whereas frequentist statistics only look at the probability of the data under a hypothesis $p(D|H)$, Bayesian methods look at the probability of a hypothesis given the observed data $p(H|D)$. The end result of a Bayesian analysis is usually a distribution that allocates probabilities to various hypotheses, which can be different values of a parameter, for example.

The benefit of using Bayesian models is that they take into account uncertainty in parameter values, and the results are easy to interpret. Traditional null hypothesis significance testing (NHST) would compare the probability of observing the data (or more extreme data) under two competing hypotheses $H0$ and $H1$, and result in an estimate of the effect size with a p-value, which is meant to answer the question "Would this data

be surprising under the null (i.e. $H0$)?". One difficulty with NHST is that confidence intervals (CIs) are hard to interpret, since by definition they only say that, for a 90% confidence interval, in the long run we have 90% chance that a randomly chosen CI generated by the repeated process contains the true parameter value. Notably, this doesn't mean that there is 90% chance that the CI of a *particular study* would include the true parameter value (Morey et al., 2016).

In contrast, a Bayesian analysis gives us *credence intervals*, which tell us the posterior distribution of a parameter estimate. The credence interval tells us how likely different parameter values are, given the data we have observed (and the prior we used). From this distribution it is easy to see the uncertainty related to the parameter, and judge how certain we are about our estimate.

In this Dissertation, Essays I and IV use Bayesian hierarchical models. Essay I uses them to estimate the posterior probability that a particular prediction model can make a correct prediction about the alternatives a decision maker prefers. In Essay I, the model takes in a set of predictions for various prediction models, and compares these predictions to the alternative the decision maker has in reality preferred. Essay IV estimates the posterior probability of making an erroneous choice, given psychological variables about the DM.

Since in both Essays each decision maker makes several choices, we can model the situation as a hierarchical model, which partially pools the information obtained from observing many decision makers. The first level of the hierarchy are the choices made by one decision maker. On the second level are all the different decision makers. After making assumptions about the variance between choices and between decision makers, a hierarchical model can be built. This model then takes into account both the shared variance between choices but within subject, and the shared variance between subjects. Such a model allows for better estimates than modeling each decision maker as a completely independent entity. On the other hand, the model also allows for the fact that different decision makers each have their own slightly different pattern of making choices, distinct from other decision makers.

3.4 Open Science and Preregistration

In recent years, especially in the domain of psychology, there has been a heated discussion about how to respond to the recent failures of replication studies, and how to generate more stable findings. One alternative put forward has been an idea broadly called Open Science. Proponents of Open Science include various parts of research under this umbrella term. Commonly, what is at least included is a recommendation to publish a research plan *before* gathering data and doing any analysis. Other proposed features of more open science are the publication of the gathered data, more careful a priori power analysis of experimental studies (Lakens & Evers, 2014; Maxwell, 2004; Cohen, 1962), better adherence to study protocols (Ioannidis, 2005), and even a reduction of the number of papers written per scientist (Nelson et al., 2012). Some journals, like the *Perspectives on Psychological Science* now also welcome registered replications as publications (Simons et al., 2014).

The Open Science proposals broadly target the same problem: reduction of findings that fail replications, being thus probably false positives. No matter the cause, the proliferation of false positives and nonreplicable findings is a problem for research. The discussion has been most vocal in the domain of psychology, but there is no reason for the improvements to be confined to that field. In general, any field would be improved by more stable findings.

The research for the final essay of this Dissertation has been done according to principles of Open Science. The research plan for the study was published before data gathering, and is available online[1]. The plan specified the a priori designed analyses for the study, the decision rule for stopping data gathering, and the hypotheses for the experiment. The predictions were frozen online, which means that changing them after having seen the data would be impossible.

