



HAL
open science

The power of context in decision-making and recommendations

Emil Mirzayev

► **To cite this version:**

Emil Mirzayev. The power of context in decision-making and recommendations. Economics and Finance. Université Côte d'Azur, 2023. English. NNT : 2023COAZ0017 . tel-04283165

HAL Id: tel-04283165

<https://theses.hal.science/tel-04283165v1>

Submitted on 13 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT

Le pouvoir du contexte dans la prise de décision et les recommandations

Emil MIRZAYEV

GREDEG

**Présentée en vue de l'obtention
du grade de docteur en** Sciences
Économiques d'Université Côte d'Azur

Dirigée par : Zakaria BABUTSIDZE
Soutenue le : 20.10.2023

Devant le jury, composé de :

Lise ARENA, Maîtresse de conférences en
Sciences de gestion, Université Côte d'Azur
Zakaria BABUTSIDZE, Professeur, SKEMA
Business School, Université Côte d'Azur
Andrea CARUGATI, Professeur, Aarhus
University
Paolo CROSETTO, Directeur de
Recherche, INRAE-GAEL
Bernard SINCLAIR-DESGAGNE,
Professeur, SKEMA Business School,
Université Côte d'Azur
Luc WATHIEU, Professeur, Georgetown
University

Le pouvoir du contexte dans la prise de décision et les recommandations

Jury :

Directeur de these

Zakaria BABUTSIDZE, Professeur, SKEMA Business School, Université Côte d'Azur

Président du jury

Bernard SINCLAIR-DESGAGNE, Professeur, SKEMA Business School, Université Côte d'Azur

Rapporteurs

Andrea CARUGATI, Professeur, Aarhus University

Luc WATHIEU, Professeur, Georgetown University

Examineurs

Lise ARENA, Maîtresse de conférences en Sciences de gestion, Université Côte d'Azur

Paolo CROSETTO, Directeur de Recherche, INRAE-GAEL

Le pouvoir du contexte dans la prise de décision et les recommandations

Résumé

Cette dissertation comprend quatre études qui examinent les effets du contexte et appliquent ces connaissances pour améliorer les systèmes de recommandation dans les marchés en ligne. Je différencie deux types de contexte : interne et externe. Les conclusions de la dissertation indiquent que les effets du contexte interne dans les choix, qui ont été principalement étudiés et observés auparavant dans des situations expérimentales, sont également présents et détectables dans des situations concrètes. De plus, l'étude examinant le contexte externe trouve une relation positive entre les outils qui permettent aux utilisateurs de signaler leurs préférences aux systèmes de recommandation et leur adoption. Les applications empiriques des chapitres composant cette thèse reposent sur quatre ensembles de données distincts, le plus grand étant des données observationnelles provenant d'une situation concrète, et les trois autres provenant d'une situation expérimentale.

La première étude applique un modèle de décision computationnel à un ensemble de données conséquent de choix dans le monde réel, en faisant la première application de cette envergure. Les résultats indiquent que le contexte influence les choix des individus. L'étude suggère que les marchés en ligne pourraient utiliser de tels modèles pour approfondir la compréhension de la composition de l'ensemble de choix et de l'interaction entre différentes options sur les décisions des consommateurs.

La deuxième étude va au-delà des études de contexte traditionnelles en développant une méthodologie visant à séparer ses trois composantes principales, connues sous les noms d'attraction, de compromis et de similarité, les unes des autres. Cette étude contribue à comprendre l'interaction entre différents effets de contexte au sein d'un ensemble de choix et préconise le développement de conceptions de systèmes de recommandation et une compréhension plus profonde de la nature hétérogène de la dynamique du choix du consommateur. Les résultats de cette étude peuvent être utilisés pour naviguer dans le problème dit de démarrage à froid rencontré par les marchés numériques.

La troisième étude présente une approche novatrice pour aborder le problème de démarrage à froid du côté de l'utilisateur dans la conception de systèmes de recommandation. Elle s'appuie sur les résultats de l'étude précédente et applique les conclusions de la littérature sur le choix à deux étapes observé chez les individus pour générer des ensembles de considération. Ses conclusions ouvrent la voie à l'examen des effets de contexte provenant de l'extérieur des ensembles de choix, à savoir les préférences des individus et les outils qui leur permettent de signaler leurs préférences aux systèmes de recommandation. Cela est présenté comme l'un des principaux mécanismes pour créer des systèmes de recommandation plus efficaces.

La dernière étude a examiné l'effet du contrôle de l'utilisateur sur l'acceptation du système de recommandation en utilisant le Modèle d'Acceptation de la Technologie comme cadre théorique. Cette étude a trouvé que les systèmes de recommandation faciles à utiliser étaient

perçus par les utilisateurs comme plus utiles et entraînaient une plus grande intention de les utiliser. Cependant, différents mécanismes de contrôle ont eu des impacts variés sur l'expérience utilisateur.

Cette thèse démontre l'existence d'effets de contexte dans des configurations multiattributs, multidimensionnelles et développe des méthodologies pour améliorer la conception des systèmes de recommandation avec ces effets de contexte. De plus, cette étude examine le contexte externe, à savoir les outils qui permettent aux utilisateurs d'exprimer leurs préférences et comment rendre les systèmes de recommandation meilleurs grâce à eux.

Mots clés : systèmes de recommandation, prise de décision, modélisation des choix, effets de contexte

The power of context in decision-making and recommendations

Abstract

This dissertation comprises four studies investigating context effects and applying the knowledge to enhance recommender systems in online marketplaces. I differentiate between two types of context: internal and external. The findings of the dissertation indicate that the effects of the internal context in choice settings, which were primarily studied and observed before in experimental settings, are also present and detectable in field settings. Furthermore, the study investigating the external context finds a positive relationship between tools that enable users to signal their preferences to recommender systems and their adoption. The empirical applications of the chapters comprising this thesis rely on four distinct datasets, the largest being observational data from a field setting and the remaining three coming from an experimental setting.

The first study applies a computational decision-making model to a substantial dataset of real-world choice, making it the first application of this magnitude. The findings indicate that the context influences the choices of individuals. The study suggests that online marketplaces could use such models to gain further insight into how the composition of the choice-set and the interaction among different options affect consumer decisions.

The second study extends beyond traditional context studies by developing a methodology that aims to disentangle its three main components, known as attraction, compromise, and similarity, from each other. This study contributes to understanding the interaction between different context effects within a choice set and advocates the development of recommender system designs and a deeper understanding of the heterogeneous nature of the consumer choice dynamics. The results of this study can be used to navigate the so-called cold-start problem faced by digital marketplaces.

The third study presents a novel approach to addressing the user-side cold start problem in recommender system design. It builds on the results of the previous study and applies the findings of the two-stage choice literature observed in individuals to generate consideration sets. Its findings pave the way for investigating context effects arising from outside the choice sets, namely, the preferences of individuals and the tools that allow them to signal their preferences to recommender systems. This is argued to be one of the main mechanisms to create more effective recommender systems.

The final study investigated the effect of user control on recommender system acceptance using the Technology Acceptance Model as a theoretical framework. This study found that easy-to-use recommender systems were perceived by users as more useful and resulted in greater intention to use them. However, different control mechanisms had varying impacts on user experience.

This thesis demonstrates the existence of context effects in multiattribute, multidimensional settings and develops methodologies of enhancing recommender systems design with context

effects. Additionally, this study investigates the external context, namely, the tools that allow users to express their preferences and how to make recommender systems better through them.

Keywords : recommender systems, decision making, choice modelling, context effects

Acknowledgements

Embarking on this scholarly journey has been both humbling and transformative, marked by countless days of meticulous study, intense deliberation, and profound discoveries. Indeed, the journey was punctuated with hurdles and moments of self-doubt. However, every trough was succeeded by a crest, each stumble by an ascent, and every challenge served as an opportunity. The path was steep, but these uphill struggles honed my perspective, expanded my horizons, and ignited my passion for knowledge. The lessons learned, the triumphs celebrated and the sense of accomplishment far outweighed the trials faced. Reflecting on these uplifting moments, I joyfully acknowledge those who accompanied me on this journey and made it not only possible, but also immensely rewarding.

Firstly, I extend my deepest gratitude to Zakaria Babutsidze, my Ph.D. supervisor. He has been a great mentor and a source of inspiration for me. His honesty and straightforwardness challenged me to think critically and creatively, shaping me into someone better day by day. I am indebted to Francesco Castellanetta, our Ph.D. program director, for his guiding hand and unwavering faith in my capabilities. I appreciate our long and intriguing discussions with Ludovic Diabaggio that transcended academia. I am grateful to Bill Rand for his invaluable insight, feedback, and co-authorship. Daniela Iubatti, your informal talks and unwavering support have been a source of inspiration. I appreciate Diego Zunino for his honest and invaluable guidance. Bruno Cirillo deserves special mention; our albeit slightly formal discussions have significantly shaped my work. My heartfelt appreciation goes to Renata Kaminska. Her support and intuitive understanding of my thoughts were essential in my pursuits. I am grateful to Benjamin Montmartin; our separate exchanges about wine (we were in France at the end) and statistics enriched my perspectives immeasurably. To Valerio Incerti, our creative discourses, even about less creative ideas, were a constant source of inspiration. Lapo Mola's valuable feedback motivated me to pursue excellence, and I value that wholeheartedly.

Aytan, my beloved wife and best friend, your unwavering support and guidance have been my Polaris, guiding me through the darkest nights and teaching me to only settle for infinity and beyond. Ecem Delicik, my dear friend, I am grateful for the shared sto-

ries, moments, laughter, and silence that have strengthened our friendship. The moment you realized that I was speaking to you in Turkish is unforgettable (I am sure you still remember it, too). Johanna Deperi, our shared moments and long conversations were invaluable for self-reflection, joy, and sometimes “justified sadness” throughout this journey. Artyom Yepremyan, our discussions helped us conquer the PhD mountain. I will forever cherish the times when I needed a hand, and you extended yours. Manon Desjardins, your constant support despite the towering stack of papers on your desk will always be appreciated. Mehdi Ibrahim, I am grateful for our coffee talks and the irresistible desserts from Morocco. Teymur Mardaliyev, your friendship and coding prowess have been invaluable; only you could convince me to settle for JavaScript there.

Finally, no words can fully express my gratitude to my parents. Father, your inquisitive nature has been my guide, instilling in me the curiosity that led me on this scholarly path. Your enthusiasm for the unexplored has been infectious, leading me to question, explore, and discover. Mother, you have been my anchor, reminding me of my roots and my identity in the academic whirlwind. Your unwavering encouragement and steadfast belief have reinforced my sense of self, allowing me to face challenges with courage and resilience. You taught me to stay grounded and true to myself, no matter how far I travel in the pursuit of knowledge. Your lessons continue to shape my journey, helping me navigate both personal and academic hurdles with grace and fortitude.

Çox sağ olun ki, varsınız!

Contents

1 Introduction	12
2 Exploring Context Effects in Multi-Attribute, Multi-Alternative Choice Environments	15
2.1 Introduction	16
2.2 Big three context effects	17
2.3 Computational decision making models	22
2.4 Multialternative decision by sampling	24
2.4.1 Mechanisms behind MDbsS	24
2.4.2 Accounting for big three context effects	26
2.5 Empirical application	29
2.6 Results	36
2.7 Conclusion	40
3 Enhancing Choice Modeling in Multi-Attribute, Multi-Alternative Settings	42
3.1 Introduction	42
3.2 Context effect and choice modeling	44
3.3 Generalizing context effect measures	45
3.4 Empirical applications	55
3.5 Conclusion	71
4 A Context-Informed Approach to Cold-Start Problem in Recommender Systems.	73
4.1 Introduction	73
4.2 Theoretical background	74
4.3 Methodology	81
4.4 Results	84
4.5 Conclusion	85
5 User Control and Its Impact on Recommender Systems' Acceptance	88

5.1 Introduction	88
5.2 Definitions and scopes	91
5.3 Theoretical background and hypotheses development	96
5.4 Models of technology acceptance	97
5.5 Methodology	101
5.6 Task details	103
5.7 Results	110
5.8 Conclusion	118
6 Conclusion	123
7 Appendices	148
A	148
B	150
C	152
D	155
E	159
F	162

List of Figures

1	Classic illustration of context effects.	18
2	Binary choice set with two options.	19
3	Attraction effect in ternary choice.	20
4	Compromise effect in ternary choice.	21
5	Similarity effect in ternary choice in two scenarios.	22
6	MDBS' account for big three context effects	27
7	Marginal effects of MDBS output	40
8	Accounting for attraction effect	47
9	Compromise effect generalization	52
10	Amazon recommender agent for laptop selection.	93
11	Preference form in RS	95
12	Critiquing in RS	96
13	TAM model framework	101
14	Experimental interface	105
15	The expectation of the recommendations	109
16	Remuneration distribution	111
17	Observed utilities of the recommendations	112
18	Main results of TAM and effect sizes	117
19	Correspondence between Ω and N .	151

List of Tables

- 1 Descriptive statistics of variables in observational data. 31
- 2 Attribute values used in experiment 32
- 3 Optimization results for observational data. 37
- 4 Outputs of Probit model for observational data 38
- 5 Optimization results for experimental data. 39
- 6 Outputs of Probit model for experimental data 39
- 7 Descriptive statistis of context variables 56
- 8 Choice model estimation results 58
- 9 Marginal effects of choice model on observational data 60
- 10 Choice model results for the reduced dataset 63
- 11 Marginal effect of choice model for the reduced dataset 65
- 12 Descriptive statistics of variables for transport experiment data 67
- 13 Choice model results for experimental data 70
- 14 Consideration set structure as unique clusters across clustering methods. 85
- 15 Top-5 and top-10 accuracy scores across clustering methods. 86
- 16 Overview of treatment groups 113
- 17 Difference tests among treatment groups. 114
- 18 Sorting and hiding statistics per feature. 115
- 19 Factor loadings of items 116
- 20 Logistic regression results for observational data 156
- 21 Mixed Logit regression results for observational data 157
- 22 Choice model results with K-means clustering 158
- 23 Logistic regression results for experimental data 160
- 24 Fixed effect Probit results for experimental data 161
- 25 Constructs. 163

1 Introduction

We all make choices, but in the end,
our choices make us.

The main objective of this thesis is to make a contribution towards understanding the choice context and applying it to improve recommender systems. The thesis makes an effort to identify context effects arising from the choice set and to develop a methodology to implement this information in the design of recommender systems. Furthermore, it analyzes the effectiveness of user control mechanisms, which allow individuals to inform the system about their preferences and also amend them.

Recommender systems have become a crucial ally in the vast landscape of the digital world. They help individuals to navigate through the plethora of alternatives and find what they are looking for. They achieve this by using sophisticated algorithms that analyze the wide space of items, users, and their interactions with each other. System designers and businesses thrive to maximize the accuracy of recommendations, which means they want to maximize the consumption of recommendations. However, recent research streams show that accuracy is not the only way to make recommender systems effective. Other factors such as diversity and serendipity are also important (Kaminskas and Bridge, 2016). Moreover, the black-box nature of the algorithms does not allow us to understand the reasons behind a particular user’s choice behavior (Kotkov et al., 2016; SAMIH et al., 2021). Therefore, more and more interest is directed towards understanding and utilizing the context around the particular choice (Adomavicius and Tuzhilin, 2005).

The choice context has long received the attention of scholars in many fields, including marketing, psychology, and economics. Because it was studied in many domains, scholars refer to it differently. This thesis concentrates on and uses two of them. The first one posits that the choice context ¹ is “the availability and nature of the choice alternatives” (Tversky, 1972; Huber et al., 1982; Simonson, 1989). Previous research has demonstrated the existence of context effects in various settings (Herne, 1997; Soltani et al., 2012; Evangelidis et al., 2018; Wu and Cosguner, 2020). The second definition of context arises

¹For the remainder of the dissertation, I will follow the previous literature (Trueblood et al., 2013) and refer to this definition of the choice context as context effects.

much later with the proliferation of recommender systems. Then, context is also referred to as “the time and content of the choice, the location or sociodemographic characteristics of the decision maker...” (Adomavicius and Tuzhilin, 2011)²

However, modeling for context effects mathematically was a challenge because the existing models suffered from the independence of irrelevant alternative criterion (Luce, 1959), which meant that these models treated each alternative in isolation. This has shifted the scholar’s interest towards computational models (Usher and McClelland, 2001; Roe et al., 2001; Trueblood et al., 2014; Noguchi and Stewart, 2018). Yet, none of these models has been applied to real-world choice data, and as a result, their applicability to field data remained an open question. I address this gap in the first study where I apply a computational model to a field data of high heterogeneity among dimensions. Building on my results in my second study, I have delved deeper into choice modeling and developed a methodology to account for the three main context effects studied in the literature (Trueblood et al., 2013).

When new users or items are introduced to recommender systems, they fail to function as intended because of the lack of information their algorithms need. In such cases, even when they do provide recommendations, the recommendations are far from personalized (Lika et al., 2014). It is considered a key challenge in recommender system design (Park and Chu, 2009). The previous literature has applied various methods to overcome this problem, including asking users to rate some items, share their preferences among others (Guy et al., 2009; Aharon et al., 2013; Bykau et al., 2013; Saveski and Mantrach, 2014). However, the limitations of these approaches are that they ignore the context effects and concentrate on scenarios where information scarcity is temporary. Using the findings from my second study and combining it with decision-making and choice modeling literature, I propose an innovative solution to address continuous information scarcity issue in recommender system design by utilizing contextual information of the menu to generate relevant choice sets through a two-step choice modeling method.

Research agrees on the importance of metrics outside the boundaries of accuracy for recommender systems (Kaminskas and Bridge, 2016). It is in the best interest of online

²To distinguish between these two definitions I will refer to this definition as external context throughout this dissertation.

marketplaces to provide users with recommendations that are not only in the category “exactly what I want”, but also “I never thought I would have liked this” (Kotkov et al., 2016). This can not only boost sales (Song et al., 2019), but also increase user satisfaction (Knijnenburg et al., 2012). However, it was observed that individual preferences are not stable and tend to change and system designers have proposed various tools to allow users to signal their preference shifts (Bostandjiev et al., 2012; Hijikata et al., 2012). However, to better understand, it is necessary to study those tools in isolation from the context arising from the choice set and concentrate purely on external context that arises from the user side, e.g., their preferences (Adomavicius and Tuzhilin, 2011). In my last study, I address this gap by conducting an online experiment and applying Technology Acceptance Model (Davis, 1985) to measure users’ acceptance of such tools.

All in all, the four studies aim to contribute to our understanding of the context and provide applications of the design of recommender systems using this information.

2 Exploring Context Effects in Multi-Attribute, Multi-Alternative Choice Environments

Abstract

Previous computational decision making models that were developed to account for context effects have only been studied with an experimental data where only one effect was produced at a time. Using data coming from strictly controlled experimental environments hinders the understanding of context effects that occur in real-world choice scenarios where items have multiple dimensions and choice sets have dozens of alternatives. In this chapter, I apply a computational model that accounts for context effects to an observational data which was not done before. The data comes from an air travel industry and is ideal to study context effects in multiattribute, multialternative choice environments. I first find optimal parameters for the computational model using the differential evolution algorithm. Then, I complement a traditional choice model with its outputs and assess the significance of its contribution. This chapter contributes to context effect and decision-making literature by providing further insights on behavior of computational decision-making models in real-world choice data.

2.1 Introduction

Context effects have been extensively studied and demonstrated in various domains, from psychology to marketing (Herne, 1997; Soltani et al., 2012; Trueblood et al., 2013; Frederick et al., 2014; Evangelidis et al., 2018; Wu and Cosguner, 2020). Some recent studies also concluded that multiple context effects may occur at the same time (Berkowitsch et al., 2014; Noguchi and Stewart, 2014). However, recent studies have also discovered boundary conditions for these effects (Liew et al., 2016; Spektor et al., 2018, 2019). Familiarity with the choice domain was found to reduce the context effects experienced by individuals (Kim and Hasher, 2005; Sheng et al., 2005). It was also found that some conditions may force these effects to completely reverse (Cataldo and Cohen, 2019). The findings described above make it necessary to call for a model which could explain these effects.

Logit and Probit models have been traditionally used in choice settings (Gensch and Recker, 1979; Kim et al., 2017). However, those models cannot account for context effects because they only account for the attributes of the focal option, not taking into account the attributes of the other options in the choice set. Tversky has proposed a model of elimination by aspects that could explain the similarity effect (1972). The foundation of the model is attention switching of individuals between alternatives and attributes and their comparisons. Once attention is received, the attribute value of a given alternative is compared to a predetermined threshold value by the individual, and if the threshold does not meet, that alternative is eliminated from the decision. The step is then continued with another attribute until the final decision is made. Another model proposed by Tversky and Simonson could explain the effect of compromise (1993). The theoretical foundations of this model posited that alternatives are compared based on a weighted sum of attribute values and a local context comprising binary comparisons among alternatives. However, these two models were unable to successfully account for all three effects. Despite their drawbacks, the sequential decision making and attention-based mechanisms in these models laid the foundations of many upcoming computational choice models (Bhatia, 2013).

In the last three decades, researchers have developed many computational models which account for context effects. Multialternative Decision Field Theory (MDFT) (Roe

et al., 2001), Leaky competing accumulator (Usher and McClelland, 2001), Multiattribute linear ballistic accumulator (Trueblood et al., 2014), Multialternative Decision by Sampling (Noguchi and Stewart, 2018) are part of them. Some of these models have been extensively tested and studied, while others are relatively new and therefore have not received much attention from scholars (Trueblood et al., 2013). However, these studies have been performed with experimental data (Trueblood et al., 2014; Berkowitsch et al., 2014; Evans et al., 2019; Busemeyer et al., 2019). Research has proven that the behavior of individuals in the laboratory choice environment is different from real-world choice environment (Hogarth and Kunreuther, 1989). Hence, the applicability of such models to field data is an uninvestigated avenue because no previous study has been done where a computational model was applied to real-world observational data. The plan is to address this gap by applying the Multialternative Decision by Sampling (MDbS) model proposed by Noguchi and Stewart (2018) to a field data from the airfare booking domain. This would allow me to assess the applicability level of this model complex field-data.

Also, applying MDbS to observational data would allow one to statistically assess the significance of the contribution of this model's ability to account for the context effect. Instead of testing this model against established choice models, I will attempt to complement them with MDbS. To do this, the random effect Probit model will be used as a variation of the Probit family and will be augmented with MDbS output. The Probit model is chosen as it does not explicitly assume IIA unlike the family of Logit models. To validate my results further, I will apply the same methodology and analysis to an experimental data.

The reasons for choosing MDbS are twofold. First, it is relatively new when compared to other models, hence it has not been further investigated before. Second, it is more robust and can account for a wider range of context effects (than other models) known to the literature (Noguchi and Stewart, 2018).

2.2 Big three context effects

We make choices all the time. Imagine the time you went to see a movie you had been waiting for a while and decided to grab some popcorn before entering. You may have seen something like that in the figure [1](#), although prices may be higher these days. You

are puzzled at first, but reminded that the movie is about to start, so you better hurry up and choose one.

The small one feels not enough for a 90 minute marathon. Then, there is a middle one which seems like an okay option at first. When you notice that big box, you immediately forget about the small one you saw moments ago. You start looking at sizes and prices of middle and large boxes and think: Well, that is easy. Big box seems like the way to go here, considering their prices are almost equal.

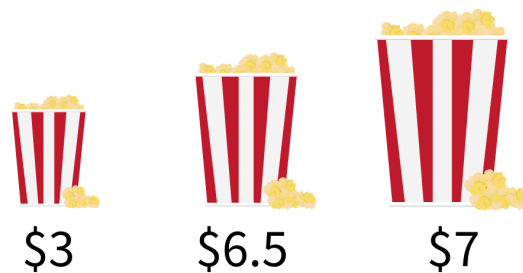


Figure 1: Classic illustration of context effects.

If this situation is familiar to you, then you have experienced the so-called context effect which can be understood as “the composition and the nature of the choice set, availability of various options in it” (Tversky, 1972; Huber and Puto, 1983). For a long time, our understanding of the choice did not extend over the borders of two related principles. The first one is independence of irrelevant alternatives (IIA), which states that when having a choice between two two options A and B if a person prefers A for example, regardless of adding a third option C to this choice set, that person’s preference must be unaltered (Luce, 1959). The second is the regularity principle, which states that the probability of choice of option A cannot increase by the introduction of option C (Luce, 1959).

However, research has concluded that the context of the choice set and the options in it have a substantial effect on how people make decisions. This effect has been extensively studied over the past five decades by many economists, marketing scholars, and psychologists³ (Kahneman and Tversky, 1979b; Simonson, 1989; Tversky and Simonson, 1993; Lichtenstein and Slovic, 2006; Dowling et al., 2020). Most of the research has focused on

³See Dowling et al. (2020) and Lichtenstein (2006) for a more comprehensive review.

three context effects, also referred to as “big three”: attraction, compromise and similarity (Howes et al., 2016).

To better understand these three context effects, let us think of a hypothetical choice set where options differ along two dimensions: Dimensions 1 and 2. We first start with a set of options consisting of two options: A that has the values 20 and 80; B that has the values 80 and 20 that figure 2 represents.

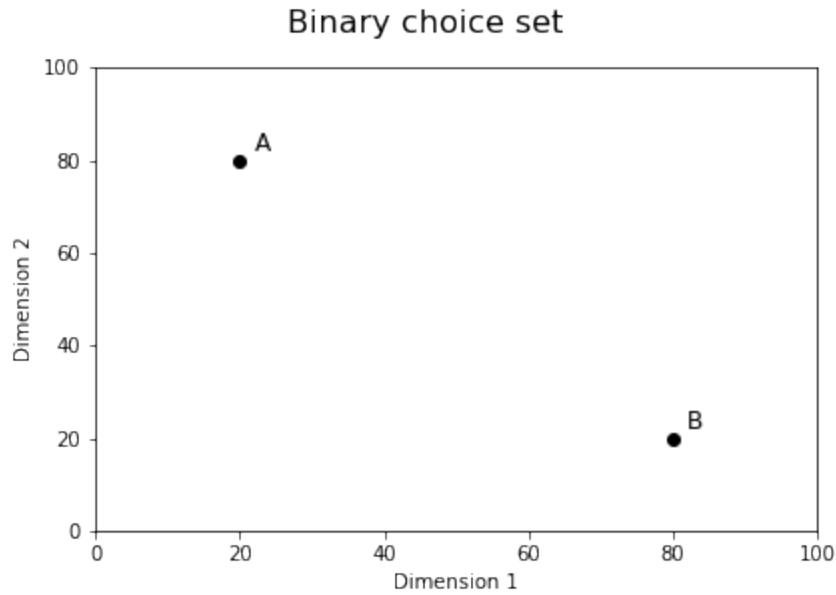


Figure 2: Binary choice set with two options.

People who prefer Dimension 1 would choose B whereas A will be chosen by people who prefer Dimension 2. This can be described as equation 1 below.

$$P(A|A, B) = P(B|A, B) \quad (1)$$

Where $P(A|A, B)$ corresponds to the probability of choosing A given the choice set A, B . Same goes for $P(B|A, B)$.

Attraction effect

Now, let us add a third option to this choice set, the option D_A to create one variation, and D_B to create the second variation of the ternary choice set. Both added options have lower values in both dimensions compared to A and B , respectively. Huber created this type of scenario and has found what he has called the “attraction effect” (1982). The attraction effect, which is also known in the literature as the asymmetric dominance

effect, is a consistent violation of the regularity principle mentioned earlier. He suggested that when having a choice set consisting of options A and B the relative probability of choosing option A can be increased if a third option with characteristics of D_A is added to the same set of choices (1982).

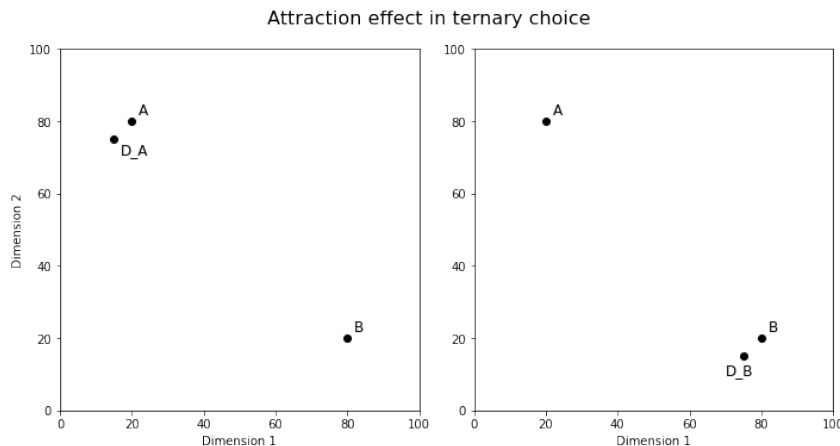


Figure 3: Attraction effect in ternary choice.

Figure 3 shows these options and their respective values in each dimension. It can be seen that option A and B are located at two different ends of the choice space. Option D_A is inferior to option A in both dimensions and D_B is inferior to B . With the attraction effect in place, equation 1 will change into 2.

$$\frac{P(A|A, B, D_A)}{P(B|A, B, D_A)} > \frac{P(A|A, B)}{P(B|A, B)} \quad \& \quad \frac{P(B|A, B, D_B)}{P(A|A, B, D_B)} > \frac{P(B|A, B)}{P(A|A, B)} \quad (2)$$

Huber noted that although other explanations are still possible, the addition of D_A to the choice set would shift the preferences of people towards dimension 2 because this is where option A appears advantageous (Huber et al. (1982); Bhatia (2013)). However, this claim has not received unanimous support in future studies in which preference changes have been observed (Wedell, 1991).

Compromise effect

When the third option we add is option C instead of D , the preference shift happens differently. C is virtually a middle option between A and B and therefore has a value of 50 in each dimension. In this case, the probabilities of choosing A and B will both decrease in favor of C , resulting in 3:

$$P(A|A, B, C) < P(A|A, B) \quad \& \quad P(B|A, B, C) < P(B|A, B) \quad (3)$$

Simonson was the first to describe such an effect (1989). He associated this with a difficulty to select: When people are not sure which attribute is important, they will find a justification to favor a compromise (Simonson, 1989). This argument can explain the reason why an individual may drift towards a middle choice in a three-option choice set. Such a compromise emerges as an important factor, acting as a tie-breaker when the decision maker is unsure between the initial two options.

It is worth noting that it is also possible to “target” a particular option from the binary choice set when adding a third option to create a compromise effect. One can add a target C which makes A a compromise option. In this case, the probability of choosing A among this triple will increase, as it is considered a compromise between the remaining two options.

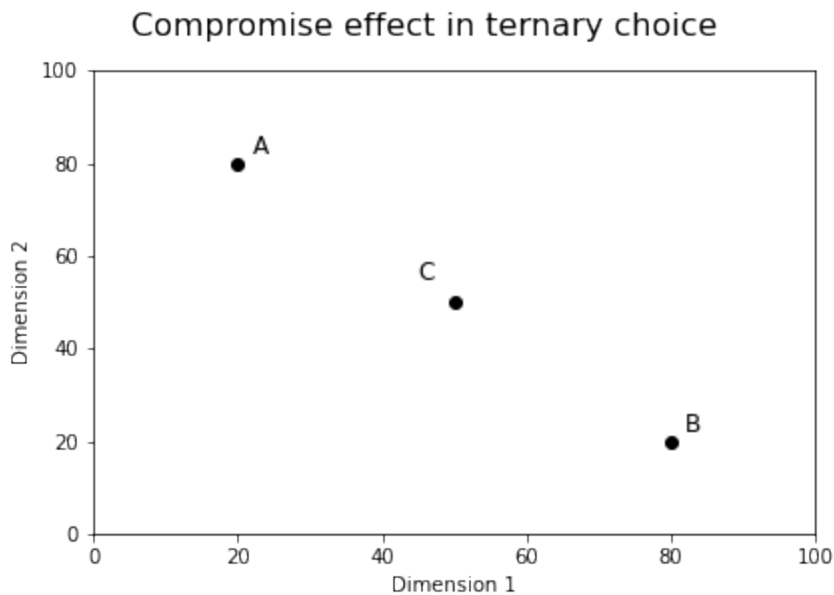


Figure 4: Compromise effect in ternary choice.

Similarity effect

Although Becker et al. (1964) have mentioned it before, the first study of the similarity effect is known to be that of Tversky (Tversky, 1972). He noted that when faced with a binary choice set consisting of A and B individuals will gravitate towards A more

than when faced with a ternary choice set consisting of A, B, S_A depicted in figure 5. He explained it by proposing the elimination by aspects theory, which states that one attribute will be chosen as elimination criteria, and all options that do not meet that criteria will be eliminated (Tversky, 1972). Therefore, in the set of choices A, B, S_A if an individual selects dimension 2 as the elimination criteria, both A and S_A will be eliminated, leaving B as a choice. In contrast, if the decision maker prefers dimension 1 more, then option B will be eliminated, leaving both A and S_A to share the “victory”, hence resulting in equation 4.

$$P(A|A, B, S_A) < P(A|A, B) \quad \& \quad P(B|A, B, S_B) < P(B|A, B) \quad (4)$$

The similarity effect in the choice set A, B, S_A will follow a similar route. Figure 5 depicts both sets of choices, where an option similar to A and B was introduced separately.

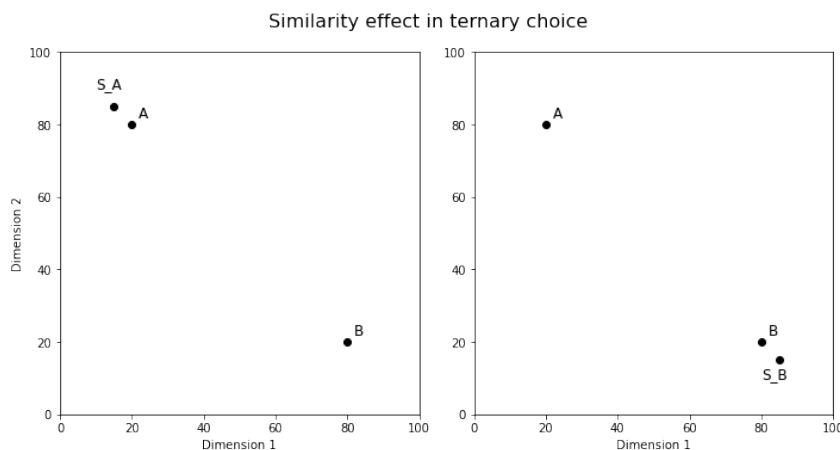


Figure 5: Similarity effect in ternary choice in two scenarios.

2.3 Computational decision making models

MDbS belongs to the attention-based choice models. These models are type of decision-making models which take into account the attention allocation mechanisms when making decisions. They generally assume that people allocate attention to different attributes and those who receive more attention have more impact on decision making (Gabaix and Laibson, 2000). Before commencing with its underlying mechanisms and assumptions, it is beneficial to discuss the other two models that MDbS influenced by and have been

studied extensively. After briefly discussing those models, I will continue with MDbS, its main assumptions, mechanisms, and how it accounts for context effects.

Multialternative Decision Field Theory

The very first computational model which could account for all three context effects was Multialternative Decision Field Theory (MDFT) developed by Roe et al. (2001) as an extension of Decision Field Theory (Busemeyer and Townsend, 1993). It is a dynamic model of decision making that accommodates multialternative preferential choice situations, which was not possible with Decision Field Theory (Hotelling and Rieskamp, 2019). MDFT assumes that decision making can be explained in three general mechanisms. First, attention allocation posits that attention switches over time between attributes stochastically. Second, the evaluation mechanism posits that the attribute value of the given option is compared with the average attribute values of other options, which makes sure that each option in the choice set participates in comparison. Third, the evidence accumulation mechanism, which is based on the results of the evaluations, gathers evidence in favor of the alternatives compared. As soon as the evidence gathered hits the externally set relative threshold, the choice is concluded (Busemeyer and Diederich, 2002). This means that as soon as the difference between the highest and the second highest evidence values is larger than the relative threshold, a choice is made. If this threshold is not met, the decision continues until the pre-set time limit is reached.

MDFT has been confirmed to account for similarity, compromise, and attraction effects in multialternative choice scenarios (Roe et al., 2001). It has previously been tested against such random utility models of choice as Logit and Probit and has been concluded to be a better fit to empirical data (Berkowitsch et al., 2014). MDFT has also been adapted to account for preference changes (Mohr et al., 2017) and decision making under time restrictions (Diederich, 2003).

Multiattribute linear ballistic accumulator

The multiattribute linear ballistic accumulator (MLBA) is another attention-based decision-making model first proposed by Trueblood et al. (2014). Similar to MDFT, it is also a dynamic model, which can be explained in three general mechanisms: attention allocation, evaluation of alternatives, and evidence accumulation. However, the two models have key differences. Firstly, MDFT emulates the search process of elimination by aspects

proposed by Tversky (1972) assuming that decision makers compare alternatives with one another over time. In contrast, MLBA assumes that individuals make comparisons and accumulate evidence of all alternatives independently from one another at the same time and then accumulate evidence (Trueblood et al., 2015). Furthermore, MDFT assumes that individuals have limited cognitive capacity to process information when comparing items together, in contrast to MLBA, which considers individuals with unlimited cognitive capacity. Moreover, unlike MDFT, where decision is made based on relative threshold, in MLBA decision is based on absolute threshold, i.e. as soon as one alternative’s evidence reaches the threshold, a decision is made in favor of that alternative. Another difference between these two models is the context effects for which they account. Although MDFT does account for attraction, compromise, and similarity effects (Hotelling and Rieskamp, 2019), MLBA additionally accounts for preference reversals arising from context (Trueblood et al., 2015).

2.4 Multialternative decision by sampling

MDbS has its origins in the theory of decision-by-sampling, which assumes that individual preferences arise from binary, ordinal comparisons of alternatives on given attribute values with reference values from the memory (Stewart et al., 2006). Unlike it, MDbS assumes that the information required for comparison also comes from the choice environment itself (Noguchi and Stewart, 2018). As in the other two previous models discussed above, its mechanisms can be explained using three stages. The next section will discuss this in detail.

2.4.1 Mechanisms behind MDbS

ATTENTION ALLOCATION

According to MDbS, when comparing two tickets between Paris and New York, for example, the price of a ticket would be compared to the prices of other tickets in the choice set and also to the ones and which an individual has previously seen but are not in the current choice set. Comparisons are ordinal, meaning that evidence accumulated toward the “winner” at a rate of one irregardless of how large the difference was.

Previously, it was concluded that people tend to compare alternatives that are similar

to each other more than dissimilar ones (Noguchi and Stewart, 2014). Similarity-based attention is one of the main assumptions of MDbS. To better understand this, let m_{ij} and m_{kj} be two attribute values with $i \neq k \in \{1, \dots, n_a\}$ and $j \in \{1, \dots, n_d\}$. MDbS defines the similarity of m_{ij} to m_{kj} as

$$s_{ij,kj} = \exp\left(-\alpha \left| \frac{m_{ij} - m_{kj}}{m_{kj}} \right| \right), \quad (5)$$

with similarity parameter α . Also, generally $s_{ij,kj} \neq s_{kj,ij}$. Consider also

$$s_{ij} = \sum_{k \neq l \in \{1, \dots, n_a\}} s_{ij,kj}, \quad (6)$$

which is the sum of all similarities for attribute m_{ij} to other attributes on the same dimension. Consequently, by dividing this value by the sum of similarities in all other attributes across all dimensions, one can calculate the probability that m_{ij} will be selected for comparison, which will be

$$p_{ij} = \frac{s_{ij}}{\sum_{l \in \{1, \dots, n_a\}} \sum_{m \in \{1, \dots, n_d\}} s_{lm}}. \quad (7)$$

EVALUATION OF ALTERNATIVES

When evaluating alternatives with each other based on pairwise comparisons, MDbS defines the probability of winning a comparison as

$$P(m_{ij} \text{ is favored over } m_{kj}) = \begin{cases} F(\beta_0(|\frac{m_{ij}-m_{kj}}{m_{kj}}| - \beta_1)) & \text{if } A_i > X_i \\ 0 & \text{otherwise} \end{cases}, \quad (8)$$

where F is a logistic sigmoid function and β_0 and β_1 correspond to the advantage value and the probability that this particular advantage value will be enough to be preferred. For example, consider the case where $\beta_0 = 0.1$ and $\beta_1 = 50$. This would mean that

the advantage of 10% would be preferred with the probability of 50%. Consequently, if the difference is 20%, then it will be preferred with the probability of 99%. The logistic function brings the notion of “soft” comparison instead of “hard” comparison, in which case small differences would be ignored, while large differences would be extremely preferred (Noguchi and Stewart, 2018).

EVIDENCE ACCUMULATION

As mentioned above, in MDbS the accumulation of evidence occurs at a rate of one. For each alternative and for each comparison, in case of winning that comparison, one evidence point is counted towards that alternative. Hence, the probability that evidence will increase by one point will be defined as

$$p_i = \sum_{j \in \{1, \dots, n_d\}} p_{ij} \cdot P(m_{ij} \text{ wins a comparison}). \quad (9)$$

In order to make a choice, MDbS sets a relative stopping rule θ following the study of Teodorescu (2013) which states that when deciding between more than two alternatives, the decision is made when the difference between the highest and the second best evidence is larger than the threshold, or the difference between the maximum and mean evidence becomes larger than the threshold. For computational feasibility, MDbS assumes $\theta = 0.1$, which means that a decision is made when the difference between the maximum and mean average evidence reaches 0.1. Other parameters given externally are α, β_0, β_1 . As a last step, evidence for each alternative is divided by the sum of evidence for the entire choice set to convert them to choice probabilities.

After discussing the main mechanisms behind MDbS, the discussion about how MDbS accounts for attraction, compromise, and similarity effects becomes necessary. The next subsection will shed some light on this matter.

2.4.2 Accounting for big three context effects

To effectively illustrate the functioning of MDbS, employing an example choice set that encapsulates the context effects discussed can be beneficial. Therefore, the example dataset depicted in figure 6 will be used. Although there are five alternatives in the figure 6, only

three of them will be discussed at a time.

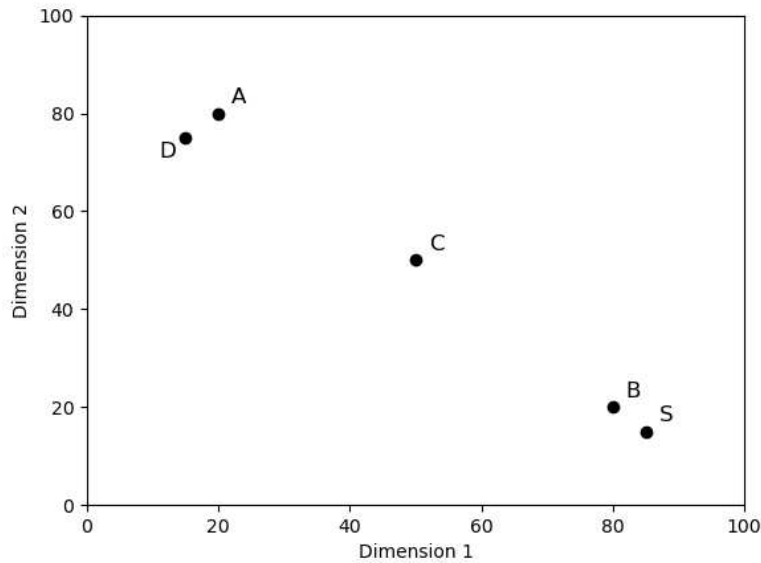


Figure 6: Example choice set to explain MDbS’ account for big three context effects. A and B are considered original two alternatives (binary choice set). D is dominated by A on both dimensions. C acts as compromise between the original two and S is a similar option to B . Although these effects would be present in different variations of the choice set (for example one can make A as compromise option), for the simplicity I will concentrate on this example.

Accounting for attraction effect

When adding the option D which is dominated by A in both dimensions to the binary choice set $A B$ (hereafter the binary choice set will be used instead of A and B) one creates an attraction effect (Huber et al., 1982; Huber and Puto, 1983). Huber et al. (1982) explained the attraction effect through weight shifts for individuals. The addition of D would cause people to weigh dimension 2 more. Therefore, A and D will have higher “interest” among the people, and this is where A wins over D .

However, MDbS takes a different approach. Adding D will increase the probability of comparison of option A , due to its similarity to D (recall equation 5). As a consequence, the probability of A winning a comparison will be greater than the probability of B winning a comparison, because A dominates D in both dimensions, while B is only better in one and A and D will be selected more often to be compared with each other. As a result, A will have the highest expected evidence accumulated, B will be the runner-up, and D will have the lowest evidence accumulated towards it. The value of α will determine

how much attribute similarity is translated into being selected for comparison.

Accounting for compromise effect

In the scenario where option C is added to the binary choice set, the compromise effect arises (Simonson, 1989). This makes C more likely to be chosen because people become uncertain about the importance of attributes and, therefore, experience a choice difficulty. This results in the choice of C as it is easier to justify (Simonson, 1989).

MDbS approach differs here as well. Recall that the probability of accumulating evidence towards an option is a product of its probability to be selected for comparison and its probability to win that comparison, as described in equation 9. C is more similar to A and B than A is similar to B (and vice versa). This will increase the probability that C will be chosen as a comparison. Although C will not win every comparison, the fact that it will be chosen more as a comparison will increase its probability of accumulating evidence.

Accounting for similarity effect

When Tversky (1972) explained the similarity effect in a choice set consisting of A , B and S , he explained it via his famous elimination by aspects theory. When B and S are similar to each other, they will be eliminated together or stay together. Hence, having S in the set of options will “steal” the probability of choice from B .

In MDbS this is explained by β_0 and β_1 from equation 8. Recall that the reason of having the sigmoid function with arguments β_0 and β_1 is to make sure that small differences would be relatively ignored. This is in line with previous literature that states that people tend to ignore small differences between alternatives when making a decision (Kalwani and Yim, 1992). Hence, the small differences between B and S would be ignored, which would translate into a decreased probability that any of them would win over the other when compared. This will indirectly increase the accumulation of evidence for A , resulting in “shared” evidence between B and S .

Not only does MDbS account for the three main context effects, it also successfully accounts for other known effects to decision-making literature, such as the attribute spacing effect (Cooke and Mellers, 1998); centrality effect (Brown and Matthews, 2011); background contrast effect (Tversky and Simonson, 1993); endowment effect (Knetsch, 1989)

and others⁴. Overall, the authors claim that MDbS can theoretically account for up to 25 variations in the context effect.

Considering its ability to account for wider range of context effects, MDbS offers a novel and more insightful path for studying multidimensional and multiattribute choice data. It also offers a fine trade-off between the complexity of the model, namely its dynamic attributes, which makes it more practical analytically.

After delving into the theoretical mechanisms of MDbS and its handling of the three main context effects, the focus now shifts to the practical component of the research: the methodology. The following section provides a deeper exploration into the specifics of the research process. It offers a brief overview of the dataset, followed by an in-depth explanation of the steps involved in the analysis.

2.5 Empirical application

Observational Data

The observational dataset is created by the merger of two sources. The first dataset constitutes a list of all air travel reservations made in Europe on European routes between December 2013 and June 2014, extracted from the MIDT database (Marketing Information Data Tapes). Besides all of the booking details (e.g., number of passengers, price), it also contains the timestamp of booking and the identity of the booking office (all offline and online outlets have unique identifiers). The second source of data contains information on all air travel searches performed on one of the most comprehensive air travel booking services operated by Amadeus S.A.S. This dataset also contains trip specifics, as well as the identifier of the office where the search was performed. Most importantly, the latter dataset contains information on all possible alternatives that could have been presented to the traveler at the time of search, but does not contain information on which of the options (if any) has the traveler chosen. The matching of these two datasets across office identifier, search / booking time, origin and destination of the trip, travel dates, and number of passengers results in a merged dataset that allows us to identify the itineraries chosen within the options offered at the search⁵. An important limitation of the data is

⁴For the full list of the effects MDbS can account for please refer to Noguchi (2018).

⁵Office ID, trip origin and destination, trip dates and number of passengers are matched

that there is no way of ensuring that the consumer has actually seen the exhaustive list of alternatives available to him/her at the time of booking. However, dataset consists of all the options that they could have seen. Although this is a drawback for a researcher, this is a standard experience for the practitioner (e.g., recommender system designer). Practitioners designing recommender systems need to create algorithms based on the set of existing alternatives without much visibility on the subset of options a particular user will be interested in, or will eventually see.