The preregistration of studies is one possible solution to combating what Uri Simonson has called p-hacking, or researcher degrees of freedom. Essentially, p-hacking means that given various variables, their statistical transformations and an array of possible tests, every study will necessarily result in some results that are under the traditional threshold of

[1]The material can be found here: https://osf.io/bq3m7/

significance $p < 0.05$ used in psychology. If hypotheses have not been pre-specified, then a researcher could in principle go through the data, and try out all these different models and tests, and only report the ones that turned out to be significant. Clearly, such a way of doing research would result in a very inflated false positive rate, and be highly unethical.

However, p-hacking does not need to be conscious unethical behavior to be a problem. Just the simple fact that authors tend to drop analyses that were not significant, and include only the statistically significant ones is p-hacking, and leads to an inflated rate of false positives. Additionally, in many cases the data offer different ways of formulating the same theoretical idea. Gelman & Loken (2013) offer the insightful example of a study regarding difference in mathematical ability of Republican and Democrat voters. The hypothesis is that the context makes a difference, so whether the subjects are working with numbers about the military or health care matters. Now, there are many ways of formulating this statistically: maybe the difference appears for the whole group, or only for men (perhaps due to men being more ideological). Or it could be that the difference in scores between contexts is significant, and so on. If the researchers formulate their hypothesis based on only the test or tests that have shown up significant, yet ignore the other alternatives, we will face an inflated false positive rate. However, the researchers are not doing anything unethical, since all the different versions of the hypothesis seem equally valid. The problem lies in the fact that their choice of test depends on the data. As Gelman & Loken (2013, p.11) put it:

> The researcher degrees of freedom do not feel like degrees of freedom because, conditional on the data, each choice appears to be deterministic. But if we average over all possible data that could have occurred, we need to look at the entire garden of forking paths and recognize how each path can lead to statistical significance in its own way.

Preregistration protects a study against inflation of false positives. Without preregistration, hypotheses could be invented by looking at the data, and fitting the hypotheses to what the data shows. Naturally, such an approach invalidates all statistical measures, such as p-values. A hypoth-

esis is interesting only when it is truly a priori. A post hoc formulated hypothesis only proves that the data is what it is, and nothing more.

Another point of discussion, and one that has been spoken of in many fields, is the so called file drawer problem. The file drawer problem is that only studies with positive results tend to be published, while negative or insignificant results are stored in the researcher's "file drawer". Note that whereas p-hacking refers to publishing significant *analyses* (one study including several analyses), file drawer refers to publishing significant *studies*. Why is this an issue? Well, naturally this also inflates the estimate on effect sizes. For example, if a drug is evaluated 10 times, but only 5 out of 10 studies find find a positive effect and are the only ones that get published, our effect size estimate is bound to be positive. However, if we would include the 5 unpublished studies that didn't find a statistically significant effect, our estimate of the true effect size would go down.

Registering research before gathering the data is one solution to the file drawer problem. If the analyses in the study are registered before data is gathered, those analyses will be included no matter their result. In the end, the study will include these analyses, because dropping them would mean not complying with the preregistered plan.

3.5 Psychological Measures

The measurement of decision making can be done from several perspectives, as noted before in Section 2.3. In Essay IV, I use measures of cognitive capability, decision making style, and maximization tendency to predict the decision making process and accuracy of subjects. The used measures are the Cognitive Reflection Test (Frederick, 2005), the decision styles scale (Hamilton et al., 2016) and the maximization inventory (Turner et al., 2012). To see the text of the questions in each measure, see Appendix 2 of Essay IV.

The Cognitive Reflection Test (CRT) is one of the most used measures of cognitive ability in the decision making literature. The purpose of the CRT is to measure how well the subject can suppress their incorrect System 1 response, and replace it with the right System 2 response. Consisting only of three questions, the measure is very easy to administer,

and hence quite popular. In recent years, there has been a discussion about the measure's validity, since the three questions are now quite well known. Toplak et al. (2014) have extended the test to ten questions. However, to keep the experiment reasonably short, I used the original version of the measure.