The matched dataset (previously used by Mottini and Acuna-Agost (2017)) consists of 13000 choice sessions with around 1 million choice alternatives in total. Every alternative is a round-trip flight and has a number of attributes including ticket price, date and times of all inbound and outbound flights, number of flights in the itinerary, number of airlines, days before booking, and a few more, less important attributes.

Menus (i.e., choice sets) with one single available alternative do not allow the consumer to make choices and are therefore discarded. Data on choices contain at most 100 alternatives for each choice session, even if more choices potentially existed. As a result, our data are truncated from the right. This creates a large number of menus, including exactly 100 alternatives, some of which may be incomplete. To deal with this oddity, we simply confine our research to menus having between 2 and 99 alternatives (excluding those with one option, since there is no choice involved). In the end, we are left with a dataset with 6,297 choice sessions with 368,723 alternatives in total.

These are the attributes that are designated as vertical in the choice process. For the purposes of this dissertation, it is assumed that consumers prefer lower values for each of them (e.g., all consumers prefer lower prices, shorter trips, fewer layovers, and not having to change airlines too frequently). Apart from vertical attributes, there are also three attributes that do not vary across alternatives within each menu. These are the number of days between when the choice was made and the start of the trip, whether the trip is domestic or international and whether it is intercontinental. In addition to vertical attributes, the time interval between the reservation and the preceding search is minimized. If, given exactly matched attributes, the booking was not performed within 24 hours after a given search, the search is declared unmatched. If, given exactly matched attributes, no search is found during the 24 hours preceding a given booking, the booking is declared unmatched. Unmatched searches and bookings are removed from the analysis

Variable	Count	Mean	St.Dev.	Min	Max
Price (in EUR)	368,723	647.12	1,105.120	59.55	16,997
Trip duration (in minutes)	368,723	518.98	555.04	70	2,715
Number of flights	368,723	2.94	0.95	2	6
Number of airlines	368,723	1.25	0.45	1	5
Menu size	368,723	58.077	30.267	2	99
Days before departure	368,723	32.36	38.03	0	340
Domestic travel	368,723	0.49	0.49	0	1
Intercontinental travel	368,723	0.06	0.23	0	1

Table 1: Descriptive statistics of variables in observational data.

attributes, the data also contains two sets of horizontal attributes, the departure times and the dates of outbound and inbound flights. These attributes are treated as horizontal, as there is no clear way of defining consumer preferences over them. To eliminate potential scale effects, z-score normalization on vertical attributes was performed as $Z = \frac{x-\bar{x}}{\sigma}$ where \bar{x} is the mean and σ is the standard deviation of the variable x .

Due to the MDbS nature of comparing dimensions with one another, for the purposes of this study, I cannot use variables that do not differ between menus. Additionally, I cannot utilize horizontal attributes in my analysis because they do not follow the standard “the greater the better mathematical approach”. However, these variables are used later in this dissertation in chapters [3](#) and [4](#). As a result, I am bound to utilize only four vertical attributes in my analysis. Table [1](#) provides descriptive information about the variables. I have also multiplied the four vertical variables by 1 to convert them to a negative scale due to the MDbS nature of comparing absolute values and the assumption that consumers prefer lower values along vertical dimensions.

Experimental data

In the course of this study, I also introduce an additional dataset sourced from a controlled experiment conducted by Noguchi ([2018](#)), distinct from the primary observational data. Although not as diverse in terms of alternatives, dimensions, and choice sets, these experimental data carry significant value, not as a principal analytic focus, but rather as a means to corroborate my main findings. I will apply the same analytical techniques employed in the observational analysis and utilize experimental dataset as a robustness check to verify the validity of my results. Henceforth, the role of these data is primarily confirmatory.

Product	Dimension	Alternative A	Alternative B
Mouthwash	Breath	4.5 hours	7.2 hours
	Germs killed	77%	56%
Exercise class	Fee	\$9.49	\$6.49
	Calories	356 kcal	259 kcal
Box of chocolate	Amount	26 oz	33 oz
	Variety	9	5
GPS	Update	3.04 Hz	5.62 Hz
	Accuracy	4.97 m	7.83 m
Mobile battery	Price	\$19.93	\$13.49
	Talk time	14.55 hours	9.25 hours
Light bulb	Life	1309 hours	1923 hours
	Price	\$1.35	\$2.50
Air purifier	Noise	64.7 dB	39.3 dB
	Efficiency	325 cfm	203 cfm
Strawberry	Quantity	407 g	452 g
	Price	\$2.58	\$2.85

Table 2: Attribute values used in the experiment. Sourced from Noguchi (2018).

This dataset comes from an experiment conducted by Noguchi (2018) in which 503 participants, aged 18 to 75 years, participated in Amazon Mechanical Turk, resulting in a total of 5295 observations. Participants faced eight randomly sampled decision scenarios with descriptions consisting of two and three alternative sets, each ternary choice set containing only one of the attraction, compromise, similarity effects. For ternary choice sets, to create a context effect, one alternative was randomly selected as focus and third alternative was generated following three scenarios: a) for attraction effect scenario, third options' both dimensions were reduced by 25% of the difference between the remaining two options' dimensions; b) for compromise scenario, the third option was generated in a way that it would make the randomly chosen target a compromise; c) in similarity scenario 2% of the difference was added to one dimension while 2% was subtracted from another dimension for the third option. The dimensions of the alternatives are described in the table 2.

Parameter optimization

Recall the three dynamic parameters for MDbS, α, β_0, β_1 , which were discussed in section 2.4.1. They allow MDbS to take into account various context effects. Noguchi (2018) demonstrates MDbS performance using a fixed set of parameters throughout the article using: $\alpha = 3, \beta_0 = 0.1, \beta_1 = 50$.

Those parameters are essential controls of the behavior of the model and they create the underpinnings of the choice set, impacting the generated choice probabilities. Therefore, it is fundamental to identify the optimal parameters that will fit the observed data. While identifying the optimal parameters could ideally be purely theory-driven, in reality the theoretical guidance will often fall short. This will leave a plethora of potential parameter values. Hence, this requires a systematic search method to explore the parameter space and identify the optimal parameters that would fit the data the best.

Parameter space definition

Before continuing further with the method, one must first define the parameter space over which the search process will commence. Recall that there are three parameters to be optimized, which were α, β_0, β_1 . Two of them, β_0 and β_1 , have theoretical boundaries.

The effect of the attribute range was first investigated by Mellers (1994), who described the tendency of people to scale the perceived attractiveness of an alternative in a given attribute using the entire range of that attribute. Therefore, β_0 in MDbS represents the fraction of the difference between attributes compared to the entire attribute range, which is bounded between 0 and 1. On the other hand, β_1 represents the percentage of preference of that difference for an individual; hence, it is also bounded with values between 0 and 100. I have created 99 β_0 values evenly spaced between 0 and 1 and 99 inclusive β_1 values evenly spaced between 0 and 100.

The parameter α on the other hand, does not have a theoretical upper bound. However, since it is used to determine which alternatives to compare with each other, it must be greater than 0 because otherwise no alternative will be selected for comparison. I have randomly generated 4,000 samples where the alpha ranged between 0.1 and 10. My observations have demonstrated that the performance of MDbS significantly deteriorates when $\alpha \geq 5$. Hence, to balance the need for a flexible model with the requirement of stable performance, an upper bound of 5 has been set for α . I have created 49 α values between 0.1 and 5.

In total, the full parameter space has been created with combinations of the three parameter values reaching 480,249 triples.

Optimization method

Parameter optimization is a task of high importance in many scientific and engineering applications, where the goal is to find the optimal values of a set of parameters that best fit a given model or system. There are various methods available for parameter optimization, ranging from differential equation-based methods to brute force and other optimization algorithms. I have chosen the differential evolution algorithm proposed by Storn (1997) for this purpose. It has several advantages over other algorithms. Firstly, it can be easily implemented. Secondly, it is ideal when the parameter space is large (Lin et al., 2019). Thirdly, it is especially suitable for complex and non-linear functions (Omran et al., 2009).

The way differential evolution works resembles other genetic algorithms. First, it creates an initial population P with a size of n within a given parameter space S and assesses its fitness using the evaluation metric F . Then, it randomly selects three members of P and creates a new member. If it is better than a randomly selected one within this triple, it replaces it. This process continues until termination criterion is met, which is either: a) F has reached its global minimum, b) the number of iterations has reached the threshold, or c) F has not improved considerably within the predefined number of iterations. The pseudocode below describes its workflow:

Algorithm 1 Simplified Differential Evolution.

- 1: Initialize population of P from the parameter space S
 - 2: **while** not met termination criterion **do**
 - 3: **for** each individual in P **do**
 - 4: **Mutation:** Select three distinct individuals from population. Compute the donor by adding the weighted difference of two individuals to the third.
 - 5: **Crossover:** Create trial individual by mixing parameters of current individual and the donor, decided by random draw and crossover rate.
 - 6: **Selection:** Compare trial and current individuals on using F . If the trial performs better, replace the current individual with the trial in the population.
 - 7: **end for**
 - 8: **end while**
 - 9: **return** Best individual from the final population as optimal parameters.
-

Differential evolution itself has parameters that must be defined in advance. The population size parameter in this algorithm defines the number of candidate solutions it

considers during each iteration. Those candidates are selected following uniform distribution in the parameter space which achieves evenly distributed candidates. There is a trade-off between high population size leading to finer exploration of parameter space and low population size leading to faster conversion, albeit not optimal. I set it to 15 to achieve both good exploration and conversion speed. The second parameter, the crossover probability, controls the extent to which the algorithm combines information from different solutions. The higher value will further diversify the population, encouraging exploration of new regions in the parameter space. On the other hand, lower values will lead to more exploitation of the current space. I have set this to 0.5. For other parameters, I will use the values suggested in the literature (Omidi and Mazaheri, 2020).

Evaluation metric

After discussing the importance of optimal parameter search and defining the optimization algorithm, the remaining question is the evaluation metric of the MDbS. Previous studies which have applied various dynamic choice models to experimental data have used the mean absolute error of aggregate choice shares for the entire dataset as the main metric. Albeit an interesting approach itself, this will not be a feasible approach for me because the experimental data these models have been applied to entailed ternary choice sets, whereas the observational data are not ternary. It comprises choice sets with minimum of 2 and maximum of 99 alternatives.

Designers and engineers of choice sets have long used “Top n ” accuracy metrics when designing choice sets or testing the performance of statistical models (Ricci et al., 2015). “Top n ” accuracy metric measures whether or not the true class of the option matches the top n predictions of the model. I will follow and adopt this metric because it is well established in the literature, mirrors the real-world decision making, and it fits the contribution of the thesis the best. I will use “Top 1” accuracy, which ranges from 0 to 1 as my optimization metric for an individual choice set. Because choice sets in the data vary significantly in size, I will use the average Top 1 accuracy metric weighted by menu size. This will ensure that smaller menus contribute proportionally to their sizes. Also, to comply with the aim of minimization of the differential evolution algorithm, I will multiply this metric by -1. As an additional measure to explore the parameter space thoroughly, I employ the multiple-run approach for the differential evolution algorithm.

Specifically, I will execute it ten times across the entire dataset. This reposition will allow me to further explore the parameter space, mitigating the risk of missing any region that can potentially contain an optimal solution.

2.6 Results

This section presents the initial results of parameter optimization, followed by the outcomes derived from choice modeling. First, an examination of the results obtained from observational data takes precedence. Afterwards, a concise discussion of the results derived from experimental data will accompany this analysis.

Observational data results

When looking at optimization results on observational data one can immediately see that the values of β_0 tend to fluctuate around 0.82 and 0.96 while β_1 is generally below 10%. which indicates that MDbS tends to be more strict in terms of defining the winners when comparing, on average preferring 90% of the “advantage” in a given dimension only a little shy of 15% of the time. Also, it appears that the α values tend to be preferred in the lower half of the parameter space, so only very similar alternatives were chosen by the model for comparison. This behavior is understandable considering that the average choice set had 55 alternatives. Table 3 contains the results of the parameter optimization using the differential evolution algorithm. It is worth noting that, its top 1 accuracy performances, albeit higher than random chance, still would fall far behind the pure statistical models, such as MNL based ones.

At first sight, such model behavior might seem surprising. Recall that the nature of MDbS is to compare alternatives with each other and collect evidence based on the won comparisons. In the choice experiments, the usual size of the menu is three, and only one of the context effects is generated at a given time. However, in observational data, the number of alternatives in the menu is much higher. The presence of a large number of alternatives potentially introduces also other context effects. Also, MDbS is bound to only dimensions which are mathematically comparable with each other. In the observational dataset there also present horizontal attributes for which only the decision maker can decide in a given scenario whether or not given the same price, flight duration, the flight which is at 5:00 in the morning is better than the one that is at 14:00 in the

Iteration	α	β_0	β_1	Average top 1 accuracy
1	2.58	0.912	8.832	0.125
2	1.622	0.948	6.001	0.124
3	1.88	0.843	6.876	0.124
4	1.883	0.832	6.83	0.124
5	1.856	0.91	8.297	0.124
6	2.154	0.954	8.021	0.124
7	2.204	0.963	8.52	0.124
8	0.234	0.859	7.03	0.123
9	0.559	0.829	55.076	0.122
10	0.235	0.844	58.893	0.122

Table 3: Optimization results for observational data.

afternoon.

I have estimated two models by using the random effect probit model with standard errors at cluster levels. The first model only included vertical attributes, whereas the second model extended the first one through the addition of the output from the MDbS model. In both cases, it seems that individuals have strong preferences for faster alternatives with lower prices and fewer layovers. This supports our initial assumption that individuals prefer lower values of vertical attributes.

Recall that variable “MDbS output” refers to the probabilities produced by MDbS. Model 2 results show a positive and statistically significant effect for the information provided. It shows that MDbS is able to capture additional information about the choice by accounting for context effects. To better understand the significance of this result, figure 7 shows the average marginal effects of the information provided by the computational model. One can immediately observe the downward trend. It is not surprising. As the number of alternatives increases, each additional alternative adds less to the likelihood of choice than the previous. In the context of menus, it could imply that when there are fewer options (smaller menus), the likelihood that any particular choice is selected is more significantly influenced by MDbS output. On average, for every 0.5 increase in MDbS output, the probability of choice has increased by 0.037 percentage points⁶. This effect was as high as 0.12 percentage points for menus containing as few as 5 alternatives.

Experimental data results

At first sight, the results of optimization indicate that the optimal parameters differ

⁶Marginal effect of MDbS output across the whole dataset was 0.074.

	Model 1	Model 2
Price	-0.309*** (0.006)	-0.282*** (0.006)
Trip duration	-0.185*** (0.007)	-0.158*** (0.006)
Number of flights	-0.195*** (0.007)	-0.178*** (0.007)
Number of airlines	-0.262*** (0.008)	-0.245*** (0.008)
MdBS output		2.085*** (0.097)
Constant included	Yes	Yes
Menu size as control	Yes	Yes
Number of observations	368,723	368,723
Akaike information criteria	48,532.341	47,968.61
Log-likelihood	-24,260.171	-23,977.305

Table 4: Outputs of Probit model with random effects for observational data. Standard errors in parentheses. Statistical significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

between datasets. Considering the nature of these two datasets, such a result is expected. Although the optimal α tends to be higher than the one in the observational data, both the optimal β_0 and β_1 values are lower than their counterparts. A higher α can indicate that for smaller choice sets, MDbS tends to be less strict in comparison criteria. While the performance metrics seem higher than for the observational data, the menus are considerably smaller. Table 5 gives further information.

Overall, considering the differing natures of these two datasets, comparing two optimal parameter combinations would not give any useful knowledge. However, this is not the case for the results from the choice modeling. These results follow the ones from observational data and confirm them. As with field data, here, the MDbS output is proven to provide statistically significant information for a choice model with coefficient in the positive direction. A 0.5 increase in MDbS’ “assessment” about the alternative resulted in a 0.46 percentage points increase in the actual choice probability among the participants. This effect did not differ between the choice sets having two or three alternatives.

Iteration	α	β_0	β_1	Average top 1 accuracy
1	2.888	0.572	4.567	0.52
2	2.918	0.577	4.598	0.52
3	2.939	0.569	4.655	0.52
4	2.747	0.579	4.349	0.519
5	2.936	0.743	3.911	0.518
6	2.925	0.715	3.967	0.518
7	2.997	0.706	4.098	0.518
8	3.342	0.61	4.93	0.517
9	2.01	0.494	3.788	0.517
10	2.01	0.494	3.788	0.517

Table 5: Optimization results for experimental data.

	Model 1	Model 2
X	0.000*** (0.000)	0.000*** (0.000)
Y	0.000* (0.000)	0.000 (0.000)
MdBS output		2.518*** (0.317)
Constant included	Yes	Yes
Menu size as control	Yes	Yes
Number of observations	5,295	5,295
Akaike information criteria	6,987.151	6,893.77
Log-likelihood	-3,489.576	-3,441.885

Table 6: Outputs of Probit model with random effects for experimental data. Standard errors in parentheses. Statistical significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

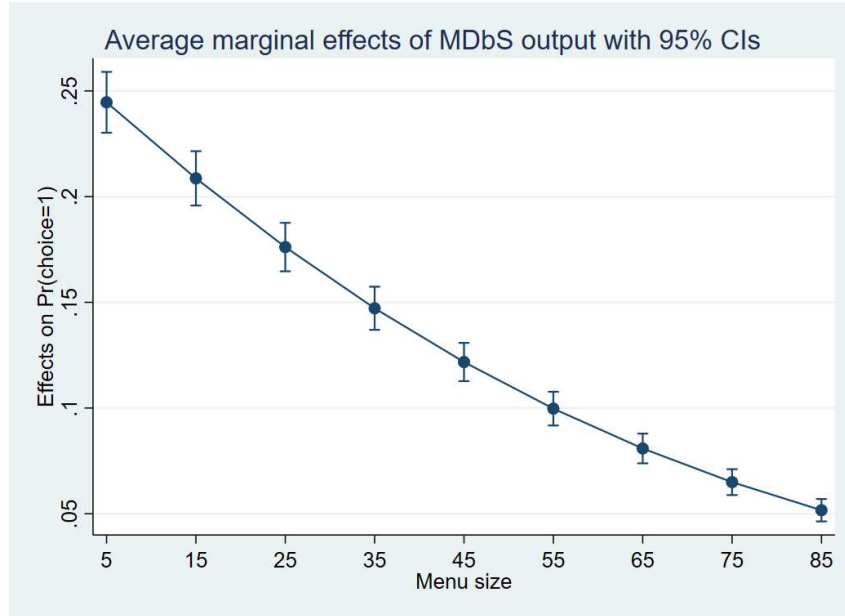


Figure 7: Average marginal effects of MDbS output with respect to different menu sizes. Horizontal lines represent 95% confidence interval boundaries.

2.7 Conclusion

In this study I have applied MDbS to observational data and showed the consistency of my findings using experimental data. This is the first account of an application of a computational model to real-world choice data of this magnitude. The results indicated that computational models can account for context effects that affect choice behavior not only in experimental settings but also in field settings.

The results of this study create implications for online marketplaces. In today’s world, these platforms aggregate immense amounts of products and services, providing consumers with dozens of choices. To help consumers in their choice, these platforms employ sophisticated algorithms which aim to curate product lists and create recommendations with a side goal of influencing the buying decisions of individuals. By applying mathematical decision-making models to choice datasets, these platforms can gain crucial information about the context within the choice sets, which might influence choice decisions towards particular alternatives. This information may also be used to create product bundles with a heterogeneous context to satisfy the needs of consumers.

This study has limitations. I have utilized only one computational model, namely MDbS. This limits the generalizability of my results. Different models “behave” differ-

ently, and although they are trying to capture the same effects, applying other decision models and investigating their differences can be an interesting avenue to pursue. Another limitation is that the proposed approach was applied only to data originating from one type of choice setting, namely, air travel. Application of this approach to other types of multi-dimensional, multiattribute choice data may help to better generalize the results.

The use of MDbS in the current study has shown it has the ability to potentially capture a wide range of context effects, including attraction, compromise, and similarity. While these results have yielded valuable insights, the general nature of MDbS and other computational models is their inability to successfully isolate these effects from one another. The main reason for that is that they have only been tested in experiments with one effect present at a time. When the number of alternatives in the dataset increases the potential interplay between options and the existence of other context effects come into play.

My findings provide a strong foundation that leads to a crucial but also challenging future direction: the development of a methodology which would allow disentangle this “general” context effect. I will computationally differentiate among three main components, attraction, similarity, and compromise in a multi-dimensional, multialternative choice setting. This goal provides great motivation for the next chapter of my thesis.

3 Enhancing Choice Modeling in Multi-Attribute, Multi-Alternative Settings⁷

Abstract

Previous approaches to modeling the effect of context on choices consider neat, compact environments, often in laboratory settings. Such an approach severely limits the study of context effects and, as a consequence, the applicability of findings. In this paper, the authors generalize the existing approach in modeling choice with context effects and apply it on large-scale observational data. The authors consider three main context effects: the attraction, compromise, and similarity effects. The proposed methodology relies on an ex ante calculation of each context effect measure for every alternative in the choice set. This approach minimizes the computational complications of estimating the resulting choice model. The proposed approach is applied to two empirical settings: the choice of airfare using observational data and the choice of daily commute mode using data from a stated choice experiment. The presence of attraction and similarity effects in both empirical settings is demonstrated. The authors also document the existence of the reverse compromise effect in airfare choice, highlighting the fact that travelers possess rigid rankings among flight attributes and are essentially maximizing their utility in terms of one (or few) attribute(s).

3.1 Introduction

The fact that behavioral biases exist in individual decision making is well established (see Dowling et al. (2020) for a recent review of evidence). One type of systematic departure from the classic utility maximization approach that seems particularly important is a set of context effects (Trueblood et al., 2013; Köcher et al., 2019). The theory behind these effects posits that the context in which choices are made influences the decision. While the choice context could have a very wide meaning, in this literature it is the availability and nature of choice alternatives which is referred to as “context” (Tversky, 1972; Huber et al., 1982; Simonson, 1989).

⁷This chapter is based on a joint work with my supervisor Zakaria Babutsidze, William Rand, Nobuyuki Hanaki, Ismael Rafai, Rodrigo Acuna Agost and Thierry Delahaye.

Context effects have been systematically studied in marketing and psychology (Kivetz et al., 2004; Rooderkerk et al., 2011; Frederick et al., 2014; Dotson et al., 2018). However, virtually all such studies have used controlled experiments in neat, compact settings. Namely, the settings where decision makers are presented with few options and (very) few attributes across which these options differ. In contrast, most actual choices take place in much messier environments. Especially today, when much of our search and shopping activity has shifted online. Proliferation of search engines allows each option to be easily compared with many alternatives across many different characteristics. In chapter 2 I focused on quantifying context as an aggregate. Although my previous study could successfully identify context in complex, multiattribute setting, the knowledge about the prevalence of context effects in these environments is still scarce. Precise measurements of context effects in multi-option and multiattribute setting is one way to contribute. Defining how to measure these effects is a minimum requirement for proceeding to evaluate the existence of context effects using observational data.

Recent attempts in computer science have been made to define some of these effects. The machine learning community has incorporated context effects in discrete choice models applied to observational data (Pfannschmidt et al., 2019; Bower and Balzano, 2020). However, the objective was to increase the prediction accuracy of the choice models (Tomlinson and Benson, 2021). As a result, incorporation of context effects takes the form of generalizing choice models to allow for departure from the strict rationality assumptions⁸. These proposed generalizations of estimated functional forms generally do not distinguish across various different types of context effects. Additionally, these approaches often run into computational difficulties, i.e. the estimation process is NP-hard (Yousefi Maragheh et al., 2020).

In this chapter, I propose measures of different context effects in multi-option and multidimensional settings. Following Rooderkerk, Van Heerde, and Bijmolt (2011), I consider three context effects - attraction, compromise, and similarity effects. Although I have discussed each of these three effects in more detail in the chapter 2.2, a very brief recap of this “trinity” seems appropriate. The attraction effect refers to the increase in

⁸Recent examples of this approach are Contextual Multinomial Logit by Yusefi Maragheh et al. (2020) and the linear context logit of Tomlinson and Benson (2021).

attractiveness of a set of options as a result of adding an alternative to the choice set, the compromise effect refers to the inclination of consumers to prefer options that represent a compromise across extreme sets of alternatives, while the similarity effect refers to the drop in choice likelihood for an alternative once another similar alternative has been added to the choice set. Each of the measures corresponding to the three aforementioned effects requires a specific approach to make the measurements applicable to the observational data. Each of these measures is calculated prior to choice estimation, which avoids computational problems. After presenting the generalized measures of the three effects, we perform an empirical analysis of the choices based on the new measures using observational data. We use an extensive dataset of airfare choices for this exercise. We identify that attraction and similarity effects influence the choices in air-travel booking data. We also detect a reverse compromise effect that seems to indicate that air travelers consistently prefer extreme alternatives (i.e., the cheapest or the shortest flight) to alternatives that constitute a compromise among extreme options.

3.2 Context effect and choice modeling

Over the years, multiple empirical models have been developed to model context effects. Empirical approaches usually model context effects in either the structural part of utility or in the error covariance part (Kamakura and Srivastava, 1984; Dotson et al., 2018). Some of these models have the capacity to take into account multiple effects at the same time (Tversky and Simonson, 1993; Orhun, 2009). These models extend a classical random utility model (McFadden, 2001) in multiple directions using discrete choice modeling (Ben-Akiva and Lerman, 1985). However, Rooderkerk, Van Heerde, and Bijmolt (2011) present a unifying model that takes into account all three context effects. Instead of using advanced statistical techniques to address violations of utility maximization assumptions associated with the existence of context effects (Luce, 1959), their approach focuses on additive specification and ex ante calculation of individual measures for each of the three context effects for each item in the menu. Namely, the authors assume that the choice estimator is additive in three context effects (along with a generic preference-driven part) and develop the methodology of quantifying three effects for each alternative prior to calculating the estimator. This is a particularly flexible approach, which also ensures that

the researcher does not run into computational difficulties (i.e., *NP* hard calculations). I follow the suite and formulate the utility that a consumer c attaches to an option i , under a given menu m , as being additive in two parts:

$$U_{c,i}^m = u_{c,i} + v_{c,i}^m$$

The first summand in this equation $u_{c,i}$ denotes an inherent utility that the consumer c can derive from the option i . This part depends only on the tastes of the consumer c towards the characteristics of the option i . It is independent of the other options contained in the menu. The second summand $v_{c,i}^m$, denotes the context-dependent utility. I additionally assume that the context-dependent part of the utility can be represented as a linear combination of three contextual effects,

$$v_{c,i}^m = a_1 \text{Attraction} + a_2 \text{Compromise} + a_3 \text{Similarity}.$$

Thus, the measures of three context effects that are necessary to estimate empirical discrete choice models based on the utility formulation above need to be computed ex ante. Measures developed by Roederkerk, Van Heerde, and Bijmolt (2011), are adapted to experimental data with a small number of alternatives in the choice set and a small number of attributes characterizing alternatives. This significantly limits the application of the unifying model of context effects. In the next section, I present a generalization of three context effect measures to multi-option, multi-attribute environment which will further allow for the application of the unifying model to observational data.

3.3 Generalizing context effect measures

Approach to generalization

Naturally, generalizing across many alternatives and many attributes presents challenges in both dimensions. The fact that theoretical underpinnings of the three effects are diverse does not simplify the task. In the following sections, I will discuss specificities involved in the generalization of each measure. First, however, I focus on common challenges.

Conceptualizations of contextual effects commonly hinge on the choice frequency com-

parisons between two alternatives. For example, in case of attraction effect, if adding a third alternative to a two-item menu induces some of the consumers to switch their choices to the other incumbent alternative - one could conclude that attraction effect is present. This is suitable for experimental setups where the researcher has control over menus and can observe choices in both cases (i.e., in case of an original two-item menu, as well as after adding the third item). However, given that the aim is to generalize context effect measures for application to a wider range of situations, and most importantly to observational data, it is necessary to take a more fine-grained view and quantify the context in which each of the alternatives is embedded. Quantifying the choice context for each alternative would create an opportunity to study the effect of the context on choice probabilities through inference across (very) different choice sets. Such an approach would be general enough to consider not only the addition of a new alternative to the menu, but also any alteration of attributes for any of the items in the menu. For example, increasing the price of an alternative could decrease the probability of its choice. This would have a direct effect on the choice probabilities of other alternatives. However, the same price increase could also change the choice context and have additional knock-on effect on choice probabilities of (at least some) alternatives.

Rooderkerk, Van Heerde, and Bijmolt (2011) take this approach for simple two-attribute products. Using the attraction effect as an example once more, the idea is to quantify how much attraction power does a given menu provide to a given alternative. If option A dominates option B (i.e., it is superior in at least some attributes and not inferior in any of the attributes), while option B is not dominated by option C , the attraction power of A compared to C could be measured by the degree to which option A is better than option B . The more pronounced the dominance, the more pronounced the attraction effect. However, once we leave a neat context of two-item menus, we wonder into a possibility that option A dominates not one, but multiple alternatives at the same time. Consider the situation depicted in figure 8. Here we have the menu with six alternatives A, B, C, D, E, F each of which is characterized by two attributes $V1$ and $V2$. In this example, option A dominates three alternatives B, C, D . To simply extend the approach by Rooderkerk, Van Heerde, and Bijmolt (2011) and calculate the attraction power of the alternative A , we could find the center among the three dominated variables

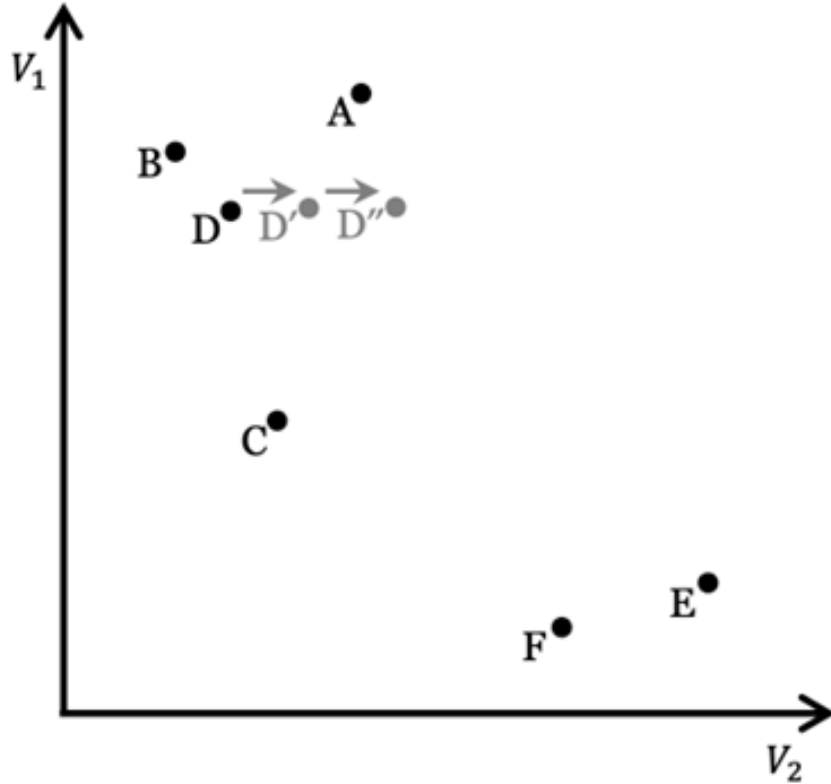


Figure 8: Visualization of accounting for attraction effect.

Note: The figure represents three alternative choice sets, each comprising of six options, described along two characteristics. Options A, B, C, E and F are common across three menus. Menus differ in the identity of the 6th option (D, D' or D'').

and then measure the distance. Such a measure would capture the difference between two choice sets A, B, C, D, E, F and A, B, C, D', E, F . In the latter case, the attraction power of A' is lower because the option D' is closer to A than D . However, such a measure would not accurately capture the difference between scenarios A, B, C, D', E, F and A, B, C, D'', E, F . In the latter case A dominates only two alternatives B, C . Therefore, the setting changes qualitatively. Such qualitative differences are avoided in experimental settings by design. However, they are prevalent in observational data. While it is acknowledged that the move from D to D' changes the choice context, I argue that the context change is more pronounced in the case of the move from D' to D'' . Although the ideal measure would combine the characteristics of the number of dominated alternatives and the (some measure of average) distance between the focal alternative and the group of dominated options, in this article I take the approach of focusing on the former, as this

is likely to have a more pronounced impact⁹.

As a result, our approach would capture the context change between A, B, C, D, E, F and A, B, C, D'', E, F or A, B, C, D', E, F and A, B, C, D'', E, F , but it will not evaluate the context difference between A, B, C, D, E, F and A, B, C, D', E, F . In what follows, the same approach is applied to similarity and compromise measures.

Once one moves towards choices which have multiple attributes, it is quick to realize that there are two distinct types of choice characteristics that our measures should potentially handle. One type of attributes constitute product characteristics over which preferences are fairly similar for all customers, and their effects can be readily anticipated from basic economic theory. These attributes can easily be ordered from most preferred to least preferred. The most obvious of these characteristics is price. One can assume that every customer would prefer to obtain a given product for a lower price. We call such product characteristics vertical attributes. These are usually attributes that can be represented using numeric values. Previous work measuring context effects only considers such (vertical) attributes (Trueblood et al., 2014; Noguchi and Stewart, 2018, 2014). This is a requirement for defining preferential relationships that are necessary to identify attraction and compromise effects. The same approach is adopted, where I consider only vertical attributes when defining attraction and compromise effects.

On the other hand, there exists another set of attributes where there is no obvious, homogeneous ordering. For example, consider the color. There is no theoretical ground to assume that all consumers would prefer a car that is blue over a car that is green (all other attributes remain constant). The same is true about attributes that at first sight are not strictly labeled as categorical, for example, time. When buying a cinema or plane ticket, there is no theoretical reason for explaining how a ticket for 15:00 is better or worse than one at 17:00. I refer to these as horizontal attributes. The potential heterogeneity between decision makers in ordering such attributes makes inclusion of such features in the calculation of attraction and compromise effects impossible. In experimental settings,

⁹Combining frequency and distance measures in one metric requires arbitrage across the two drivers of context effects. It is not clear how to solve such a problem (that is, it is not clear if dominating one option that is at a certain distance from a focal alternative generates more or less attraction than dominating two alternatives that are at a half that distance).

these attributes are often constant between treatments to avoid confounding effects. However, in the field, this usually cannot be done. Therefore, the study of context effects with observational data requires them to be statistically controlled.

However, unlike the measurement of attraction and compromise effects, measuring the similarity across the alternatives does not require the existence of a single universal ranking. In fact, many clustering methods can identify options that are more or less similar to each other based on a wide range (numeric and categorical) of variables. Therefore, in what follows, I will incorporate all (vertical, as well as horizontal) attributes in the measurement of similarity between a pair of alternatives.

Attraction effect

Previous studies of the attraction effect concentrate on carefully designed small choice sets in experimental settings (Huber et al., 1982; Huber and Puto, 1983). In such settings, an alternative is added to the choice set in a position that is unequivocally inferior to (only) one of two items already present in the menu. Notice again that identification of inferiority requires the attribute under consideration to be vertical, and this cannot be achieved with horizontal attributes. This manipulation introduces an asymmetry between the two incumbent alternatives; one alternative now dominates the decoy, while the other does not. The attraction effect implies that such manipulation increases the attractiveness of the dominant incumbent option with respect to the other incumbent alternative.

A standard measure of the attraction effect considers a trade-off between two (vertical) characteristics. Let us consider $i \in \mathbb{N}$ vertical attributes V_i for a set of two options A and B . In two dimensions $n = 2$, we start with $V_1(A) > V_1(B)$ and $V_2(A) < V_2(B)$, and then introduce an alternative C such that $V_1(A) > V_1(C) > V_1(B)$ and $V_2(C) < V_2(A) < V_2(B)$. Under such circumstances, C is dominated by A , but not by B . This introduces asymmetry in consumer considerations and increases the probability that the consumer will choose option A . Generalizing this concept to multiple (vertical) attributes is straightforward. For $N > 2$, we again start with A being preferred over B in some $j > 0$ dimensions, while B is preferred to A in some others $k > 0$, such that $j + k \leq N$. Then we need an alternative C that will be strictly worse than A in at least one dimension, while not being better in any other dimensions and being better than B in some dimensions while being worse in some others. As long as these two conditions are satisfied, the

attraction effect states that C will result in A being favored.

Generalizing this approach to multiple alternatives is somewhat more challenging. The reason for this is that, instead of one comparison (A vs. B in the case above), for a set of choices with M alternatives, there are $\frac{M(M-1)}{2}$ potential comparisons to consider. Under real-life circumstances, it is easy to identify situations where more than one of the $\frac{M(M-1)}{2}$ relationships has the potential for an attraction effect. Besides, for any given pair of choices, we could have multiple decoy options generating attraction effect. The final complication is that option A may have one set of decoy alternatives and option B another set of decoy alternatives. In these contexts, it is not clear which option the attraction effect favors.

To quantify the attraction effect generated by the menu for a given alternative, I propose to calculate the number of options present in the menu that the focal alternative dominates. This is done across all vertical dimensions. Then, two alternatives present in the same menu can be compared by examining how many choices they dominate. Under such circumstances we can consider different positions option C can take with respect to options A and B . If option C is neither superior (dominant) nor inferior (dominated) by any of the options A and B , or if it dominates both focal options, then it does not generate an attraction effect for A or B . If option C is dominated by both options in the focal pair, it generates an attraction effect for both of them (compared to other alternatives). In all of these cases, the location of option C contributes similarly to the choice probability of both options A, B . Finally, if option C is dominated by only one of the two focal alternatives (say by A , but not by B) - it generates a discriminatory attraction effect favoring option A and increasing its probability of being chosen. As a result, the number of options that the current alternative dominates in a menu (appropriately normalized by the size of menu for a comparison across different choice settings) measures the (relative) extent of the attraction effect generated by the menu. For example, compare the probability of choosing option A versus E in figure 8 between two sets of menus A, B, C, D, E, F and A, B, C, D'', E, F . This probability is higher in the former situation (where A dominates three alternatives, while E dominates one) than in the latter case (where A only dominates two alternatives, while E still dominates one). Although in these cases both alternatives do have some attraction effect, the relative attraction effect of option A compared to

option E is stronger in the former scenario. Therefore, I measure the attraction effect that favors the focal option F as

$$Attraction(F) = O(Dominated),$$

where $O(Dominated)$ measures the number of alternatives in the menu that the focal option F dominates. Given the measure, we expect that the higher the attraction effect in favor of the focal option, the higher the probability of choice of the focal option (*ceteris paribus*).

Compromise effect

The compromise effect is traditionally understood and operationalized in a three-option, two-attribute (experimental) setting (Simonson, 1989; Dhar et al., 2000). It is worth mentioning again here that these two attributes need to be vertical so that we can define universal preference relationships. Let us consider the similar starting situation of options A and B as in the previous subsection: $V_1(A) > V_1(B)$ and $V_2(A) < V_2(B)$. The addition of option C to this menu such that $V_1(C) > V_1(A) > V_1(B)$ and $V_2(C) < V_2(A) < V_2(B)$, makes option A a compromise between two extreme options B, C . The compromise effect maintains that such an alteration of the menu would disproportionately benefit alternative A compared to alternative B .

To formulate the general measure of the compromise effect, let's first consider the case of multiple options M in two dimensions (attributes, $N = 2$). The compromise effect calculation over multiple options is visualized on figure 9 with $M = 7$ case. To quantify the extent of the compromise introduced by the focal option F in the menu, I propose to split all other alternatives $M - 1$ into four groups. Let group 1 contain all alternatives for which $V_1(G_1) \leq V_1(F)$ and $V_2(G_1) \leq V_2(F)$. These are the alternatives dominated by the focal option. In the case of figure 9, this set contains only the option A . Let group 2 contain all alternatives for which $V_1(G_2) \geq V_1(F)$ and $V_2(G_2) \geq V_2(F)$. All these options dominate the focal option. This set contains option E in figure 9. Clearly, the focal option cannot constitute a compromise between any pair of alternatives which is included in any of these first two groups of alternatives. Next, let group 3 contain all alternatives for which $V_1(G_3) > V_1(F)$ and $V_2(G_3) < V_2(F)$, and group 4 contain all alternatives for which $V_1(G_4) < V_1(F)$ and $V_2(G_4) > V_2(F)$. In the case of figure 9, group 3 contains

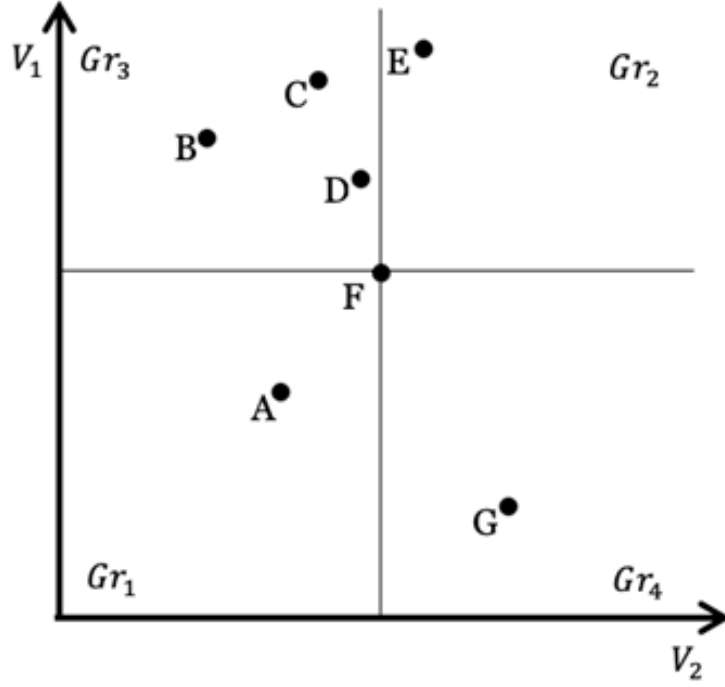


Figure 9: Visualization of the compromise effect generalization.

Note: The figure represents a generalization of the compromise effect across multiple alternatives. F represents a focal option. Gr_1 collects alternatives dominated by F , Gr_2 collects alternatives dominating F . Focal option represents a compromise between alternatives in Gr_3 and Gr_4 .

options B , C and D , while group 4 contains option G . The focal alternative can be viewed as a compromise between groups 3 and 4. For the quantification of the extent of such a compromise, I define

$$\text{Compromise}(F) = \frac{\min(O(G_3), O(G_4))}{\max(O(G_3), O(G_4))} * (O(G_3) + O(G_4)), \quad (10)$$

where $O(G_i)$ measures the number of alternatives in the group i . The first multiplier (the ratio) in the measure quantifies the asymmetry across the sizes (in terms of number of alternatives) of the two groups, while the second multiplier (the sum) quantifies the joint size of two groups across which the focal option is a compromise. For the option F in figure 9, this value is $\text{Compromise}(F) = \frac{1}{3} * 4 = 1.33$. If any of the two concerned groups are empty, the value is zero, corresponding to the fact that the focal alternative is at the extreme edge of one of the dimensions and therefore is not a compromise. As a

result, our compromise measure will be strictly zero for options B , C , E and G . On the other hand, the better the balance between the size of the two groups, the more valuable compromise alternative F provides. So, the same measure for option D in figure 9 is 4. Alternative D also corresponds to the compromise between 4 alternatives (like option F), but the comparison groups are better (in this case, perfectly) balanced. Notice that the measure also increases the number of total options in two comparison groups. Notice that the same measure for option A is 2, even though (similar to option D) it also exhibits the correct balance between the sizes of two comparative groups. This reflects the fact that option D is a compromise between larger sets of extreme alternatives¹⁰.

Extending the compromise measure to multiple dimensions is somewhat more challenging. The challenge relates to the fact that increasing number of dimensions (i.e., vertical dimensions) presents exponentially increasing opportunities of different ways a given option can be a compromise. The $N = 2$ case has one pair of groups to compare. However, in the case of $N = 3$, a focal alternative can be a compromise between multiple pairs of option groups. For example, option F can be a compromise between two groups Z and Y such that all options in group Z are superior to option F in dimensions 1 and 2, but inferior in dimension 3, while options in group Y are inferior to option F in dimensions 1 and 2, but superior in dimension 3. Permutation calculus guarantees that there are three such potential comparisons. However, this is not all. Option F can also be a compromise between two groups X and W such that all options in group X are superior to option F in dimension 1, but inferior in dimension 2, while options in group W are inferior to option F in dimension 1, but superior in dimension 2, as long as dimension 3 is constant in all options in groups X and W , as well as F . The permutation calculus guarantees three additional such comparisons.

As a result, moving from 2 to 3 dimensions increases the number of potential compar-

¹⁰An alternative way to quantify the compromise between two sets of extreme options is to count the number of all possible pairs for which a given focal option is a compromise. This would result in $Compromise'(F) = O(G_3) * O(G_4)$. This measure behaves very similarly to the one discussed in the paper. In fact, the correlation between the two compromise measures in the dataset used in this paper is 0.825. All the results reported in the paper are qualitatively unaltered by the replacement of the compromise measure with this alternative. However, I prefer to work with the compromise measure in the paper, as it takes a more “collective” view of the choice process.

isons to calculate the value of the alternative as a compromise option in the range from 1 to 6. Appendix [B](#) derives the number of comparison alternatives necessary to cover all potential ways in which a focal alternative can be a compromise as a function of the number of dimensions. However, as all the groups defined above are mutually exclusive (i.e., each alternative can only belong to one and only one of such potential comparison groups), generalization of the compromise measure in N dimensions would require summation of the specific comparison group’s compromise measure over all comparison groups. And thus,

$$\text{Compromise}(F) = \sum_j \text{Compromise}(F)_j, \quad (11)$$

where j runs over all possible comparison groups. Summation, instead of averaging, is used in order to reward options that constitute a compromise across multiple (many) comparison groups. Given the measure of the compromise effect, I expect that the higher the compromise effect, the higher the probability of choice of the focal alternative.

It is worth mentioning here that as both attraction and compromise measures only generalize across vertical dimensions, it is important to control for all relevant horizontal dimensions in choice models employing these measures of the two context effects.

Similarity effect

Operationalizing the measure of the similarity effect in three options and two vertical dimensions is straightforward ([Roederkerk et al., 2011](#)). Increasing the size of the menu introduces an important challenge of defining the border between options that are similar to the focal alternative and those that are not similar to it. At the same time, unlike the previous two context effects, the theory pertinent to the similarity effect does not require dimensions to necessarily be vertical ([Tversky, 1972](#)). The sufficient condition to quantify the similarity effect requires detecting the number of other alternatives that are similar to the focal option.

Clustering, using machine learning, gives a possibility to operationalize the measure of the similarity effect across all dimensions. Several clustering algorithms have been developed that can take multidimensional lists and partition them into groups of similar

objects. Clustering algorithms are unsupervised machine learning techniques that do not require explicit guidance on the definition of similarity. They use different internally consistent evaluation criteria in order to partition the input group of objects into multiple subgroups. Items belonging to the same group are judged to be similar to each other, while items belonging to two different groups are regarded dissimilar. Some algorithms, like K-means clustering (Lloyd, 1982), require additional input on (or an optimization layer for calculating) how many subgroups the user would like to detect. Others, such as Affinity Propagation (Frey and Dueck, 2007), automatically calculate the optimal number of detected clusters. Appendix C provides a summary of two popular clustering algorithms that can be used for this purpose. We argue that being able to autodetect the number of clusters is significant in terms of minimizing necessary input, as well as minimizing computational power, and use Affinity Propagation in the empirical application below.

As a result, I propose using a clustering algorithm (in this case, Affinity Propagation) in order to detect clusters within the menu of proposed options. Once such clusters have been identified, the size of the cluster to which the focal option belongs can be used as a straightforward measure of similarity. Hence, I measure the similarity effect as

$$\text{Similarity}(F) = O(\text{Cluster}_F),$$

where, Cluster_F refers to the cluster to which the focal option belongs. Given this measure of the similarity effect, we expect that the higher the similarity, the lower the choice probability of the focal alternative.

3.4 Empirical applications

In this section, I present two empirical applications using the generalization of three contextual measures and an estimate unifying model of context effects. Both applications come from a travel context in Europe. The first application uses a large set of observational data on airfare booking. This is a very heterogeneous dataset, and choice sets vary in terms of the number of alternatives, as well as between city pairs of origin-destination. The second application uses experimental data of stated choice in urban commutes. These data are less exciting in terms of menu variability, but it allows one to address several

potential concerns with the main observational dataset. As a result, this is used as a validation exercise.