The decision styles scale (DSS) (Hamilton et al., 2016) measures decision styles with a 10-question measure. The scale is subdivided into two parts, with five items assessing rational decision making and five items assessing intuitive decision making. The DSS defines rational style as *"being characterized by a thorough search for information and a systematic evaluation of all choices and potential alternatives"*, while intuitive style is quick decision making, mostly based on hunches and feelings.

The maximization inventory (Turner et al., 2012) consists of 34 questions that generate three separate scales measuring decision difficulty, alternative search, and satisficing. The first scale, satisficing, reflects the tendency of the decision maker to settle for options that are good enough, instead of looking at maximizing alternatives. Decision difficulty measures the ability to make decisions, even when there are plenty of alternatives, thus tapping into the emotional distress generated by decision making. Finally, alternative search focuses on the behavioral aspect of gathering more information to make a choice.

4. Overview of the Findings

4.1 Essay I: Road to robust prediction of choices in deterministic MCDM

In MCDM research, finding the true form of the decision maker's value function can be an arduous task. For this reason, in many cases it is convenient to assume that the DM's preferences can be represented by a linear value function.

In the study, linear value functions with various weights were used to predict choices of student apartments. The apartments were defined by four criteria: size, montly rent, distance to the university, and distance to the city center. The used weight schemes were weights estimated with linear programming, weights estimated with the AHP methodology (Saaty, 2008, 1980), and equal weights. Additionally, we used a logistic regression and a lexicographic model to compare the linear value function approach to two popular models that do not share the same linear value function assumption.

A Bayesian hierarchical model was used to plot the posterior probabilities of each model making a correct prediction. We were able to show that the weights estimated with linear programming achieved the best prediction accuracy. However, the AHP weights were a close second, and the difference was not statistically significant. All other models and weight schemes had lower accuracy rates, and were statistically significantly worse. These findings are in contrast to what Dawes (1979) and Lovie & Lovie (1986) call the *flat maximum effect*. The flat maximum effect means that in many cases with linear models it's only necessary to pick the right

variables - the weights are less important. However, in our experiment we showed that the unit-weighted model fared worse in prediction than AHP or estimated weights, contrary to the flat maximum effect.

We also found that the results depended on whether the DM was theoretically consistent with a linear value function. When there existed theoretically a linear value function that would be consistent with the DM's choices, the prediction rate was statistically significantly higher.

Additionally, the difference in value of alternatives A and B (estimated with the linear value function with weights by linear programming) was a variable affecting the prediction accuracy. The higher the difference in value between A and B, the higher the rate of successful predictions. This was not surprising, since when A is a very attractive alternative, and B has some very bad attribute values, it is easy to predict that A will be chosen. On the other hand, prediction is hard, e.g. when A and B are very similar, and the choice comes down to for example whether the DM would rather have a slightly larger apartment or a smaller distance to school.

4.2 Essay II: Context matters: The impact of product type, emotional attachment and information overload on choice accuracy

Decision research sometimes treats all choice situations as independent of the context in which the choosing happens. Whether the decision maker is choosing between cars, apartments, or employees, in all cases they are trying to maximize value, and therefore similar theoretical frameworks can be applied.

However, research in marketing has shown that different types of products can be desired for different reasons. Hedonic products generate emotions in us, while utilitarian products are useful. In this essay, we show that this division of products to hedonic and utilitarian products is meaningful also for multiple criteria decision making research. When subjects choose between bundles of products, the product type influences the rate of erroneous choices the subjects make. We find that high emotional attachment to hedonic products increases choice accuracy, while low emotional attachment decreases it. For utilitarian products, there is no relationship between emotional attachment and choice accuracy.

Furthermore, we are able to show that, in contrast to some research in marketing (Iyengar & Lepper, 2000; Scheibehenne et al., 2010), choice accuracy does not depend on the quantity of information in a straightforward manner. In fact, also the quality of information matters. When we add a third criterion that does not change the dominance relations of alternatives, choice accuracy does not go down. This shows that adding information that makes dominance relationships easier to notice can in fact result in a leveling of choice accuracy, despite the added burden of extra information.