Observational data

Observational data I use are the same as that used in the previous chapter. The dataset is described in detail in section 2.5. These data have been subject to preprocessing rules, which are also discussed in the section 2.5. The descriptive statistics of the context variables are shown in the table 7. The descriptive information for other covariates can be found in the table 1.

Variable	Count	Mean	St.Dev	Min	Max
Attraction	368,723	19.78	20.33	0	98
Compromise	368,723	1.73	3.96	0	63.01
Similarity	368,723	11.27	5.88	1	77

Table 7: Descriptive statistics of context variables.
Note. Statistics before normalization.

Measurement of context effects

The measurement of the attraction and compromise effects is pretty straightforward. I follow the methodology outlined in the previous section. For the attraction effect, I count the number of alternatives dominated by a given option within the menu. This is implemented across all four vertical attributes. The compromise effect, given by equation 11, is also measured across all four dimensions. This results in 25 pairs of comparison groups for each alternative (see equation 17 in Appendix B).

However, before proceeding to the measurement of similarity effect, it is needed to normalize the flight departure variables for the clustering algorithms. In order to identify similar alternatives within the menu, the clustering method needs a variable that allows it to measure the distance between any two departure values. This is achieved by transforming these variables into the Coordinated Universal Time format preserving dates, hours, and minutes of departure time. This way, the algorithm is able to measure the distance between any pair of alternatives in minutes. For normalization purposes, I also subtract the timestamp of the earliest flight in a menu from the departure times of every flight in that menu; thus, all times are measured as times after the earliest time.

After this transformation, I use Affinity Propagation for obtaining sets of similar options within each menu. I feed the clustering algorithm with the data on all the vertical

and horizontal variables for each alternative. The algorithm returns an identifier for the group of options that comprises each of the options. Affinity propagation detects on average 7.62 clusters within the choice sets. In order to develop the measure of the similarity effect, I calculate the number of alternatives in the cluster to which the focal alternative belongs.

Choice modeling

To examine the context effects on choices in the airline booking data, I estimate random effects Probit models augmented by the context effect measures. These models have the crucial advantage of interpretability. Another advantage (over, for example, Logit) is the feature that Probit does not explicitly require the assumption of the independence from irrelevant alternatives. If the augmented model perfectly accounts for all context effects (IAA), this would not be a concern. However, as one cannot guarantee that human choices are not affected by any other context features (that have not yet been hypothesized and examined), having this feature is an additional advantage. I, however, also present robustness checks by fitting alternative statistical models in the Appendix [D](#) (Logit, Mixed Logit, and Fixed Effects Probit) [11](#).

An important point to note here is the fact that the context variables incorporate menu size effects. For example, the attraction variable cannot take any value higher than 5 in a menu of size 6. However, the same variable can take the value of 49 in the 50 size menu. One way to deal with this feature would be to normalize the context variables by the size of the menu. Another alternative is to account for this feature statistically by controlling for the size of the menu in the regression equation. I opt for the latter because it guarantees higher flexibility in the empirical model structure. It also allows to account for menu size effects that could go further than context effects (for example potential choice overload). An additional advantage is that it is much simpler to interpret marginal effects of unscaled context variables.

¹¹An additional robustness check in terms of the usage of the clustering method is also presented in the same appendix. There I use K-mean clustering (augmented with the use of Silhouette score ([Rousseeuw, 1987](#)) to calculate the optimal number of clusters) to calculate the similarity measure. The results are robust to this alteration as well.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Price	-0.225*** (0.006)	-0.232*** (0.006)	-0.200*** (0.008)	-0.227*** (0.006)	-0.219*** (0.006)	-0.184*** (0.008)	-0.186*** (0.007)	-0.184*** (0.007)	-0.186*** (0.007)
Trip duration	-0.136*** (0.009)	-0.116*** (0.009)	-0.087*** (0.010)	-0.112*** (0.009)	-0.124*** (0.009)	-0.093*** (0.010)	-0.092*** (0.010)	-0.094*** (0.010)	-0.092*** (0.010)
Number of flights	-0.342*** (0.008)	-0.322*** (0.008)	-0.325*** (0.008)	-0.309*** (0.008)	-0.289*** (0.009)	-0.279*** (0.009)	-0.289*** (0.009)	-0.281*** (0.009)	-0.290*** (0.009)
Number of airlines	-0.208*** (0.010)	-0.209*** (0.010)	-0.197*** (0.011)	-0.210*** (0.010)	-0.205*** (0.010)	-0.193*** (0.010)	-0.199*** (0.010)	-0.193*** (0.010)	-0.199*** (0.010)
Attraction			0.003*** (0.001)			0.003*** (0.001)		0.003*** (0.001)	
Compromise				-0.038*** (0.003)		-0.034*** (0.003)	-0.031*** (0.003)		
Similarity					-0.020*** (0.001)	-0.020*** (0.001)	-0.031*** (0.002)	-0.020*** (0.001)	-0.031*** (0.002)
Attraction within cluster							0.025*** (0.002)		0.0245*** (0.002)
Attraction outside cluster							0.002*** (0.001)		0.002*** (0.001)
Compromise within cluster								-0.149*** (0.020)	-0.136*** (0.020)
Compromise outside cluster								-0.028*** (0.004)	-0.027*** (0.004)
Constant included	YES	YES	YES	YES	YES	YES	YES	YES	YES
Horizontal variables as controls	NO	YES	YES	YES	YES	YES	YES	YES	YES
Number of observations	368723	368723	368723	368723	368723	368723	368723	368723	368723
Number of choices	6297	6297	6297	6297	6297	6297	6297	6297	6297
Consistent Akaike information criterion	49592	48906	48878	48764	48716	48565	48477	48555	48469
Log likelihood	-24761	-24356	-24335	-24278	-24254	-24165	-24114	-24153	-24103

Table 8: Choice model estimation results.

Notes: Outputs from random effects Probit regressions. Standard errors in parentheses. Statistical significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Before going to the estimation, one needs to transform the departure-time variables into outbound and inbound flight pairs. The transformation that was performed for the clustering exercise cannot be used directly because it estimated coefficients that are not interpretable. To make this information as tractable as possible, I generate a set of variables. First, a day-of-the-week variable is generated for the outbound flight. Second, a variable is generated that measures the duration of stay at the destination ¹². These two variables together describe the inbound and outbound flight timing characteristics at the level of the day. However, consumer preferences can be defined on a smaller scale. Therefore, I also generate two variables that describe the exact time of the day of the outbound and inbound flights. These variables, $t_{out}, t_i \in [0; 1)$, are measured as a fraction of a day such that $t_i = 0$ corresponds to midnight, while $t_i = 0.5$ corresponds to midday. I further apply a cosine transformation to these variables, i.e., $\cos(t_i) = 2\pi t_i$. This confines the departure time variable to the interval $[-1; 1]$, and ensures the smooth transition in departure times at midnight. These transformations result in a total of four variables describing departure timestamps for outbound and inbound flight pair - the horizontal attribute of the alternative.

I estimate a sequence of 9 models and present the results in table 8. I estimate these models by using random-effects Probit regressions with robust standard errors. First, I start out by fitting two simple baseline models of consumer choice. Model 1 is the simplest estimation, which includes only the four vertical attributes as independent variables. Model 2 further extends this model by adding four horizontal attributes. In both cases, with or without horizontal attribute controls, all vertical variables generate meaningful results. Consumers clearly have preferences for shorter, cheaper flights with fewer layovers and airline changes. Travelers also seem to have preferences for the outbound flight during the day and for the inbound flight during the night (recall that the cosine transform variables reach a maximum at midnight and a minimum at midday).

To further extend model 2, three models (3 through 5) that each incorporate one of the context effects, and one model that incorporates all three context effects at once (model 6) were estimated. Table 8 indicates the consistency between the coefficient estimates of

¹²For the regression analysis, similar to other numeric variables, in order to eliminate any scale effects, I perform a z-score transformation of duration of stay variable.

Variable	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Attraction	0.0187		0.0170		0.0173		
Compromise		-0.1900		-0.1646		-0.1509	
Similarity			-0.0747	-0.0756	-0.1222	-0.0740	-0.1207
Attraction within cluster					0.1042		0.1040
Attraction outside cluster					0.0133		0.0135
Compromise within cluster						-0.6039	-0.5444
Compromise outside cluster						-0.1377	-0.1323

Table 9: Average marginal effects for relevant models.

Note. Average marginal effects implied by various models. All p-values were significant at $p < 0.001$ level.

model 6 and those of models 3-5. This set of models also allows one to evaluate the effect of the three context effects on consumer choice. In line with the theory, the presence of attraction and similarity effects is observed. That is, if the attraction measure increases for a given option, this increases the likelihood that the option is chosen. On the other hand, when the similarity measure increases, it decreases the probability that the option is chosen. Both of these effects are statistically highly significant and are in the hypothesized direction. To better understand the economic significance of the estimated effects, table 9 presents (average) marginal effects of the relevant models. From table 9 one can read that if the attraction measure increases by one unit (that is, having one more option dominated by the focal alternative, *ceteris paribus*) the likelihood of an option being chosen goes up by about 0.02 percentage points on average. On the other hand, if the similarity measure increases by one unit, the likelihood that a given option is chosen goes down by about 0.08 percentage points on average.

Tables 8 and 9, however, also indicate the existence of a reverse compromise effect. The compromise effect posits that if an option represents a compromise between extreme alternatives, it will have a higher likelihood of being chosen. On the contrary, our results indicate that increasing our compromise measure decreases the likelihood of an option being chosen. This effect is again statistically and economically significant. From this

we can conclude that in the context of airfare choice, consumers prefer extreme options to those that represent compromise. This implies that the preferences of individual consumers are strongly anchored to one of the four vertical attributes. For example, if a traveler attaches particular importance to price, she will be reluctant to trade away an option that is cheap for increases in the attractiveness in any other (vertical) dimension. This, in fact, is rather understandable given the context of current empirical exercise: The two largest groups of air travelers are holidaymakers, who are price-sensitive and do not readily trade away price advantage for shorter travel time, and business travelers, who are time-sensitive and do not trade away flight duration for a decrease in price.

Next, I investigate the interaction between several context effects. Previous literature has hypothesized and demonstrated the interaction between attraction and similarity effects in laboratory environments (Huber et al., 1982; Huber and Puto, 1983; Roederkerk et al., 2011). The interaction between similarity and compromise effects has not been studied in literature; however, one can consider that if the similarity effect efficiently identifies comparable alternatives that could constitute a consideration set of the consumer, the compromise effect, which considers options outside the consideration set, will not constitute an adequate guide for consumer behavior. Given that the similarity measure hinges on identifying clusters of similar options, the interplay between the similarity effect and the other two context effects is rather straightforward to study.

For this, four additional measures are calculated for each option that decomposes attraction and compromise effects along the cluster lines identified by the similarity measure. More precisely, the attraction and compromise measure for a given option is calculated: 1) by taking only the alternatives that belong to the same cluster to which this particular option belongs, and 2) by considering only the alternatives that do not belong to the same cluster. This way, one can get a measure of attraction and compromise effects of an alternative within the cluster (i.e., among comparable alternatives, or within the consideration set) and outside the cluster (i.e., among relatively noncomparable alternatives, or outside the consideration set).

In models 7 through 9 I study the comparative effects of pairs of these effects. Model 7 decomposes the attraction effect of model 6 into two parts (inside and outside the cluster). Model 8 decomposes the compromise effect along the same lines, and model 9

estimates the model that includes both decompositions simultaneously. The results are again consistent and meaningful. Models 7 and 9 imply that attraction effect within the cluster, i.e., among comparable alternatives, has a much stronger impact on the purchase likelihood than that of the impact of the measure calculated based on nonsimilar alternatives. Table 9 indicates a difference in the size of the order of magnitude. Similarly, as indicated by models 8 and 9, being a compromise among comparable alternatives has a much higher detrimental effect on purchase likelihood than being a compromise among remote alternatives. The p values for all tests of coefficient pair equality (i.e., estimated coefficient for attraction within cluster being equal to that of the attraction outside cluster, and coefficient for compromise within the cluster being equal to the compromise coefficient outside the cluster) are below 0.001, indicating that cluster-based measure of similarity may be an efficient indicator of the consumer's consideration set.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Price	-0.249*** (0.008)	-0.259*** (0.009)	-0.253*** (0.013)	-0.257*** (0.009)	-0.249*** (0.009)	-0.238*** (0.013)	-0.245*** (0.013)	-0.238*** (0.013)	-0.245*** (0.013)
Trip duration	-0.167*** (0.013)	-0.143*** (0.013)	-0.137*** (0.016)	-0.143*** (0.013)	-0.147*** (0.013)	-0.139*** (0.015)	-0.144*** (0.015)	-0.138*** (0.015)	-0.143*** (0.015)
Number of flights	-0.414*** (0.014)	-0.392*** (0.014)	-0.393*** (0.014)	-0.386*** (0.014)	-0.373*** (0.014)	-0.369*** (0.015)	-0.373*** (0.015)	-0.371*** (0.015)	-0.375*** (0.015)
Number of airlines	-0.161*** (0.015)	-0.163*** (0.015)	-0.160*** (0.015)	-0.164*** (0.015)	-0.164*** (0.015)	-0.162*** (0.015)	-0.165*** (0.015)	-0.161*** (0.015)	-0.164*** (0.015)
Attraction			0.002 (0.003)			0.002 (0.003)		0.003 (0.003)	
Compromise				-0.066*** (0.013)		-0.063*** (0.013)	-0.064*** (0.013)		
Similarity					-0.023*** (0.003)	-0.023*** (0.003)	-0.029*** (0.004)	-0.022*** (0.003)	-0.029*** (0.004)
Attraction within cluster							0.016*** (0.005)		0.016*** (0.005)
Attraction outside cluster							-0.001 (0.003)		-0.001 (0.003)
Compromise within cluster								-0.182*** (0.048)	-0.180*** (0.048)
Compromise outside cluster								-0.029 (0.018)	-0.031* (0.018)
Constant included	YES	YES	YES	YES	YES	YES	YES	YES	YES
Horizontal variables as controls	NO	YES	YES	YES	YES	YES	YES	YES	YES
Number of observations	79080	79080	79080	79080	79080	79080	79080	79080	79080
Number of choices	3954	3954	3954	3954	3954	3954	3954	3954	3954
Consistent Akaike information criterion	25568	25124	25134	25106	25088	25085	25087	25101	25104
Log likelihood	-12759	-12482	-12481	-12467	-12458	-12444	-12439	-12446	-12441

Table 10: Choice model estimation results from the reduced dataset.

Notes: Outputs from random effects Probit regressions. Standard errors in parentheses. Statistical significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

An important drawback of the dataset is the feature that there is no way to know which available options on the market reached the eyeballs of the consumer. One way of thinking about this problem is to consider the most likely way such menus are delivered to decision-makers. Most online booking sites and flight aggregators use specific and proprietary algorithms to rank available menus at the search point. These rankings decide which options are shown to the customer. Although the information on specific ranking algorithms is not public, we know that the attribute that usually plays the most important role is the price. Given the robust findings that price negatively affects choice probability, the best guess for a simple ranking mechanism that would capture a wide variety of sorting mechanisms would be options sorted in decreasing order with respect to price. Assuming that each user was reached by the same number of options, we could construct a reduced dataset for a sensitivity check. For this exercise, I construct a dataset that only contains menus with more than 20 options, and only retain 20 cheapest alternatives per menu. There are also cases where the chosen option is not part of the set of 20 cheapest alternatives in the menu ¹³. I also eliminate these choice cases from the reduced dataset. This leaves me with about four thousand choice cases.

¹³20 alternatives are chosen so that there are enough entries to have variance in key variables. Results are robust to different menu sizes, with the characteristics that, as I reduce menu size, more effects seem to lose significance

Panel A	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Attraction	0.0138		0.0191		0.0010		
	[0.512]		[0.362]		[0.317]		
Compromise		-0.5507		-0.5239		-0.5335	
		[< 0.001]		[< 0.001]		[< 0.001]	
Similarity			-0.1916	-0.1890	-0.2425	-0.1859	-0.2375
			[< 0.001]	[< 0.001]	[< 0.001]	[< 0.001]	[< 0.001]
Attraction within cluster					0.1357		0.1329
					[0.001]		[0.001]
Attraction outside cluster					-0.0111		-0.0080
					[0.628]		[0.728]
Compromise within cluster						-1.5127	-1.4961
						[< 0.001]	[< 0.001]
Compromise outside cluster						-0.2374	-0.2566
						[0.107]	[0.083]
<hr/>							
Panel B	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Attraction	0.0359		0.0291		0.0096		
	[< 0.001]		[< 0.001]		[< 0.001]		
Compromise		-0.4150		-0.3482		-0.3112	
		[< 0.001]		[< 0.001]		[< 0.001]	
Similarity			-0.2037	-0.2082	-0.3084	-0.2038	-0.3034
			[< 0.001]	[< 0.001]	[< 0.001]	[< 0.001]	[< 0.001]
Attraction within cluster					0.2454		0.2436
					[< 0.001]		[< 0.001]
Attraction outside cluster					0.0177		0.0182
					[< 0.001]		[0.006]
Compromise within cluster						-1.5331	-1.3553
						[< 0.001]	[< 0.001]
Compromise outside cluster						-0.2851	-0.2655
						[< 0.001]	[< 0.001]

Table 11: Average marginal effects from the reduced dataset (Panel A, top) and marginal effects from the full dataset at menu size = 20 (Panel B, bottom). P-values in the square brackets.

Table 10 presents the estimation results of all 9 models, while panel A in table 11 presents the corresponding marginal effects. For the sake of comparison, panel B in table 11 also presents marginal effects implied by the models estimated on the original dataset at the menu size being equal to twenty. One difference due to the move from full to reduced dataset is that the total attraction effect seems to lose significance. However, once this effect is decomposed along the borders traced by the similarity measure in model 9, it is clear that the in-cluster attraction effect does attain statistical significance. Naturally, there are differences in terms of the size of marginal effects from the original and reduced datasets that one can read by comparing panels A and B in table 11. However, these differences are fairly small, and two exercises seem to produce consistent results.

Experimental data

The airfare choice data presents an excellent opportunity for the application of the proposed methodology. It is a large dataset of actual choices made in a natural environment by consumers, the product under question is relatively complex (i.e., characterized by more than two attributes), the menu size is not constant across different choice cases, and menu sizes are sufficiently large to get sufficient variance in all context variables. However, the dataset also has shortcomings. Firstly, although one knows what was available on the market when the choice was made, there exists no accurate information on which of the options was in fact considered by the consumer. Secondly, there is no information on the identity and characteristics of the consumers. Without such information, one is not able to account consumer-side features that could systematically drive choice outcomes that is observed.

To remedy these shortcomings, in what follows, the same methodology is applied to an experimental dataset. These data, like the observational dataset, come from a travel context. Similarly, the product studied is relatively complex. Unlike an observational dataset, however, the dataset comes from a stated choice experiment. Here, there is no variance in menu sizes, and these menus are relatively small (five alternatives). Importantly, this dataset contains information on a set of variables that describe the demographics of the subject. An added advantage of the dataset is that each subject makes 12 choices and these 12 choice cases are constant across all subjects. This allows to control for menu-specific, as well as subject-specific characteristics.

The data comes from a discrete choice experiment administered to residents and daily commuters to the city of Ljubljana, Slovenia, by Gerzinic et al. (2021). 108 subjects were sequentially presented with 12 5-alternative menus and asked to choose the best alternative in each case¹⁴. This represents 1,296 recorded choice cases. Each alternative described a commuter trip with a 'park and ride facility choice' to the city with respect to the following five characteristics: price, car ride duration, public transport ride duration, public transport (average) wait time, and the mode of public transport (either bus or train). Subjects were overwhelmingly Slovenian nationals (91.67%), 58% female, with a mean age of 36 years (*St.Dev.* = 12.5). Further information on education, income, household size, and the number of cars in the household was also obtained. The descriptive statistics of the choice variables in the experimental dataset is presented in table 12. The characteristics of the subjects are used as control variables.

Variable	Count	Mean	St.Dev	Min	Max
Price	6480	5	3.266	1	9
Car ride duration	6480	15	8.166	5	25
Public transport ride duration	6480	20	8.166	10	30
Public transport wait time	6480	16.67	10.275	5	30
1[Public transport is train]	6480	0.5	0.500	0	1
Attraction	6480	0.067	0.249	0	1
Similarity	6480	2.633	1.080	1	4

Table 12: Descriptive statistics for choice and context variables.

Measurement of context effects

A significant disadvantage of this particular dataset is the small menu size (five) and the large number of choice variables (also five). These circumstances, together with the fact that the experiment was not designed for specific purpose of studying context effects and that one of the choice variables is categorical (mode of public transport), restricts

¹⁴In fact, the experiment consisted of multiple choice rounds within each menu. In the first round, subjects needed to choose the best option of the five presented alternatives. Consequently, this alternative was removed from the menu, and in the second round they needed to choose the worst option out of four remaining alternatives. Then this option was removed, and subjects needed to choose the best of the three remaining and ultimately the worst of the two remaining options. The experiment was designed for a different purpose; see Gerzinic et al. (2021). For the purpose of this study, only data from the first round of choices are used.

possibility of variance in context variables. So much so that the procedure applied to observational data does not identify a single case with these data where we can observe a compromise option. As a result, there is no variance in the compromise measure. In addition, the presence of categorical variable drives the AP clustering algorithm (as well as K-Means algorithm, for that matter), which always results in two groups of similar options; one consisting of all options using bus as the mean of public transport, and the other using train. Given that calculating the dominance relationship also requires constancy of categorical variable across a pair of options, all dominance relationships (and hence all attraction) are only within the cluster. The consequence of all this is that I cannot estimate model 4, and that models 6 through 9 become equivalent. Table [12](#) also presents descriptive statistics of the context variables.

Choice modeling

The choice modeling exercise takes a very similar approach to that using observational data. The only difference is that in the current case regressions also include menu-level fixed effects. This is necessary as the 12 choice cases are constant across all subjects. The results of the random effects Probit estimation are given in table [13](#) ¹⁵. At the level of choice variable, it is not surprising that one finds negative effects of all vertical variables (price, duration of car ride, duration of public transport ride, time of waiting for public transport). I also find that commuters prefer the train over the bus as a mode of public transportation.

The results with respect to context variables are consistent with those obtained from the observational data set in that significant attraction and similarity effects are observed. Although in model 3, the attraction effect is insignificant (and goes in the “wrong” direction), in the unifying model of context effects it achieves statistical significance ($p = 0.021$). Marginal effects indicate a much stronger impact of the context in the experimental setup. A one-unit increase in attraction results in 4.8% increase in the probability of choice of an option. An increase in similarity of one unit, on the other hand, decreases the probability of choice by 3.2%, *ceteris paribus*. Such a difference in marginal effects is not surprising given the much smaller menu size compared to the observational

¹⁵Robustness checks with random effects Logit, and fixed effects Probit with experimental dataset are presented in Appendix [E](#).

dataset.

Variable	Model 1	Model 2	Model 3	Model 5	Models 6-9
Price	-0.221*** (0.008)	-0.228*** (0.008)	-0.228*** (0.008)	-0.214*** (0.008)	-0.208*** (0.009)
Car ride duration	-0.067*** (0.003)	-0.070*** (0.003)	-0.070*** (0.003)	-0.069*** (0.003)	-0.066*** (0.003)
Public transport ride duration	-0.046*** (0.002)	-0.047*** (0.003)	-0.047*** (0.003)	-0.041*** (0.002)	-0.039*** (0.003)
Public transport wait time	-0.048*** (0.003)	-0.050*** (0.003)	-0.050*** (0.003)	-0.043*** (0.003)	-0.040*** (0.003)
1[Public transport is train]	0.189*** (0.042)	0.179*** (0.043)	0.179*** (0.043)	0.147*** (0.044)	0.140*** (0.044)
Attraction			-0.005 (0.089)		0.226** (0.098)
Similarity				-0.122*** (0.025)	-0.149*** (0.027)
Constant included	YES	YES	YES	YES	YES
Control variables included	NO	YES	YES	YES	YES
Number of observations	6480	6180	6180	6180	6180
Number of choices	1296	1236	1236	1236	1236
Number of subjects	108	103	103	103	103
Consistent Akaike information criterion	5206	5105	5113	5089	5092
Log likelihood	-2537	-2386	-2386	-2374	-2371

Table 13: Choice model estimation with experimental data.