Our results are based on an experiment with incentivized hypothetical choices. In the experiment, subjects had to choose between bundles of products including various amounts of either music songs from different genres, or office supplies. These represented the hedonic and utilitarian product types, respectively. The choice problems were generated so that in each choice situation, some alternatives were dominated by other alternatives. Whenever the DM chooses a dominated alternative, we count this as a decision error, because they could have chosen an alternative with higher values. In this, we only need to assume that more is better for each criterion, which is plausible, since pretests had shown the products to be desirable. Additionally, since they have low storage and maintenance costs, it would be hard to argue that they would have negative value for the DM.

4.3 Essay III: Judgments of importance revisited: what do they mean?

Decision makers often make statements about the relative importance of criteria. For example, when looking for an apartment, they can say that *"rent is more important than location"*. However, under current theoretical frameworks, utilizing such statements is difficult. Some authors even say that such statements are completely meaningless, and they ought to be ignored (Morton & Fasolo, 2009).

Goldstein (1990) argued that judgments of importance can be represented as impact, which is defined as the product of the criterion weight and the criterion variation. However, Goldstein left open the mathematical definition of impact. Variation, for example, can be defined with the

criterion range, variance, coefficient of variation, or average range. Likewise, criterion weight can refer to the estimated weights of some value function, or weights that have been elicited with some other method.

We had subjects make pairwise judgments of importance between four criteria about rental apartments: size, price, distance to the university and distance to the city center. We used various definitions of impact and measured the correlation of the pairwise importance judgments that these imply with the original elicited pairwise importance judgments. We discovered that when impact is defined as the product of weights elicited with the AHP method and the coefficient of variation of the criterion, the correlation was very high. In fact, it was almost as high as the correlation of AHP weights and the original judgments, which was used as a benchmark. We did not find any differences according to whether the subject was consistent with a linear value function or not.

The results provide a possible avenue for interpreting judgments of importance. Since subjects can easily make such statements, it would be good if we had a method for utilizing them in analysis. Our results are a first step in that direction, showing the plausibility of impact as an interpretation.

4.4 Essay IV: Explaining choice quality with decision style, cognitive reflection and decision environment

Research psychologists have spent considerable effort in designing psychological measures that tap into the individual differences of decision makers. For example, Hamilton et al. (2016) measure the decision style of the individual, which means their tendency to use an intuitive or rational decision style. The Cognitive Reflection Test (Frederick, 2005) measures the decision maker's ability to suppress wrong System 1 responses, and replace them with a System 2 response. The maximization inventory of Turner et al. (2012) measures whether the DM is a satisficer or a maximizer, how commonly they experience difficulty in making these decisions, and whether they tend to expend a lot of effort in searching for the best solution among alternatives.

Essay IV was completed according to the procedures of open science. This means that the research plan was published beforehand online, and

it stated the data collection and storage plan, and listed the hypotheses and analyses we had planned a priori. Unfortunately, a programming error in the experiment resulted in loss of data that rendered some of the hypotheses impossible of investigation.

In the analysis, we investigated the link of psychological measures to errors in decision making. In the experiment, subjects were provided with six candidates for a job, and they had to select their two best candidates. Each candidate was defined by six numeric criteria. We had designed the alternatives so that there were always some candidates that were dominated. Selecting a dominated candidate was counted as a mistake, since another candidate would have had a better profile.

We found that the CRT was related to errors, namely that the higher the subject's CRT score, the lower the probability of making an error. Additionally, higher alternative search was related to higher rate of errors. Despite good statistical power, no other psychological variables were connected to higher error rate. Also contrary to previous literature (Bröder, 2003) we did not find an interaction between cognitive ability, and the ability to adapt to the environment. In our study, those with a high cognitive ability score were equally likely to make errors in both decision environments.

5. Summary and Conclusions

This chapter presents a summary of the findings of the Essays and comments on managerial implications, limitations of the studies, and further research topics arising from the studies.

5.1 Summary of Findings

Essay I showed that a linear value function can be used to predict preferences in a multiple criteria choice situation. A Bayesian hierarchical model provided evidence that weights estimated with the epsilon method, and weights generated with the AHP method were the two best weight schemes for preference prediction. In contrast to Korhonen et al. (2012), we found that whether the subject was consistent with a linear value function had an impact on the probability of a successful prediction.

Essay II showed that the context of decision making matters. Product type and emotional attachment interact, so that subjects made fewer errors with high emotional attachment and hedonic products. Furthermore, providing more information did not result in information overload.

Essay III found that the most promising mathematical interpretation for judgments of importance was the product of AHP weights and the coefficient of variation. Other mathematical definitions had lower correlations. However, there was still quite a lot of variation between individuals, and it could be that there are other definitions that work better for some subjects.

Essay IV found that most tried psychological measures predict errors only very weakly. The best predictors all in all were the compensatory nature of the decision environment, the score on the Cognitive Reflection

Test, and the score on the alternative search subscale. The higher the score on the CRT, or the lower the score on alternative search, the lower the probability of choosing a dominated alternative in the experiment.

Taken all together, the results presented in this Thesis show the importance of accounting for behavioral and psychological issues in Multiple Criteria Decision Making. The results point to interesting and important differences between what formal models assume and predict on the one hand, and how real decision makers behave on the other hand. Essay I showed that decisions can be predicted with a linear value function, even though the formal model's assumption of linearity was not met. Essay II showed that, whereas formal models assume description invariance, real decision makers are impacted by the context of the decision. Essay III showed that connecting judgments of importance - a variable that seems to make sense for real decision makers - to a formal model is fraught with difficulty. Finally, Essay IV took a step in the other direction, and found that psychological variables that a priori could be important in explaining decision behavior, failed to have an impact in the prediction of errors.

The so-called Rationality Wars have been a hot topic in the decision making literature for decades (Sturm, 2012; Samuels et al., 2002). While the heuristics and biases stream found an ever increasing number of cases of decision makers' irrationality (e.g. Kahneman, 2003), the ecological rationality researchers claimed that in many cases accounts of irrationality were overblown, and simply the result of bad formatting or instruction (Gigerenzer & Gaissmaier, 2011; Gigerenzer, 2008). This thesis, while not specifically discussing these issues, does at least lend credence to the idea that the context of the decision matters. While traditional value theories assumed that choices are description invariant - and hence largely context-independent, Essay II in this Thesis provided one counterexample to that reasoning.

5.2 Managerial Implications

Even though the research was conducted in the laboratory, the findings still allow for some interesting managerial insights. Essay I showed that assuming a linear value function is a feasible method, even though for

some subjects there is no consistent value function even in theory. When businesses want to predict consumers' preferences, a simpler value function model may work as well as a more complex and computationally intensive alternative.

Essay II found that emotional attachment and product type interact when choice accuracy is measured by dominance. This might be of interest for businesses in ecommerce. In such systems, buying hedonic products is possibly more error prone, if the subject does not feel a high level of emotional attachment. Paying attention to generating enough emotional attachment might increase the quality of consumers' choices, and thereby decrease the number of returns, when consumers are unhappy with what they ended up ordering online.

Essay III investigated possible explanations for judgments of importance. We discovered that they are related to the impact of a criterion, and could perhaps be used in MAVT. For managers, this experiment provides evidence that judgments of importance are not meaningless statements, and could perhaps be used in holistic consumer evaluations of products.

Essay IV showed that cognitive reflection and alternative search explain decision errors. Further, despite high statistical power, the experiment found no difference between subjects who consider themselves very analytical, and subjects who were less analytical. This suggests that in such a numeric task, self-perceived analytical ability is not related to performance. These results and measures could be used by managers to evaluate job candidates, and predict their performance in some tasks.