Note: Five of the subjects did not provide information on all control variables (i.e., demographics). As a result, the data from those subjects is missing in regressions with control variables. Standard errors in parentheses. Statistical significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

3.5 Conclusion

This chapter has built on findings described in chapter 2 and further extended established measures of context effects with the aim of making contextual choice modeling applicable to observational data from a wide range of environments. Previous approaches to the study of context effects were limited to choices among a small number of options (usually two or three) of simple products (usually characterized by two numerical attributes). This meant that much of the previous literature used experimental settings and examined relatively simple product line design questions, e.g., Orhun et al. (2009) and Rooderkerk, Van Heerde, and Bijmolt (2011), where the choice was among a set of similar products with marginally varied characteristics.

One important limitation of this approach is that (in case of attraction and compromise effects), it relies on ordinal relationships between the available options. In other words, the information that option A is better than option B in terms of attribute X is used. This is a departure from previous literature that has relied on cardinal measurements (e.g., Rooderkerk, Van Heerde and Bijmolt (2011)), that is, how much one thing differs from another. When dealing with only two or three options, cardinal measurement is the only way to quantify context effects. In the proposed framework, on the other hand, cardinal measurement is not a necessity. However, given that cardinal measurement is possible in most contexts (at least for a set of attributes), using ordinal measurements may hide some information from the modeling approach. One could imagine developing context effect measures that will consider both ordinal and cardinal measures. In this way, for example, a researcher could measure the attraction effect not only based on how many alternatives a focal option has, which is an ordinal measure, but also by how much the focal option dominates those alternatives (at least on average); which is a cardinal measure. In a similar vein, information on how central the focal option is in the group of alternatives could be incorporated in the measure of the compromise effect, and information on the spread of options identified as similar to the focal option could be incorporated in the measure of similarity effect. Moreover, clustering has proved itself being a reliable method to account for similarity effect.

This framework opens up possibilities for studying context effects in a much wider range of settings. The presented methodology can handle very large menu sizes of products

that are relatively complex (i.e., characterized by many and varied types of attributes). Equally importantly, the approach does not require consistent menu sizes in the dataset. In the paper, only one such empirical application is presented. However, this methodology could be applied to larger and richer datasets generated from electronic commerce websites, as well as offline environments. This enables the study of context effects in the field without explicit experimental interventions, which usually prove to be very expensive. As a result, the methodology could contribute to designing optimal offers on larger scale (e.g., optimizing recommender systems) especially in online settings where each shop outlet can carry hundreds of substitute options to choose from. Considering that in an increasingly crowded digital marketplace, online portals often leverage recommender systems to assist consumers in the decision-making process. Sometimes, even the number of alternatives recommended may seem overwhelmingly large. Prior research has established that when faced with many alternatives, people tend to use decision-heuristics (Fishburn, 1974b). The findings of this study can be utilized in an attempt to design consideration sets that encompass empirically similar alternatives. This might especially be useful when recommender systems have very little information about the user, which are known as cold-start problem.

The results of this study provide a good foundation for the next chapter of my thesis. There, I will utilize cluster based similarity measure to propose a recommender system design which can especially be useful in user information scarce environments.

4 A Context-Informed Approach to Cold-Start Problem in Recommender Systems¹⁶.

Abstract

The cold-start problem has become a significant challenge in recommender systems. To solve this problem, most approaches use various user-side data and combine them with item-side information in their system design. However, when such user data are not available, those methods become unfeasible. We provide a novel recommender system design approach, which is based on two-stage decision heuristics. Using only the characteristics on the item side, I first identify the structure of the final choice set and then generate it using stochastic and deterministic approaches.

4.1 Introduction

With the rise of the Internet, interaction with recommender systems has become a common part of human activity. When there are many options to choose from, recommender systems save consumers time and effort by matching them with items (Bobadilla et al., 2013). Making successful recommendations requires knowledge of demand-side factors, such as consumer taste, historical interactions, purchasing power, and sociodemographic characteristics, along with supply-side factors, such as item characteristics.

Because recommender systems are online services implemented by providers, supply-side information is generally available at all times. However, this is not always the case with demand-side information. On most occasions, users are not identified either because it is not feasible or because interaction with the system does not require them to identify themselves.

This lack of information is referred to as the cold-start problem (Adomavicius and Tuzhilin, 2005). Some services, for example Netflix, solve this problem by providing general suggestions until they can gather enough information about the user. Others, such as Goodreads, explicitly survey the new user to seek this information.

¹⁶This chapter is based on a joint work with my supervisor Zakaria Babutsidze, William Rand and Thierry Delahaye. It has been published in the proceedings of the 54th Hawaii International Conference on System Sciences

With current regulations and user awareness of security and privacy on the internet (Antón et al., 2010), systems face continuous cold-start problems (Wong et al., 2014). In such cases, the use of contextual information from the supply side becomes crucial (Adomavicius and Tuzhilin, 2005). One way of using this information is to use random utility models. However, such models are based on the notion that perfectly rational consumers have well-defined preferences (Babutsidze et al., 2019). Hence, they are not able to account for context-dependent preferences (Tversky and Sattath, 1979). Under such circumstances, clustering-based approaches can be used to enrich the context in recommender systems. Previous research (Babutsidze et al., 2019) has demonstrated that such an approach is flexible enough to be extended over imperfectly rational or context-dependent consumer behavior.

In this paper, I incorporate insights from the decision literature into recommender system design using the dataset of European flight choices (Lh eritier et al., 2019). I argue that the choice process occurs in two sequential stages: consumers first identify the small subset of choices that they would “consider” and then make a choice from that subset. Combining findings in marketing, management, and consumer behavior, and using clustering to quantify the contextual information of the choice set, I propose a user-side, two-step “consider-then-choose” (Gilbride and Allenby, 2004; Liu and Arora, 2011) approach to recommender system design to tackle the cold-start problem.

4.2 Theoretical background

Choice heuristics

Previous literature in psychology and economics has suggested that people tend to use various decision heuristics to reduce cognitive load during the decision-making process (Fishburn, 1974a; Bettman, 1979; Johnson et al., 1989). Because consumers tend to behave as *satisficers* rather than *maximizers*, they do not perform an evaluation of all the alternatives available to them, but stop when they find an option that has overall better attributes and satisfies their needs (Simon, 1956). For simplicity, let us consider the case of buying flight tickets. There are N tickets and each have k attributes, which can include the price, duration, time of the day of flight, number of connections, and others. A consumer has a single objective function

$$O = O(T_1, T_2, \dots, T_N), \quad (12)$$

where T is the linear transformation of the attributes of the flights Z^k with some random parameters θ . We can rewrite equation 12 as

$$O = O\left(\sum_k \theta_k Z_1^k, \dots, \sum_k \theta_k Z_j^k, \dots, \sum_k \theta_k Z_N^k\right), \quad (13)$$

which would allow a consumer to explicitly compare the marginal contribution of each attribute to the maximization of the objective function (de Dios Ortúzar and Willumsen, 2011). However, most of the time, choice attributes do not require or allow for such trade-off calculations, as some of them are too valuable or there are too many attributes to consider. For example, a consumer flying for business purposes may value the flight duration more than a budget traveler, who would value price above everything else. In such cases, even when consumers are perfectly rational and well informed, when faced with options that simultaneously differ across many attributes, or when it is difficult to calculate such trade-offs, they use various heuristic approaches (Hauser and Wernerfelt, 1990).

Heuristics are mathematical formulas that describe different rule-based decision steps taken by individuals to reduce their potential decision effort (Bettman et al., 1998). One can distinguish several types of approaches based on heuristics: lexicographic rule (Fishburn, 1974a), conjunctive/disjunctive (Coombs, 1951), elimination by aspects (Tversky, 1972), to name a few.

The lexicographic rule is the simplest deterministic rule in the heuristic approach. Here, individuals choose the alternative that has the highest value of the feature they want. If there are several options with equal values, individuals compare those options based on the second most valued feature. This loop continues until there is one option remaining. For example, a person looking for flight tickets from Paris to New York will have different options that vary in time, price, number of connections, baggage allowance, transfer time, and so on. Attributes of the choices are first ranked based on their importance to the consumer: cheaper than 600 euros, checked and carry-on baggage included, one layover, maximum transfer time of four hours. Then, a filtering stage occurs. After filtering on

the price, if there are multiple options remaining, the consumer will switch to baggage allowance, connections, and transfer time. As soon as the choice set contains only one option, the search stops.

The conjunctive and disjunctive heuristic approaches are related (Coombs, 1951). In the conjunctive rule, consumers first establish the list of features they consider relevant to the choice problem. Then, they establish various thresholds on those features. If an alternative passes all those thresholds, it is chosen. In contrast, in the disjunctive approach, an option is chosen that exceeds the threshold in at least one of the features (Coombs, 1951). The results of the lexicographic approach and the conjunctive approach appear to be similar. The only difference is that, instead of evaluating options based on the first aspect, then the second, and so on, in the conjunctive approach, the consumer evaluates options based on all aspects simultaneously. In some cases, consumers may be willing to use a conjunctive approach that generalizes both conjunctive and disjunctive approaches (Hauser, 2014). It allows some variation in the desired aspects. For example, if the consumer valued 4 aspects as mentioned in the example above, he or she might be willing to accept an option which satisfies 3 of those 4 aspects. This approach is particularly useful when there are time constraints or a fully conjunctive rule would result in no choice (Hauser et al., 2009).

Elimination by aspects is another heuristic approach that has been proposed in the literature (Tversky, 1972). The basic setup of this approach is that an individual chooses one attribute and eliminates options based on this attribute and repeats this procedure for other attributes if necessary until the remaining options do not share common attributes anymore. Then, as a last step, the final option is chosen according to Luce's choice axiom (Luce, 2012), which states that the probability of selecting one option over the others in a choice set is not affected by the presence or absence of other options. Most of the results on this topic (Batsell and Polking, 1985; Gensch, 1987; Currim et al., 1988; Manrai and Sinha, 1989) indicate that the use of multiphase heuristic processes can increase the accuracy of the estimation and result in improved interpretability of the models. Elimination by aspects is considered a heuristic method with stochastic rules due to the nature of the comparisons made by an individual and because the selection process is not based on the relative importance of the features (Aribarg et al., 2018). Such models are also hard

to apply successfully because they require a tremendous number of parameters to be estimated (Batsell and Polking, 1985). Although these models can theoretically capture the essence of the two-stage choice process, they are unable to identify the results of separate stages (Gilbride and Allenby, 2004).

Two-stage choice

During the choice process, consumers usually face a large number of options (Payne et al., 1988). Evaluating that many options drastically increases cognitive load during the decision process and, therefore, to reduce this load, consumers first select a small subset during an initial consideration stage (Paulssen and Bagozzi, 2005) and then make their choice from that subset in the final stage (Bettman, 1979; Gensch, 1987; Paulssen and Bagozzi, 2005). First, this allows users to remove unrealistic options from thorough consideration. Second, because the set of choices is much smaller in the final stage, users are able to invest more cognitive effort to analyze individual options with more care (Gensch, 1987). In addition, the decision strategies used in the two stages differ considerably and are therefore not interchangeable. The main reason for this is that the cognitive costs of the decision rules should not outweigh their potential benefits during each stage (Bettman et al., 1990).

In the information processing literature, the small subsets from which consumers make their final decisions are called consideration sets. There are several definitions of a consideration set. (Shocker et al., 1991) defines a consideration set as a “set of alternatives that satisfy a specific objective and are accessible to a consumer on a particular occasion”. Hauser (1990) refers to it as a “set of options that receive a significant amount of consideration during the decision-making process”. However, in marketing, scholars generalize these definitions and refer to consideration sets as a “subset of alternatives that survive the initial screening phase” (Häubl and Trifts, 2000).

Despite the fact that consumers may not always use such a two-stage process to select products (Hauser et al., 2009), the use of consideration sets is justified because they represent the choice process more realistically and better explain consumer behavior (Horowitz and Louviere, 1995). Potentially up to 80% of the decision process, the uncertainty can be resolved if one correctly determines the consideration set (Hauser, 1978).

For an empirical study of consideration set formation, information on consideration

sets can be elicited in multiple ways (Gaskin et al., 2007; Yee et al., 2007; Ding et al., 2011). However, for modeling purposes, the literature discusses two main ways of consideration set formation: deterministic (Coombs, 1951) and stochastic (McFadden et al., 1973; Urban et al., 1984). Although stochastic modeling makes all potential sets possible by attaching nonzero choice probability to each of them, deterministic approaches may render some outcomes impossible (Aribarg et al., 2018). Because we cannot know which answer is the best and the decision maker, or consumer, is the final arbitrator of the “correct” choice (Hauser, 2014), the use of either of these two approaches must consider the environment of choice, time frame, future value (or loss) associated with the correct and incorrect choice, and so on (Punj and Moore, 2009). For example, let us consider the flight booking case again and suppose that the consumer lives far from the airport and can reach it only in the afternoon. Consequently, all tickets with departure time before noon will not be considered. When forming a consideration set in this case, one must consider not only the characteristics of available options, but also the characteristics of the consumer and the choice environment. While using a purely stochastic approach might yield sets which include some options that the consumer would indeed consider, there will also be options which will have zero probability for consideration. In contrast, applying some deterministic rules derived from this particular choice environment, such as departure time, in the consideration set formation, will completely exclude those options.

When there are not many options to consider, the options have few attributes, or the utility of the final choice is not evenly distributed among the attributes, the consideration sets can be modeled using simple deterministic rules, because there is not much cognitive load and the decision rules are relatively simple (Lee and Lee, 2004; Hauser, 2014). When forming consideration sets, it is also important to consider their size. It is very difficult to decide on an optimal size based on the choice environment and individual processing capabilities (de Dios Ortúzar and Willumsen, 2011).

With the current progress in computer science, mathematics, and behavioral economics, recommender systems are ideal tools to solve this information overload problem and provide users with the most relevant consideration sets (Breese et al., 2013).

Recommender systems

Recommender systems (RS) have been an important part of our daily lives thanks to

the rise of the Internet. RS are software tools and / or algorithms that match users to items (Mahmood and Ricci, 2009)¹⁷. An example is Netflix, which recommends a movie similar to the one the user just watched. The general purpose of any RS is to help users who do not have sufficient knowledge or experience or the capacity to evaluate the item pool fully.

I distinguish between personalized and general recommender systems. Personalized RS may suggest different items to different users or user groups. General RS, in turn, are usually directed towards the general public and might be relevant only to some part of it, for example, Billboard Hot 100, IMDB Top 250, or the front page of the New York Times (Ricci et al., 2010).

When RS faces new users or new items, it may not provide personalized content due to the sparsity of information (Lika et al., 2014). Because such RS mainly utilize historical interactions of similar users on similar items, and their ratings, facing a new entity about which it has no information makes it impossible to generate recommendations. This problem is referred to in the literature as the cold-start problem (Adomavicius and Tuzhilin, 2005) and is considered a key challenge in the design of RS (Park and Chu, 2009).

The literature distinguishes three main cold-start settings (Park and Chu, 2009): a) recommending existing items for new users (user-side), b) recommending new items for existing users (item-side), c) recommending new items for new users (user- and item-side). However, when trying to address this problem, scholars have focused mainly on settings in which the challenge was to recommend new items to existing users (Zhang et al., 2010).

Recently, some progress has been made in solving the user-side cold-start problem after the introduction of contextual information into recommender systems. As a result of this effort, context-aware recommender systems were introduced (Adomavicius and Tuzhilin, 2011). In this approach, the context refers to the time and content of the choice, the location or sociodemographic characteristics of the decision-maker, and so on. Some approaches have been very successful by combining contextual information with

¹⁷Although there are other tools which create suggestions for users (such as Interactive Decision Aids, Recommender agents and etc.), for the purposes of this study I refer to recommender systems and bound the definition using the one by Mahmood (2009). These other tools have some theoretical and implementation differences, which makes them orthogonal to the purposes of this study. However, I discuss, differentiate between them, and define the boundary conditions in the chapter 5.2.

collaborative filtering (Aharon et al., 2013; Bykau et al., 2013; Saveski and Mantrach, 2014). The use of baseline information for new users (Kluver and Konstan, 2014) and the use of data from social networks (Guy et al., 2009) have also been proven to overcome the cold-start problem to a some extent.

In practice, however, these cold-start problems often transform into continuous cold-start problems (Kiseleva et al., 2016). This happens when:

1. The user stays “inactive” for a long period before the initial interaction
2. The user’s interactions have a significant time window
3. The user creates a “one-time” account
4. It is not possible or permitted to track users, or (under GDPR) the user has requested their personal information to be removed from the system (Hildebrandt, 2022).

In the case of the continuous cold-start problem, the solutions suggested in the literature discussed above are not feasible. The first reason is that users generally do not need to create an account to interact with some services, for example, watching videos on YouTube, searching for items on Amazon, or looking for airline tickets. Because of this, systems commonly treat different sessions by the same user as new users. The second reason is that due to increasing awareness of internet security and privacy, people tend to use incognito mode when they make searches (Antón et al., 2010), which disables most of the tracking and user identification.

My approach addresses the user-side continuous cold-start problem, which has not been thoroughly researched before. By utilizing only characteristics at the item and search level, I propose a novel RS design which is able to tackle the information sparsity. First, I use clustering (Rokach and Maimon, 2005) to quantify contextual information both at the individual level and at the search level, and I cluster empirically similar items together. Then, a hypergeometric sampling technique is used to generate the structure of the final choice set, meaning how many options from each cluster should be in the final choice set. Because my goal in this study is to provide the RS design that is not aimed at providing accurate recommendations per se, the final stage of the choice set generation

will consist of applying both stochastic (McFadden et al., 1973; Urban et al., 1984) and deterministic rules (Hauser, 2014; Lee and Lee, 2004; Coombs, 1951).

4.3 Methodology

Observational data I use are the same as that used in the previous chapter. A detailed discussion of this dataset can be found in section 2.5. These data have been subject to preprocessing rules which were also described in section 2.5 and in section 3.4. The descriptive statistics of the vertical variables are shown in the table 1.

Clustering

Clustering is used to divide the data into different groups where empirically similar elements belong to the same group and dissimilar ones are assigned to different groups (Rokach and Maimon, 2005). By using clustering, my aim was to identify options that were similar in their context. I used two mainstream clustering algorithms: Affinity propagation (AP) and KMeans (KM). Both clustering methods are described in detail in the Appendix C.

To quantify the context and clusters, I created two variables that captured the characteristics of clusters: relative cluster size and relative cluster dispersion. The first represents the normalized number of options within that cluster. The second is derived using

$$\frac{\sum_{i=1}^m (x_i - \mu_k)^2}{\sum_{i=1}^N (x_i - \mu_M)^2},$$

where N is the number of options within the menu, m is the number of options within the cluster, μ_k is the centroid of the cluster k to which x_i belongs and μ_M is the center of mass of the menu.

Two-stage choice

The modeling process consisted of two stages. In the first stage, I modeled the structure of the final choice set and determined how many elements of each cluster should be present in the consideration set. Next, I used simple stochastic and deterministic rules to select options following the structure obtained during the first stage.

First, the attractiveness measure of clusters within the menu was calculated. I defined the attractiveness measure as the probability that the given cluster contains an actual

choice and calculated it using the traditional multivariate logistic model (Ben-Akiva et al., 1985). Utilizing both descriptive information of options within clusters and the aforementioned cluster-level characteristics as covariates, I estimated those probabilities for every cluster in the menu according to

$$a_k = Pr(Y = 1|X_k) = \frac{\exp(\beta X_k)}{1 + \exp(\beta X_k)}, \quad (14)$$

where a_k is the attractiveness measure, X_k is the feature vector of the cluster k and β is a vector of coefficients.

Using cluster-level characteristics allowed me to embed the contextual information of options within a cluster into my model. Then, using this metric, the structure of the final choice set was determined via hypergeometric sampling.

Let N be the number of options within the menu that belong to k unique clusters and $m_i \in M$ be the number of options that belong to cluster i , so that $\sum_{i=1}^k m_i = N$. If we sample n random options from that menu without replacement, we get a set $J = \{j_1, j_2, j_3, \dots, j_k\}$ which follows the hypergeometric distribution, and the probability of getting such a vector J is determined by

$$P(j_1, j_2, \dots, j_k) = P(J) = \frac{\binom{m_1}{j_1} \binom{m_2}{j_2} \dots \binom{m_k}{j_k}}{\binom{N}{n}}, \quad (15)$$

where j_k is the number of elements belonging to cluster k in our sample.

However, using M and N does not allow one to quantify the menu context in terms of its clusters, which was the goal. One way to avoid this limitation is to use the attractiveness measure instead of M in sampling. However, because the attractiveness measures are in the range of zero to one, it was impossible to use them directly in our sampling. So, I define the *attractiveness score* of a cluster as $s_k = a_k * 1e6$, where a_k is the attractiveness measure of a cluster k . The constant $1e6$ was chosen to account for the smallest differences between two almost identical a_k . Consequently, N was replaced by $D = \sum_{i=1}^k s_i$. Therefore, equation 15 became

$$P(j_1, j_2, \dots, j_k) = P(J) = \frac{\binom{s_1}{j_1} \binom{s_2}{j_2} \dots \binom{s_k}{j_k}}{\binom{D}{n}}. \quad (16)$$

To save computational time and overcome the sparsity of the vector J , during sampling I used only attractiveness scores of the top n most probable clusters. Because n was assumed to be relatively small, I was able to define all possible vectors J in advance such that $j_1 \geq j_2 \geq j_3 \dots \geq j_k; k \in n$ using integer partitioning.

In order to make the sampling results also dependent on M I used M as a constraint for J , so that

$$\forall j \in J, m \in M : j_i \leq m_i.$$

If this condition could not be satisfied for some i , then I did the assignment $j_i \leftarrow m_i$ and the remainder $j_i - m_i$ was added to the leftmost possible element of J . However, for 108 menus, it was still the case that there was no valid J that complied with these rules. I simply removed those menus from the analysis.

To better understand the approach, let J be $[4, 3, 2, 1, \dots, 0]$ and M be $[8, 2, 4, 1, \dots, 6]$. Then, $j_2 \geq m_2$, which violates the above constraint. So, the assignment $j_2 \leftarrow 2$ is made and the remainder 1 is added to the leftmost possible element of J . The final result becomes $[5, 2, 2, 1, \dots, 0]$.

Finally, by randomly sampling according to equation [16](#) one hundred thousand times, I picked our most likely J by finding the most repeated sample. Then, by selecting the top j_1, j_2, \dots, j_k options from the top n most probable clusters based on the attractiveness score of the options obtained using equation [14](#).

After identifying the clusters and the number of elements to select from, two methods were applied to generate the final choice set. The first method was stochastic and consisted of randomly selecting elements according to the vector J . The second method was deterministic and used the price of the option as a determinant. The cheapest options were selected according to J . As a baseline, I used the same two approaches, but the selection was made without regard to J . Therefore, the baseline of the first method was the random selection of one option from every cluster. The baseline of the second method was the selection of the cheapest option from every cluster. Recall that there were two different clustering methods, AP and KM, used. Hence, I applied four models per clustering method:

- *Model1*. Random selection following J
- *Model1b*. Random selection (baseline of model one)
- *Model2*. Selection of the cheapest options following J
- *Model2b*. Selection of the cheapest options (baseline of model two)

Performance metrics

To evaluate each model’s performance, I used accuracy at top-N, which is a commonly used metric not only in classification tasks but also in RS design studies, especially for context-based recommendations (Ricci et al., 2010). In classification, it measures whether the actual class is in the top N predicted classes of the model. Similarly, in the RS design it measures if the chosen option is among the top N suggestions of the system. I complied with the existing literature and selected the accuracy in the top-5 and top-10 as our evaluation metrics (Cremonesi et al. (2010)).

Consequently, $n = 5$ and $n = 10$ were chosen. Therefore, all possible J values were found via integer partitioning of five and ten, which gave us seven and 42 possible variations accordingly.

4.4 Results

Clustering results

The results of the clustering methods are clearly different. While AP tended to create fewer but larger clusters (7.62 on average), KM generally identified more clusters (10.29 on average) with relatively smaller sizes. This indicates that both algorithms were able to identify the contextual information but in different ways.

The runtimes of these algorithms also differed considerably. Because AP did not need an initial number of clusters, while for KM we had to compute the optimal cluster count in each menu, for the same menu, AP converged on average 7.2 times faster. This makes AP more viable for larger choice spaces.

First stage results

Table 14 gives descriptive information about the structure of the consideration sets for the different clustering methods. We notice the similarities between AP and KM in

terms of the average number of clusters present in the choice sets. Despite the different contexts identified by those algorithms in the clustering phase, both algorithms appeared to identify the “important” clusters. One can also see that KM resulted in more variance, yet generated less diverse consideration sets in general. In contrast, AP appeared to be more robust when it came to different configurations of choice environment and was able to generate consideration sets that were more distinct.

	$n = 5$		$n = 10$	
	AP	KM	AP	KM
Mean	2.74	2.59	3.62	3.41
Standard deviation	0.84	1.16	0.98	1.61
Minimum	1	1	1	1
Maximum	5	5	7	10

Table 14: Consideration set structure as unique clusters across clustering methods.

Second stage results

One can see that both clustering methods were also robust to the selection methods used in the second stage. This indicates that item-side contextual information helps capture the choice environment better and also provides meaningful insights into the consumer behavior.

Both models considerably outperformed their baseline counterparts. Stochastic models performed in general better in KM than in AP, which is not surprising. The main reason for this is that KM identified smaller clusters, and therefore the chance of randomly selecting a correct option was higher. This difference decreased in cases where the selection was made based on deterministic rules.

The performance of models that use a deterministic rule to make the selections may indicate that consumers use multiple determinants as criteria during the decision-making process, which also complies with previous findings (Bettman, 1979; Lee and Lee, 2004). Table 15 summarizes the results from the second stage.

4.5 Conclusion

I have proposed a novel approach to tackling the user-side continuous cold-start problem in RS design. By using the contextual information of the menu, we were able to generate relevant choice sets using a two-step choice modeling approach. The structural approach

	$n = 5$		$n = 10$	
	AP	KM	AP	KM
Model 1	0.39	0.40	0.56	0.55
Model 1b	0.21	0.23	0.21	0.25
Model 2	0.49	0.48	0.63	0.62
Model 2b	0.32	0.32	0.32	0.34

Table 15: Top-5 and top-10 accuracy scores across clustering methods.

to choice set generation proved to be robust not only to selection criteria, be it stochastic or deterministic, but also to the clustering method used. Because in an online environment, the calculation time is critically important, using AP as the clustering method appears to be advantageous.

The findings of this work can be implemented by various systems that face continuous cold-start problems. They also help to understand the decision-making process of consumers, and hence reduce their search cost by introducing the most relevant alternatives. This also benefits the supply side via the reduction of the overall time spent by users on the platform.

This work has some limitations. RS design using simple one-stage MNL probabilities would result in 52% and 65% in top-5 and top-10 accuracy, respectively. Such random utility models violate Luce’s choice axiom (Luce, 1959), which states that the choice probabilities of the options in the choice set must be equally affected by the introduction or removal of a new option. However, one possible way to improve our approach could be to integrate those probabilities into models in this study. Another possible avenue for future research could be using more complex characteristics derived from the choice set along with clustering.

In this study, I demonstrated the importance of context effects derived from choice sets in shaping user decisions, as well as the feasibility of incorporating these effects into the design of recommender systems. Albeit enhancing the system’s ability to align its recommendations on the account of the user’s two-step decision-making process is a significant step forward, one must also recognize that context does not arise from the choice set alone. In previous chapters, I have used Tversky’s (1972) definition of context being “the composition and the nature of the choice set, and availability of various options in it”. A more recent stream of research has also established that the context of choice also

depends on user preferences, among other things (Dey, 2001; Adomavicius and Tuzhilin, 2011).

Research agrees that user preferences are not static and rather fluctuate according to individual-specific factors (Song et al., 2019). In light of this, it becomes clear that to help users choose, recommender systems must have tools that enable users to signal their preferences to them. Overall, this could be summarized as “help me help you” approach.

However, a significant challenge arises when investigating tools that can account for dynamic preferences, as the existing literature has studied it in specific domain settings. The lack of generalizability hinders the creation of more versatile recommender systems, which can adapt to varying contexts and user preferences.

To bridge this gap, my next study ventures into the realm of domain-neutral choice settings. By controlling for the choice context, I can focus on the investigation of tools that “let users help”, ultimately striving to contribute to the design of recommender systems that are responsive and adaptive. This argument creates strong motivation and positions my study in the next chapter as a necessary progression from this one.

5 User Control and Its Impact on Recommender Systems' Acceptance

Abstract

This study examines the influence of user control on recommender systems' acceptance in a context-independent experimental setting, using the Technology Acceptance Model as a theoretical framework. It confirms the original Technology Acceptance Model relationships, demonstrating that easy-to-use and useful recommender systems lead to higher user adoption rates. User control is found to be a crucial factor in explaining users' behavioral intention. The findings also reveal that different control methods have varying effects on user experiences, suggesting a need for dynamic user controls that align with user requirements. In conclusion, the study highlights the importance of user control in recommender systems and encourages further research into dynamic control mechanisms and more innovative approaches to increase user adoption.

5.1 Introduction

Internet has brought stores from all over the world to one's computer screen. It revolutionized e-commerce by connecting users with products. While making e-commerce widely accessible, it also created difficulties for users with a large number of products to choose from (Ricci et al., 2011). To mitigate this, recommender systems (RS) have been adopted (Kotkov et al., 2016). Recommender systems are specialized information filtering tools that aim to suggest relevant items to users (Adomavicius and Tuzhilin, 2005). The sheer amount of information present online make their existence essential, and in today's world, they have become omnipresent in our digital lives. When we try to choose the next Christmas gift, movie to watch over the weekend or a place to go during the next vacation, we use their recommendations to help us. RS are not only useful for consumers (or end users) to make a choice, but it is also essential for businesses as they offer more relevant items to users that otherwise would be undiscovered (Ricci et al., 2011). Using RS, businesses increase user interaction, have a strong effect on sales volumes and diversity (Song et al., 2019).

Recommendation generation can be categorized into three main categories: collaborative filtering, content-based filtering, and hybrid approaches (Burke, 2002). Collaborative filtering uses the historical interactions of users with items and defines similar users and similar items. Then, it suggests items to users based on the items similar users have consumed (Schafer et al., 1999). For example, Spotify would suggest a new rap song to the user based on the rap songs other users with similar profiles have listened to (Jacobson et al., 2016). Although quite useful, this approach has two major drawbacks: it assumes that user preferences are static and do not change; it relies heavily on historical user data (Wang and Blei, 2011). Content-based filtering is another technique which is widely used. It uses the properties of the items to decide on similar items and suggest them to users (Pazzani and Billsus, 2007). For example, Google News analyzes articles based on their content and then suggests similar articles to users (Das et al., 2007). This approach, too, while quite useful, suffers from drawbacks. Recommending only similar items to those already purchased, it may lead to the notion that users are exposed to only a limited circle of items, also commonly known as filter bubble (Tintarev et al., 2018). The hybrid approach, on the other hand, implied from its name, uses both content-based and collaborative techniques. Modern hybrid RS also utilizes more sophisticated deep learning-based approaches (Bahrainian et al., 2020). Netflix can be considered the most widely known hybrid RS. Because hybrid techniques are based on content-based and collaborative filtering, they have inherited the main drawback of those two techniques, namely, adapting to evolving preferences.

It was previously observed that users' preferences are prone to change over time. They become bored with the similarity of items they have consumed and suggested and want more personalisation (Song et al., 2019). This makes the RS adaptation to dynamic user preferences extremely important. To mitigate RS low ability to "evolve" alongside with the users' preferences, system engineers have introduced user control mechanisms to allow users to participate in personalisation.

User control is referred to as the "extent to which users can influence recommendation generation" (Jannach et al., 2019). It was defined as the ability of users to influence the recommendation process and its results (Knijnenburg et al., 2012). User control is necessary for several reasons. Firstly, it allows system designers to directly tackle one of

the biggest drawbacks of RS, namely, adapting to evolving user preferences. It is achieved by allowing users to provide initial preference information, explicit or implicit feedback, or modify the recommendation algorithm to better align it with their current interests (Tintarev and Masthoff, 2015). Secondly, it improves the overall user experience, allows users to explore various aspects of the recommendation space, and come across new and serendipitous items (Tintarev and Masthoff, 2015; Kotkov et al., 2016).

However, studies investigating the effects of user control mechanisms on RS usage reached differing conclusions. The study on user control preference in conference RS setting derived that participants did not treat recommendations from control enabled systems differently from RS without user control (Jameson and Schwarzkopf, 2002). The importance of user control to users varies among consumers depending on “the nature of the application, its adoption, and individual characteristics” and users were found to prefer simple control mechanisms to more sophisticated ones (Knijnenburg et al., 2011, 2012). The ability to control recommendation generation was found to positively affect user experience in music RS (Bostandjiev et al., 2012; Knijnenburg et al., 2012). Another study on different implementations of control mechanisms in music RS concluded that despite the positive correlation between the user satisfaction and usage of such mechanisms, user control does not always end with satisfaction. The authors also found that people who had below average domain interest were satisfied even with non-accurate recommendations (Hijikata et al., 2014).

Because of the complexity of RS and the nature of the studies, it is not surprising that most of the studies about user control in RS have involved either music, movie, or news RS. The main reason for that has been their availability. Those systems were already pre-trained, and one had to either download the data about user profiles, ask users to log-in with their existing accounts, or create a new account. This approach has several drawbacks. Firstly, relying on pre-trained systems and using existing user profile data can lead to selection bias. These systems were trained and designed only for use in a given domain, and hence conclusions of such studies may have limited applicability to other domains. Secondly, asking users to use their own accounts or create new ones can cause privacy concerns as users may be reluctant to share their personal preferences, which can result in self-selection bias, thus reducing the representativeness of the study (Bélanger

and Crossler, 2011). Thirdly, using existing RS would limit the ability of researchers to amend the specific features or algorithms of such systems, hence constraining the scope of the experimentation.

Another aspect to keep in mind is the domain knowledge. Users with a high domain knowledge generally tend to be more resistant to RS suggestions. Additionally, users with extensive domain knowledge may not be a good representation of the general user population and may tend to engage with RS in a way that the recommendations align with their pre-existing beliefs, leading to reduced diversity and “bubble” effects (Möller et al., 2018).

The limitations discussed above make the domain-neutral setting significantly important in RS studies. While it is nearly impossible to come up with a pure “domain-neutral” setting, it is still possible to create a setting which has eliminated the most of domain-specific constraints. It would allow for examining the behavior of users with varying expertise levels in RS and achieve a more comprehensive understanding of how users with heterogeneous expertise levels interact with RS. This will not only facilitate the development of more adaptive RS, but also allow the conclusions of studies to be more generalizable. Having a setting not tied to any domain would also allow researchers to fine tune different features of RS, its algorithm, which is hardly possible when using domain-specific RS. Lastly, the domain-neutral setting alleviates the impact of potential biases, such as confirmation bias, anchoring bias, and overconfidence, which are highly likely among users with high domain knowledge (Hijikata et al., 2012).

To the best of my knowledge, no other study has investigated user control and its effect on RS usage in a domain-neutral setting. I aim to fill this gap by designing an experimental setting which is not only domain-free, but also has an RS with no specific assumptions about users’ preferences. I used the Davis Technology Acceptance Model (TAM) which is widely used in studies regarding the usage of a specific technology (Davis, 1985) to test our hypotheses.

5.2 Definitions and scopes

The importance of precisely defining the concepts and constructs used in this study is crucial. Defining the key constructs not only establishes a common understanding, but

also provides boundary conditions and ensures the reliability of the research findings (Creswell, 2013).

By clearly defining RS one is able to distinguish among related concepts such as recommendation agents (RA) and interactive decision aid tools (IDA) which while sharing some commonalities, also entail some important differences in their goals, functionalities, and user interactions (Xiao and Benbasat, 2007). By specifying the characteristics of RS in our particular context, my aim is to ensure the relevance of the findings. This is also important considering the methodology and framework.

Similarly, defining user control is also important for understanding the various methods and tools the user can use to influence the recommendation generation process. Different conceptualizations and methods have been explored in the literature (Pu et al., 2012). Although these methods have a common end goal, they differ in complexity, effect time frame, and technical knowledge required to understand the causal links between recommendations generation manipulations and the outcomes. By specifying the exact definition of user control in this study, the objective is to ensure that the construct is measurable and generalizable.

Henceforth, in the following subsections, I discuss the definitions of recommender systems and user control in detail and justify our choices for this study.

Recommender systems

The emergence of systems which provide personalized or tailored recommendations is closely connected to the rise of e-commerce and Web-based technologies. In the literature such systems are described using different terms such as recommendation agents (RA), interactive decision aids (IDA) and RS. Thus, I will address the overlap between these terms and clarify the definition of RS used for this study.

RA has been conceptualized in the literature as “an interactive decision aid that helps consumers in the initial selection of alternatives that are available in an online store” (Häubl and Trifts, 2000; Xiao and Benbasat, 2007). They are predominantly implemented in e-Commerce, education, and organizational knowledge management contexts (Xiao and Benbasat, 2007). In the e-commerce setting, RAs are implemented in the initial product search phase on e-commerce websites, like Macy’s or Amazon’s laptop selection assistant shown in the figure 10. The user is asked about their preferences in a conversational,

filter/table style, or some other format, and then, based on the stated preferences, RA generates the “recommendations”. One distinct feature of RA is that it recommends options that satisfy all preferences stated by the user. If one option does not meet one of the preference requirements, it has zero chance of appearing in the recommended list (Häubl and Trifts, 2000). Another distinct feature of RA is that they consider the focal user’s preferences in “isolation” , which means other potentially relevant information, such as choice context, other characteristics of the user or the user group are left unused (Wang and Benbasat, 2008).



Figure 10: Amazon recommender agent for laptop selection.

Interactive decision aids, on the other hand, are a “web-based decision support system that elicits user preferences, performs a search on their behalf, and provides them with a subsequent product list” (Maes et al., 1999). They have a predefined set of decision rules which can give users a feeling of restriction in expressing their preference while interacting with the system and dissatisfaction with the overall process (Wang and Benbasat, 2009; Silver, 1988).

Recommender systems are tools that use information about users, their preference, historical user-item interactions, and contextual information (Adomavicius and Tuzhilin, 2005). Context here is referred to as “any information that describes the characteristics of the user or entity that is relevant to the interaction, e.g., a person, place, object, attributes, or the application itself” (Dey, 2001). While RS is conceptually similar to IDA and RA there are differences between them.

Firstly, RS covers all stages of users’ interaction with the system. Whereas RA mostly focus on the initial phase of the interaction. Hence, RS can adapt to changing or evolving user preferences and needs (Song et al., 2019). This broad scope allows RS to provide a more comprehensive and satisfying user experience. For example, Netflix’s RS regularly adapts its recommendation algorithms based on users viewing history, the genres they prefer, and interaction with the platform over time.

Secondly, RS offer more dynamic decision rules when compared to IDA. While both systems use similar ways of eliciting user preferences, RS has more flexible and dynamic rules in doing so. RS allows users to go back to previous elicitation stages and change their preferences. Such flexibility enhances the relevance of provided recommendations and leads to greater engagement. For example, Spotify uses RS that dynamically adjusts its algorithm based on users' listening habits, discovered artists, and changing moods ¹⁸.

Thirdly, RS use a more complex set of information to generate recommendations. Unlike RA which only considers the focal user's preferences, RS also utilizes the preference information of the user group it thinks the focal user belongs to. Also, unlike IDAs which predominantly use dialogue like elicitation, RS uses different combinations of elicitation methods. For example, in the same Spotify, one can use like dislike icons to state the preference for a given song, artist, and genre. In addition, they can generate recommendations based on the sole item within a playlist. Such interactions often lead to more serendipitous recommendations and greater overall satisfaction (Kotkov et al., 2016). All in all, for the scope of this study, I refer to online systems which have: a) dynamic decision rules, b) elicit user preferences in all stages of users' interaction with the system, and c) uses item features in recommendation generation as RS.

User control

User control is referred to as "the extent to which users have control over recommendation generation" (Knijnenburg et al., 2012). Depending on the interaction history of the user with RS, two main stages of user control can be distinguished: 1) control in the preference elicitation phase, 2) control after recommendation generation (Jannach et al., 2017).

The first stage commonly consists of preference forms, interactive conversation, and critiquing (Jannach et al., 2019). In the preference form, users are asked predefined questions about their search. The answers may come in the form of selection from drop-down list, user input, or slider-style input. Figure 11 shows an example of this method. Users can express their preferences clearly and directly. However, when the preference form is too long or too complex, users may disengage (Jannach et al., 2017). Moreover, it

¹⁸Here mood is not necessarily referring to user's personal mood, but to the mood of the songs.

might be difficult to interpret the underlying meaning of these preferences. For example, when users utilize slider-based inputs, it might not be obvious to them what the effect is of having all sliders in the middle, far left or far right (Jannach et al., 2017). Another disadvantage of this method is when the user preference does not exist in the form at all.

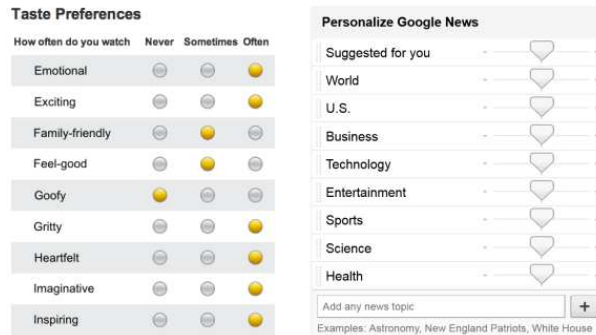


Figure 11: Preference form on movie and news RS. Sourced from from Jannach et. al (2017).

The interactive conversation, on the other hand, converts the preference building process into a multistage dialogue where the next steps are generated based on the responses to previous ones (He et al., 2016). It ensures that users provide feedback more naturally and less intrusive. This method is especially used in the so-called “cold-start” events, where RS has no previous knowledge about the user (Gu et al., 2019). However, the drawback of this method is that the results are not immediately visible and preference updating occurs only upon completion of the procedure. Also, if the user wants to change their previous answers, they have to restart that stage again. Designing a good dialogue-based system is not possible without taking into account the trade-off between the relevance of the dialogue and the cognitive requirements of the procedure (Gao et al., 2021).

In contrast to interactive conversation, when critiquing, users state their preferences directly on features of the products or on the products themselves, as seen in figure 12. It is a more straightforward way to provide focused feedback. However, users may not always be able to articulate their preferences this way, as they might have to review and critique multiple items. Moreover, it might not always be clear how critiquing a particular recommendation affects the algorithm. For example, in the figure 12 it might not be obvious to a user what the word “this” refers to in terms of algorithm change if

she selected “See fewer stories like this”.

The second stage of user control entails functionalities which allow users to influence recommendations after they are generated. Previous approaches have used sorting, feature- and weight-based filtering, and critiquing. (Swearingen and Sinha, 2001; Schafer et al., 2002; Bostandjiev et al., 2012; Schaffer et al., 2015; Jannach et al., 2017).

Considering the number of various tools available to achieve user control in different stages of user interaction with RS, it is important to define the scope of user control for this study. I follow the user control requirements from previous studies (Jannach et al., 2017). User control must be easy to understand, have an immediate effect on recommendations, and must not interfere with user’s interaction with the system. Hence, in this study, I will refer to critiquing and sorting as user control.

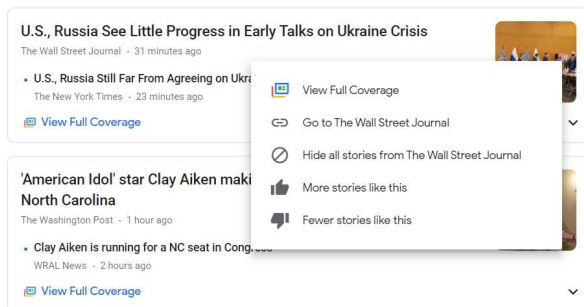


Figure 12: Critiquing on Google News. Sourced from Jannach et al. (2017).

5.3 Theoretical background and hypotheses development

It is essential to base this study on a theoretical framework. As mentioned above, I use the technology acceptance model of Davis (1985). Currently there are two dominant technology acceptance models used in the literature: Unified Theory of Acceptance and Use of Technology (UTAUT) and Technology Acceptance Model (TAM) (Davis, 1985; Venkatesh et al., 2003). Both models provide valuable insight into understanding the acceptance of information systems by users. In the following section, I will briefly discuss both TAM and UTAUT and argue why I have ultimately selected TAM for our study. Then hypotheses will be developed.

5.4 Models of technology acceptance

The continuous quest to understand the acceptance of a certain technology by users is an ongoing challenge in management studies (Schwarz and Chin, 2007; Williams et al., 2009). Rapidly increasing information system implementations, their crucial role in the modern business world, and their under-utilization problems have made the issue of acceptance of a technology central (Lancelot Miltgen et al., 2013). During the last few decades, increased interest of the research community in addressing this question has resulted in the development of two major theories and models of technology acceptance.

The Technology Acceptance Model (TAM), which was proposed by Davis (1985), is a widely accepted and influential model to understand and predict the acceptance and adoption of information systems by users. It was derived from the social-psychology based Theory of Reasoned Action and Theory of Reasoned Behavior that have explained individuals' behavioral intentions and actions. TAM simplifies and adapts these two models and assumes the mediating functions of “perceived ease of use” and “perceived usefulness” in explaining the relationship of the characteristics of the system and the actual use of the systems (Marangunić and Granić, 2015).

Perceived Usefulness (PU) - According to Davis, PU refers to “the degree to which a user believes that the use of a particular technology will improve his or her job performance or provide benefits” (Davis, 1985). The construct of PU is detrimental to TAM as it directly influences the intentions of users to adopt and use the technology. Empirical evidence has suggested that users are likely adopt a system “if they perceive it useful and capable of improving their performance” (Davis, 1985).

Perceived Ease of Use (PE) - It represents “the degree to which a user believes that the use of this specific technology will be easy to use” (Davis, 1985). That is, it captures the user's perception of the ease of the system's use. It not only directly impacts user's intentions to adopt a technology, but it also has an indirect effect on those intentions through PU, i.e. users may perceive easier to use systems more useful.

Due to its simplicity and versatility, after introduction, TAM has been used with a number of extensions and modifications to better suit a specific research context and include relevant constructs (Featherman and Pavlou, 2003; Amoako-Gyampah and Salam, 2004; Burton-Jones and Hubona, 2006). Subsequently, those modifications have led schol-

ars to build on TAM. TAM2 (Venkatesh, 2000) and TAM3 (Venkatesh and Bala, 2008), UTAUT (Venkatesh et al., 2003) have been introduced. A detailed review of TAM can be found in the study of Marangunic (2015).

UTAUT is another model introduced by Venkatesh which is a more complex and integrative model. It is a synthesis of eight established models, including TAM (Venkatesh et al., 2003). UTAUT considers four main constructs, performance expectancy, effort expectancy, social influence, and facilitating conditions as main determinants of behavioral intentions and usage behavior. The definitions of performance expectations and effort expectancy are heavily influenced by Tam's PE and PU. Social influence refers to the extent to which a user perceives that important others believe that they should use a specific technology. Facilitating conditions is the degree to which a user believes that an organizational and technical infrastructure exists to support the use of a technology. UTAUT also entails four moderators - age, gender, experience, and voluntariness as moderators. Like TAM, UTAUT has also been extensively used in various research contexts with different technologies, from communication (Wu et al., 2007), specialized business systems (Kijsanayotin et al., 2009), to general purpose technologies (Abu-Shanab et al., 2010). For a more comprehensive review, see the work of Williams (2015).

TAM was chosen as a theoretical framework for this study. TAM is a more suitable model than UTAUT for the current study because it focuses on the cognitive and affective aspects of user acceptance that are most relevant and measurable in this context. TAM is also a simpler and more parsimonious model than UTAUT, making it easier to operationalize and measure in an online experiment environment. TAM requires only two main factors to be manipulated and measured: perceived usefulness and perceived ease-of-use. Both of these factors can be altered by changing the level of user control over the RS. UTAUT, on the other hand, involves four main factors. These factors are more complex and may not be easily varied or measured in an online experiment. For example, social influence may depend on the presence and feedback of other users, which is not available in an online experimental setting, as participation is simultaneous. Facilitating conditions may depend on the availability and quality of technical support, which may not necessarily be relevant or consistent given the online nature of the experimental setting.