Although the research in this Thesis is mainly descriptive in its nature, it allows for some prescriptive implications. The fact that Essay I showed the feasibility of using a linear value function in prediction signifies that finding the true model is not always necessary. Especially for companies, the aim in prediction is usually good inference, not the truth of the underlying model. Even though these two aims are often aligned, cost considerations may well favor using a simpler model, even if the model isn't strictly true.

5.3 Limitations

All the Essays are based on laboratory experiments, conducted mainly with students. Naturally, the limited subject pool limits the generalizability of the findings. However, we consider that subjects can be used as representative subjects, as long as it is ensured that they find the task interesting and meaningful. For example, making strategic business acquisitions is not a very understandable problem for students, but evaluating student apartments is.

While laboratory experiments enable good control of the variables that influence the phenomena under study, the experiments are also necessarily simplifications of real life decision making tasks. In the lab, subjects' attention can be directed only to the task at hand, and they cannot wander off to do something else for a while, returning to the task later. Additionally, when the reward for participating in an experiment is too low, subjects can fail to be motivated, which can jeopardize the quality of the data.

In the experiments done for this Dissertation, we had to simplify decision tasks considerably for the experiments. The apartment choice task used in Essay I only featured four criteria: price, size, and two location parameters of the apartment. In real life, apartments are defined by tens of criteria, not all of which can be easily reduced to numeric form. For example, the layout of the apartment is an important criterion that was left out of the experiment.

The product bundle choice in Essay II was done by reducing the bundles to numeric tables, ignoring the content of the songs. Even though the style of music was accounted for, we ignored the wide variation in pop music, for example.

The recruitment task in Essay IV only represented candidates by six numeric criteria. Ignored were all attributes of how the candidate presents himself, how they discuss, or what kind of impression they made in the interview. Naturally, this is not very representative of a real recruitment situation.

However, what we lose in the real-life representativeness of the experimental tasks, we gain in control of the variables. Setting the variables for a choice experiment allows us to make sure that all subjects receive

same amounts of information (and even exactly the same content, when desired). In a field study, such fine control is harder to achieve. Additionally, generating data in numeric form on the computer makes measuring repeated choices easier, something which once again would be hard to do in the field.

5.4 Further Research Suggestions

The experiments done for these Essays suggest several possible directions for future research. This section will outline some promising research questions and hypotheses.

Essay I showed that a linear value function approach is suitable for prediction purposes. However, the models that were used for comparison were necessarily limited, and the lexicographic model was the only one that did not involve a value function approach. It would be intriguing to compare the epsilon method against other models from the machine learning field, for example. Such models can be quite flexible, and might provide a suitable way of predicting preferences.

Essay II showed the importance of contextual considerations in MCDM research. A replication of the experiment with a larger sample would enable more robust findings. Additionally, trying out a more realistic version of the choice problem would enable investigation of the influence of emotional attachment and the possibility of ruling out any experimental artefacts arising from a tightly controlled laboratory experiment.

Essay III investigated interpretations for judgments of importance. A study that would focus solely on the different meanings of importance, and the use of impact in preference prediction would be worthwhile. The major drawback of the reported study was that importance was only measured with one method. A more multifaceted study could enable deeper inspection of the content of importance statements, and recheck whether the impact interpretation still holds true in these circumstances.

Essay IV discovered that most of the psychological measures used in the study are not useful in prediction of errors. The effect sizes of the used measures are simply too small to help in prediction. However, one major question is whether decision styles would have more influence in a

task with more ambiguous information. The experiment of Essay IV was tightly controlled, and each subject was presented with exactly the same information. Perhaps differences in decision style only become apparent, when subjects have to make tradeoffs between satisficing based on little information, or spending time and effort in analyzing more information.

References

Andersen, J. A. (2000). Intuition in managers: Are intuitive managers more effective? *Journal of Managerial Psychology*, *15*, 46–63. doi:10.1108/02683940010305298.

Baron, J. (2000). *Thinking and Deciding*. (3rd ed.). New York: Cambridge University Press.

Baucells, M., Carrasco, J. A., & Hogarth, R. M. (2008). Cumulative Dominance and Heuristic Performance in Binary Multiattribute Choice. *Operations Research*, *56*, 1289–1304. doi:10.1287/opre.1070.0485.

Bröder, A. (2003). Decision Making With the 'Adaptive Toolbox': Influence of Environmental Structure, Intelligence, and Working Memory Load. *Journal of Experimental Psychology. Learning, Memory & Cognition*, *29*, 611–625.

Choo, E. U., Schoner, B., & Wedley, W. C. (1999). Interpretation of criteria weights in multicriteria decision making. *Computers & Industrial Engineering*, *37*, 527–541. doi:10.1016/S0360-8352(00)00019-X.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: a review. *Journal of abnormal and social psychology*, *65*, 145–153.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*, 1–73. doi:10.1016/0010-0277(95)00664-8.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American psychologist*, *34*, 571–582.

Dewberry, C., Juanchich, M., & Narendran, S. (2013). Decision-making competence in everyday life: The roles of general cognitive styles, decision-making styles and personality. *Personality and Individual Differences*, *55*, 783–788. doi:10.1016/j.paid.2013.06.012.

Dyer, J. S., Fishburn, P. C., Steuer, R. E., Wallenius, J., & Zionts, S. (1992). Multiple Criteria Decision Making, Multiattribute Utility Theory: The Next Ten Years. *Management Science*, *38*, 645–654. doi:10.2307/2632682.

Edwards, W., & Fasolo, B. (2001). Decision Technology. *Annual Review of Psychology*, *52*, 581–606. doi:10.1146/annurev.psych.52.1.581.

Ehrgott, M., Figueira, J. R., & Greco, S. (Eds.) (2010). *Trends in multiple criteria decision analysis* volume 142 of *International series in operations research & management science*. New York ; London: Springer.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*, 223–241. doi:10.1177/1745691612460685.

Franco, L. A., & Hämäläinen, R. P. (2016). Behavioural operational research: Returning to the roots of the OR profession. *European Journal of Operational Research*, *249*, 791–795. doi:10.1016/j.ejor.2015.10.034.

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

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. URL: http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf.

Gigerenzer, G. (2008). *Rationality for Mortals: How People Cope with Uncertainty*. New York: Oxford University Press.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*, 451–482. doi:10.1146/annurev-psych-120709-145346.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological review*, *102*, 684.

Goldstein, W. M. (1990). Judgments of relative importance in decision making: Global vs local interpretations of subjective weight. *Organizational Behavior and Human Decision Processes*, *47*, 313–336. doi:10.1016/0749-5978(90)90041-7.

Hamilton, K., Shih, S.-I., & Mohammed, S. (2016). The Development and Validation of the Rational and Intuitive Decision Styles Scale. *Journal of Personality Assessment*, *98*, 523–535. doi:10.1080/00223891.2015.1132426.

Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, *114*, 733–758. doi:10.1037/0033-295X.114.3.733.

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*, e124. doi:10.1371/journal.pmed.0020124.

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*, 995–1006. doi:http://dx.doi.org/10.1037/0022-3514.79.6.995.

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

Katsikopoulos, K., & Fasolo, B. (2006). New tools for decision analysts. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, *36*, 960–967. doi:10.1109/TSMCA.2006.871798.

Katsikopoulos, K. V. (2014). Bounded rationality: the two cultures. *Journal of Economic Methodology*, *21*, 361–374. doi:10.1080/1350178X.2014.965908.

Korhonen, P. J., Silvennoinen, K., Wallenius, J., & Öörni, A. (2012). Can a linear value function explain choices? An experimental study. *European Journal of Operational Research*, *219*, 360–367. doi:10.1016/j.ejor.2011.12.040.

Lakens, D., & Evers, E. R. K. (2014). Sailing From the Seas of Chaos Into the Corridor of Stability Practical Recommendations to Increase the Informational Value of Studies. *Perspectives on Psychological Science*, *9*, 278–292. doi:10.1177/1745691614528520.

Lovie, A. D., & Lovie, P. (1986). The Flat Maximum Effect and Linear Scoring Models for Prediction. *Journal of Forecasting*, *5*, 159–168.

Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: causes, consequences, and remedies. *Psychological methods*, *9*, 147–163. doi:10.1037/1082-989X.9.2.147.

Montibeller, G., & von Winterfeldt, D. (2015). Cognitive and Motivational Biases in Decision and Risk Analysis. *Risk Analysis*, *35*, 1230–1251. doi:10.1111/risa.12360.

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, *23*, 103–123. doi:10.3758/s13423-015-0947-8.

Morton, A., & Fasolo, B. (2009). Behavioural decision theory for multi-criteria decision analysis: A guided tour. *The Journal of the Operational Research Society*, *60*, 268–275. doi:http://dx.doi.org/10.1057/palgrave.jors.2602550.

Nelson, L. D., Simmons, J. P., & Simonsohn, U. (2012). Let's Publish Fewer Papers. *Psychological Inquiry*, *23*, 291–293. doi:10.1080/1047840X.2012.705245.

Saaty, T. L. (1980). *The Analytic Hierarchy Process: Planning, priority setting, resource allocation*. McGraw-Hill.

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*, 83–98.

Salo, A., & Hämäläinen, R. P. (2010). Multicriteria decision analysis in group decision processes. In *Handbook of group decision and negotiation* (pp. 269–283). Springer.

Samuels, R., Stich, S., & Bishop, M. (2002). Ending the Rationality Wars: How to Make Disputes About Human Rationality Disappear. In R. Elio (Ed.), *Common Sense, Reasoning and Rationality* (pp. 236–268). volume 2.

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, *37*, 409–425. doi:10.1086/651235.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*, *69*, 99–118.

Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An Introduction to Registered Replication Reports at Perspectives on Psychological Science. *Perspectives on Psychological Science*, *9*, 552–555. doi:10.1177/1745691614543974.

Stanovich, K. E. (1999). *Who Is Rational?: Studies of individual Differences in Reasoning*. Mahwah, N.J: Psychology Press.

Sturm, T. (2012). The "Rationality Wars" in Psychology: Where They Are and Where They Could Go. *Inquiry*, *55*, 66–81.

Thunholm, P. (2004). Decision-making style: habit, style or both? *Personality and Individual Differences*, *36*, 931–944. doi:10.1016/S0191-8869(03)00162-4.

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

Turner, B. M., Rim, H. B., Betz, N. E., & Nygren, T. E. (2012). The maximization inventory. *Judgment and Decision Making*, *7*, 48–60.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458. doi:10.1126/science.7455683.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, *39*, 1179–1189.

Wallenius, J., Fishburn, P. C., Zionts, S., Dyer, J. S., Steuer, R. E., & Deb, K. (2008). Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. *Management Science*, *54*, 1336–1349. doi:http://dx.doi.org/10.1287/mnsc.1070.0838.

Weber, M. (1987). Decision Making with Incomplete Information. *European Journal of Operational Research*, *28*, 44–57.

This Dissertation combines ideas from psychology with theory of Multiple Criteria Decision Making. Experiments showed that incorporating these ideas into models of decision making can help in making better predictions about decision making errors.

Aalto-DD 207/2017

ISBN 978-952-60-7682-9 (printed)
ISBN 978-952-60-7683-6 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Business
Department of Information and Service Economy
www.aalto.fi

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL DISSERTATIONS