Hypothesis development

Personalization has been proven to be one of the key elements enhancing user satisfaction with RS. Numerous studies have emphasized the importance of personalization in delivering tailored recommendations that better reflect user needs, resulting in greater satisfaction (Bostandjiev et al., 2012; Hijikata et al., 2012; Knijnenburg et al., 2012; Song et al., 2019). The underlying idea is that personalized recommendations reflect better on individual preferences, thereby streamlining decision-making processes (Adomavicius and Tuzhilin, 2005). A consistent finding in the research is that end-users are more satisfied with information systems when they consider these systems more useful (Mahmood et al., 2000; Venkatesh et al., 2003). As described in the Technology Acceptance Model (TAM), perceived usefulness is an important factor in determining an individual's intention to adopt a technology and, as a result, its acceptance (Davis, 1985). Systems perceived as more useful are believed to be better equipped to respond to users' needs and expectations, resulting in increased satisfaction and acceptance levels (Venkatesh and Bala, 2008).

User control is recognized as an essential mechanism to achieve personalization in RS (Jannach et al., 2017). By enabling users to influence recommendation processes, RS can better suit individual preferences, thus producing more tailored and relevant suggestions. The relationship between personalization, perceived usefulness, and user control can be elucidated by examining how user control improves personalized experiences. When users have greater control over RS, they can refine the system to better match their preferences and requirements (Chen and Pu, 2014). Considering the associations between personalization, perceived usefulness, and user control, I am positing the following hypothesis:

H1: *Higher degree of user control in RS leads to increased perceived usefulness of RS.*

User control represents an important component in modern recommender systems, as it empowers end-users to influence the recommendation generation process and provide feedback on suggested items (Pu et al., 2012). Such control mechanisms include setting preferences, employing preference filters, and providing explicit and implicit feedback on recommended items. A main concern with the incorporation of user control into recommender systems is the transition from passive recommendation consumption to active system participation (Knijnenburg et al., 2011). This change requires that users make decisions and provide input, including specifying preferences and implementing preference

filters. Such supplementary steps might be time-consuming and may not be fully aligned with user expectations of seamless system interactions (Xiao and Benbasat, 2007). While increased user control may yield more personalized and precise recommendations as a result of fine-tuned preference profiles, it simultaneously requires users to take additional steps, potentially undermining the system’s ease of use. Additionally, while feedback contributes to the refinement of the recommendation algorithm, users may exhibit reluctance or an inability to invest time and effort consistently in assessing recommendations and delivering feedback to the system (Herlocker et al., 2000). This additional step also imposes cognitive demands, as users must assess the relevance of the recommendation, determine suitable feedback, and articulate it. (Knijnenburg et al., 2012). Hence, I derive the following hypothesis:

H2: *Higher degree of user control will lead to a reduction in ease of use of RS.*

As the Information Systems (IS) domain expanded, scholars have turned their focus to the role of other factors in improving user adoption and continued use of innovative technologies. Users have been found to intend to use specific IT when they believed that they could control the way technology behaves (Teo, 2019). Having the ability to navigate the potential hazards and uncertainties associated with the use of unfamiliar technologies is crucial (Gefen et al., 2003). In addition to controlling the process, controlling the content that software generates has been associated with increased usage in IT systems (Lee, 2006). Within this context, the interplay between user control algorithm-driven frameworks, such as recommendation systems, has emerged as a focal point of interest.

Recommendation systems represent IT systems that utilize algorithms to provide personalized content, merchandise or services to users, based on their preferences, historical data, and additional contextual factors (Portugal et al., 2018). As these systems become increasingly ubiquitous and permeate diverse aspects of everyday life, it is essential for both researchers and practitioners to comprehend the elements that shape user intention to utilize such systems.

Scholars have concluded that users showed a higher propensity to adopt algorithms when given even slight control over the algorithm’s operations (Dietvorst et al., 2018). In view of these insights, I advance the following hypothesis:

H3: *H3: User control leads to higher behavioral intention (BI) to use RS.*

TAM also proposes that perceived ease of use and perceived usefulness predict the intention of the behavior to use a specific technology. Moreover, empirical evidence also shows that perceived ease of use also has a direct effect not only on intention to use, but also through perceived usefulness (Davis, 1989). Hence, in this light, I presume that the same would be true in RS:

H4: *H4: Higher perceived usefulness leads to higher intentions to use RS.*

H5: *H5: Higher perceived ease of use leads to higher intentions to use RS.*

H6: *H6: Higher perceived ease of use leads to higher perceived usefulness of RS.*

Figure 13 summarizes the theoretical model and hypotheses in this study.

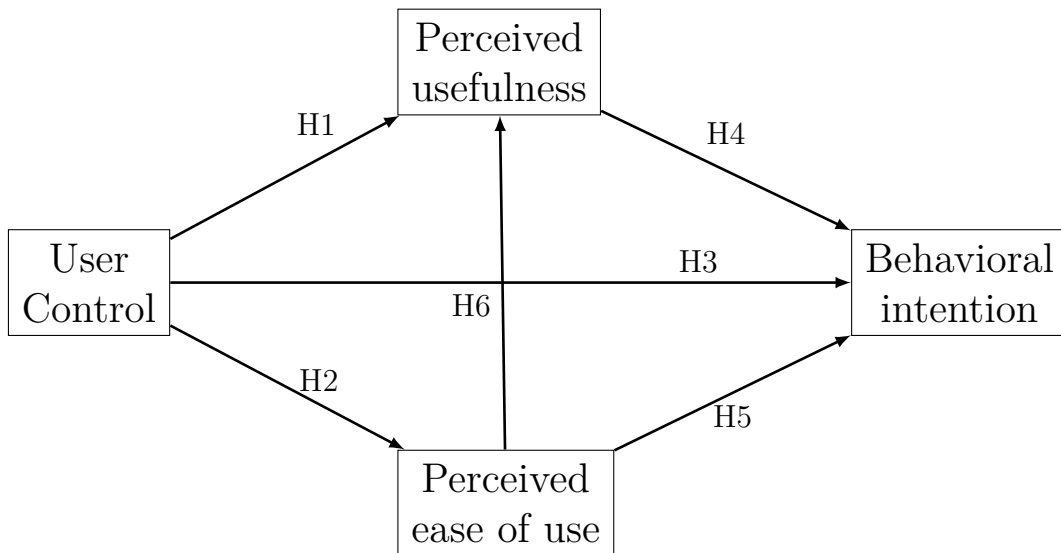


Figure 13: Model framework used in this paper. Sourced from Davis (1989). H4, H5 and H6 are the original relationships in Davis' work.

5.5 Methodology

To test these hypotheses I have conducted an online experiment with participants (hereafter users and participants are used interchangeably) residing in US recruited through Amazon MTurk. The experiment was hosted on a third-party website and consisted of an information search task and a questionnaire. I found information search task suitable for this study firstly because information search tasks can represent real-world scenarios where consumers are deciding between various options and also it is where the application of recommender systems is at its highest. Secondly, such a setting encourages users to

interact with the recommender system, as active involvement is crucial to understand users' acceptance of such technology. Third, an information search task provides an opportunity to directly measure the user's performance, such as the time taken to complete the task, the number of recommendations used, and the quality of the results obtained. These metrics not only can be used to evaluate the effectiveness of the recommendations, but also create an opportunity for us to tie users' performance to their remuneration. In previous studies, the experiments that involved recommender systems have either used an already existing mainstream recommender system where participants were asked to log in to those systems with their personal accounts, or used recommender systems with preference-elicitation phase that would require the user to rate the required number of items before proceeding with the experiment (Millecamp et al., 2018).

One can argue that the information search task may not be able to capture a variety of possible use cases of recommender systems. Recommender systems are used in various domains, such as e-Commerce, entertainment, social media, to count a few. Focusing only on an information search task may limit the generalizability of the study to other contexts. I address this concern in two ways. Firstly, I use randomly generated data in a multidimensional and multiattribute setting. It makes the task less dependent on any potential domain-specific knowledge, making the findings more generalizable. Secondly, I only recruit participants from MTurk who have masters qualification as they are more adaptable to varying tasks and are less likely to be affected by the task content.

Another argument can be that users might have varying expertise in performing online tasks and tying their remuneration to their performance might not give them enough intentions to perform well. By recruiting only participants with masters qualification and implementing two-stage remuneration process, I address both parts of this argument.

Amazon uses complex criteria including the variety of tasks users have previously participated in, the approval rate, and the relative timeframe between participations to assign master's qualification to MTurkers (commonly used self-identification term by people participating in tasks on MTurk). I implement a two-stage remuneration process where for completion of the task, the participant is paid the base amount no matter their performance. Then, based on performance, an additional amount is paid up to the base amount as a bonus, i.e., based on your performance, you can get double the base amount

as the total payment ¹⁹. Therefore, users are also eager to get higher bonuses.

Ultimately, with the proliferation of online experiments and portals like MTurk and Prolific where one can run such experiments, attention was drawn to various AI or simple rule-based bots, participants who would try to cheat in one way or another (Aguinis et al., 2020). To minimize such risks, I have implemented IP-based protection systems that would block the same user from participating again, or simultaneously participating in more than one experiment using different browsers ²⁰. Also, users trying to use any custom scripts were blocked and not paid. Only participants who resided in the US were eligible for the study. In total, I collected 400 usable observations between November 2022 and February 2023. The sample size was determined prior to the study and analysis was conducted only after the data gathering process ended.

After participants completed the search task, they would proceed to the questionnaire stage. Their answers did not affect the amount of remuneration, but only those who completed the questionnaire and submitted the code they were given in the end to Mturk were considered entitled to payment. The questionnaire consisted of 18 questions; three demographic, one attention check, and fourteen questions to measure variables of interest with seven-point Likert scale measurement were used: from 1-strongly disagree to 7-strongly agree. Two questions were used to measure behavioral intention and four per each of the remaining variables: user control, perceived ease of use, and perceived usefulness. The detailed description of the questions and measurement scales are available in the Appendix ^F.

5.6 Task details

In this section, I will discuss the information search task in detail. As soon as participants enter the portal where the experiment is hosted, they are provided with instructions on what to expect and how to perform. If any part of the instructions was unclear to them, they could navigate back to that part and read it again. They were informed that

¹⁹Also, if the task has optional bonus and the user is getting little fraction of the bonus, this is also flagged by Amazon's internal rating systems and it affects the user's master qualification directly.

²⁰All IP-based information has been deleted after the completion of the data collection process.

they would have 10 minutes to perform the search task, which would be followed by a questionnaire. Only those who complete the questionnaire would be eligible for payment. In the end, they would receive a “survey code” to provide to Mturk, which is the only way to signal the completion of the experiment.

After the participants read the instructions, they saw the main screen of the information search task. In this task, users were presented with a choice task to select five options from the set of choices given according to predefined “user preferences” within 10 minutes. The choice set which was located on the left hand side of the screen and consisted of hundred randomly generated options with a feature vector $\vec{f}(F_1, F_2, F_3, F_4, F_5)$, where $F_i \in U(0; 500)$. I decided to set the upper bound of the values in favor of 500 to increase the variance in the dimension to better reflect the real-world choice scenarios. Figure [14](#) shows a randomly generated choice set inside a blue rectangle.

For this search task, participants were asked to maximize the simple utility function $U = F_1 + F_3 + F_5$. This type of utility function was selected for several reasons. First, this utility could be related to different choice settings in the real world. For example, in vacation planning, travelers often consider multiple factors such as destination, accommodation, activities, budget, and climate. A pragmatic traveler might put a premium on the destination, activities, and budget while being more flexible on accommodation and climate. For example, they might look for a trip to an exotic location with adventurous activities within their budget, even if it means staying in a more modest accommodation or traveling during the off-season. Another example could be related to music selection where the user could be satisfied with three genres she likes, for example, rock and rap and orchestral classic music, only when the next song has all of them (And We Run by Within Temptation, for example). And if we assume that the features $F_1 \dots F_5$ are inverse representations, it can also be applied to airfare choice, where the customer would search for the options which has the lowest price, flight duration, and layovers. Second, such explicit definition of utility allows to evaluate subjects performance by comparing the average u over five chosen options with the accurate “ground truth” which one can calculate given the choice set (i.e., the average of the top five options that maximize u). This is important as it allowed me to tie subject’s remuneration to her performance and provide economic incentives for solving the task. Third, one can argue that we could

REMAINING TIME:
09:32

Please select five options with the highest possible $\rightarrow F1 + F3 + F5$ value. You can either select options from recommendations, or from the entire choice set. You have 10 minutes to complete the task. Do not forget to confirm your choice before the time runs out. Based on random assignment, you can neither hide columns nor sort them.

Hide F1 Hide F2 Hide F3 Hide F4 Hide F5

	F1	F2	F3	F4	F5
<input type="checkbox"/>	247	73	57	50	319
<input type="checkbox"/>	287	271	258	47	177
<input checked="" type="checkbox"/>	56	395	348	53	187
<input type="checkbox"/>	231	485	286	151	220
<input type="checkbox"/>	121	6	13	116	372
<input type="checkbox"/>	356	3	54	413	308
<input checked="" type="checkbox"/>	453	309	69	68	73
<input type="checkbox"/>	289	224	344	126	455
<input type="checkbox"/>	6	393	286	197	277
<input type="checkbox"/>	55	47	69	398	92
<input type="checkbox"/>	39	174	459	346	317
<input checked="" type="checkbox"/>	371	139	143	154	363
<input type="checkbox"/>	391	64	68	51	468
<input type="checkbox"/>	181	354	104	130	370
<input type="checkbox"/>	206	220	52	292	76
<input type="checkbox"/>	394	254	474	250	309
<input type="checkbox"/>	278	380	315	460	482
<input type="checkbox"/>	373	144	479	102	105
<input type="checkbox"/>	376	33	183	86	276
<input type="checkbox"/>	256	244	199	208	175

Total selected: 5

RECOMMENDATIONS

	F1	F2	F3	F4	F5
<input type="checkbox"/>	499	448	193	24	88
<input type="checkbox"/>	56	499	78	318	91
<input type="checkbox"/>	164	242	498	111	206
<input checked="" type="checkbox"/>	318	374	215	488	238
<input type="checkbox"/>	278	380	315	460	482
<input type="checkbox"/>	499	97	348	418	117
<input type="checkbox"/>	85	490	186	166	424
<input checked="" type="checkbox"/>	238	53	489	247	323
<input type="checkbox"/>	110	126	155	483	404
<input type="checkbox"/>	391	64	68	51	468

FINAL SELECTION

	F1	F2	F3	F4	F5
<input checked="" type="checkbox"/>	318	374	215	488	238
<input checked="" type="checkbox"/>	453	309	69	68	73
<input checked="" type="checkbox"/>	371	139	143	154	363
<input checked="" type="checkbox"/>	238	53	489	247	323
<input checked="" type="checkbox"/>	56	395	348	53	187

CONFIRM & PROCEED TO NEXT STAGE

Figure 14: Experimental interface: selected options are highlighted in green. Neither sorting or hiding feature is available for this participant. Rectangles were added ex-post for identification purposes and were not part of the interface. Left rectangle corresponds to the choice set, upper right to recommendations and lower right to final selection.

select utility as a function of only two or any other combination of three features, for example, $U = F_1 + F_2 + F_4$. While I agree that this could be done, I have selected the utility function as it is to prevent two features of interest from being besides each other. This would add some level of complexity to the task which would require the participant to gaze focus on nonneighboring dimensions. Studies have previously concluded that eye movements occur in real-life choice scenarios between dimensions (Noguchi and Stewart, 2018).

I also decided to round the values F_i to the nearest integer when presenting to reduce the cognitive capabilities required to perform the calculations. Although this may not fully mitigate cognitive difficulties in performing the task, having recruited only participants with masters qualification certainly helps to neutralize it more (Aguinis et al., 2020).

As depicted in the figure 14, along with the choice set, on the left-hand side of the screen, the user is presented with ten “recommended options” that they can use to solve the task. Users could select and deselect options from either the choice set or from the recommended options. Unlike previous studies about recommender systems that have used readily available systems like Movielens (Kumar et al., 2018), Spotify (Millecamp et al., 2018) that use collaborative filtering and other complex machine learning-based algorithms, the decision was made to take a different approach. The main criteria in deciding were that it had to be a simple and natural recommendation algorithm, which resulted in diverse recommendations, while being easy for users to understand and also to make changes to. The algorithm was somewhat disconnected from the user’s preferences. It simply collected options that have high value in each of the five option characteristics (F_i s). Given the attributes of the products, RS would assume by default that the user is interested in at least one of the attributes, as is the case in many real-world systems (Gu et al., 2019). The algorithm presents participants with options possessing the highest values in each dimension, generally resulting in a higher expected utility compared to the expected utility of the random selection from the set of choices. However, the algorithm will not supply users with a set of options that directly corresponds to the search task’s genuine solution, as it neglects the necessary pairing across three dimensions mandated by the task. The design of such an algorithm mirrors real-life decision-making scenarios in which the recommendation system lacks comprehensive information about the user’s

preferences and must make assumptions about preference weights across all dimensions (Schein et al., 2002).

Another feature of the recommender system is that users are also presented with the control mechanism to interact with the RS aiming to improve the recommendation set and thus simplify her search task. Users can eliminate up to three features from algorithm's consideration. Removing four out of five features would have resulted in options corresponding to top 10 values in the remaining feature, which would eliminate the usefulness of recommendations to users. At the beginning, the recommender algorithm takes all five features into account. Based on how many features the user has removed, the number of elements that correspond to the top values in each included dimension changes. Here is the summary of how the algorithm generates recommendations with different scenarios:

1. All features are present. In this case, options corresponding to the top two values for each feature are included in the final recommendation. There could exist cases in which one or more options correspond to top values in different features. I refer to this as an overlap event. In case of overlap, let us say that for the second top value, the algorithm will randomly decide which feature to "keep" for that option, and for the other feature(s) it will select the next order, which will be three. This process will continue until there are no overlaps and ten options are selected.
2. Four features are present. In this case, the algorithm will select the top two values per each feature and solve any overlap events. Then, for the remaining two options, it will randomly decide for which feature(s) it should include the remaining two variables which will randomly be either the third, or both the third and fourth top values. Lastly, it will solve any overlap events.
3. There are three features. Similarly to the case above, the top three elements are selected per feature. The overlap events are solved. The remaining option will be the fourth top element based on randomly selected features of the three. Finally, it will solve any overlap events.
4. There are two features. The algorithm will select five options for each feature

corresponding to the top five elements in each feature. Then, the overlap events will be solved.

The expected utility of ten recommendations could be calculated by $U_R = \sum_{r=1}^{10} U_r$ where U_R is the utility of the recommendations and U_r is the utility of the single recommendation. The utility of a single recommendation could be calculated using $U_r = U_{F_1} + U_{F_3} + U_{F_5}$.

Assume that there is a vector X that has 100 elements, where $X_i \sim U(0, 1)$. If one orders this vector in ascending order, the expectation of the order statistic of k th is calculated as

$$\begin{aligned}
E(X_{(k)}) &= \frac{n!}{(k-1)!(n-k)!} \int_0^1 x^k [1-x]^{n-k} dx \\
&= \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} \int_0^1 x^k [1-x]^{n-k} dx \\
&= \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} \cdot \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} \int_0^1 \frac{\Gamma(n+2)}{\Gamma(k+1)\Gamma(n-k+1)} x^k [1-x]^{n-k} dx \\
&= \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} \cdot \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} \\
&= \frac{\Gamma(k+1)\Gamma(n+1)}{\Gamma(k)\Gamma(n+2)} = \frac{k}{n+1},
\end{aligned}$$

where k is the order of the element we are interested in, n is the number of elements in the vector. Subsequently, in case $X_i \sim U(0, \theta)$ which is this case ($\theta = 500$), the transformation ends up with

$$E(X_{(k)}) = \theta \frac{k}{n+1}.$$

By applying this formula, I can calculate the mathematical expectation of each recommended option and as a result, the entire set of recommendations. The default recommendation set would have an expected utility of 894. If, for example, the user hides all useless features, F_2 and F_4 , then the expected utility of the recommendation set would increase to 981 or by almost 10 percentage points. If the opposite happens, i.e., user hides all useful features, F_1 , F_3 and F_5 , then the expected utility of recommendation set would decrease significantly to 750 or by 16 percentage points from the default.

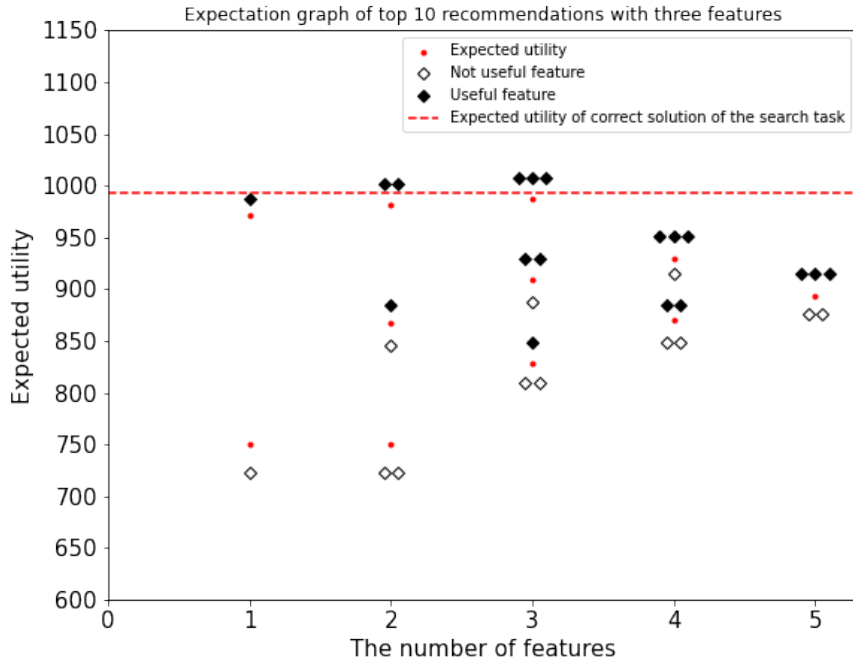


Figure 15: The expectation of the recommendations given different scenarios. Diamond symbols represent the features affecting the algorithm. Their positions are arbitrary. The special case where only 1 feature affects the algorithm is included for convenience.

An important aspect of this design is that, depending on user behavior, the usefulness of the recommendation system can actually be measured by measuring the change in overall return across all recommended options. Figure 15 shows the expectation of the recommendations given the different scenarios. So, not only would I be able to study the differences in user behavior between experimental treatments but also could study which part of the performance differential is actually due to objective change in the quality of recommended options. Refer again to figure 14 for additional details of the experimental interface.

Modern search interfaces usually have tools to facilitate a smoother user experience along with recommendations. These may include sorting and filtering features that are not part of the recommender system per se. However, they offer users help to find what they are looking for. For this reason, I have added a sorting feature as an additional characteristic. Users could sort options in increasing or decreasing order along each of five dimensions in both the choice set and the recommended options table on the right-hand side. This potentially helps them to solve the task, maximize U , but does not

change the set of recommendations. Although this feature does not directly improve the recommendations, it makes it easier to solve the task and can be used in conjunction with the RS to complete the task faster. Therefore, this can be seen as a feature that creates control over the search task, but does not have an impact on the generation of recommendations (Jannach et al., 2017).

Participants were randomly assigned to one of four treatments based on whether subjects have sorting and feature hiding capabilities. Users were rewarded with a fixed 2\$ and additional bonus up to 2\$ which was calculated as

$$R = 1 + \frac{U_u}{U_a},$$

where U_p is the performance of the user, which is the average utility of the five final options selected by the user and U_m is the maximum performance, which is the average utility of five options that correspond to the actual options with the highest values of U . I performed 1000 simulations assuming participants would randomly select five options from a) default provided recommendations; b) recommendations as a result from hiding unnecessary columns (i.e. best case); c) recommendations as a result hiding necessary columns (i.e. worst case); d) choice sets without considering recommendations. Even in the worst-case scenario (scenario d), the average remuneration for an approximately 15 minute experiment is around 3.2\$ which is greater than the Federal minimum hourly wage of 7.25\$ per hour²¹. Figure 16 compares the remuneration of the simulations and the actual remuneration of the participants in the experiment.

5.7 Results

Descriptive results

I obtained 400 usable samples. The average respondent was a 42-year-old male with a bachelor's degree. Considering the average age in the United States being 38 and the median age ranging between 31 and 45²², I can state that the sample corresponds well

²¹Retrieved from <https://www.dol.gov/general/topic/wages/minimumwage>. Applicable from March 2023.

²²Retrieved from <https://worldpopulationreview.com/state-rankings/median-age-by-state>.

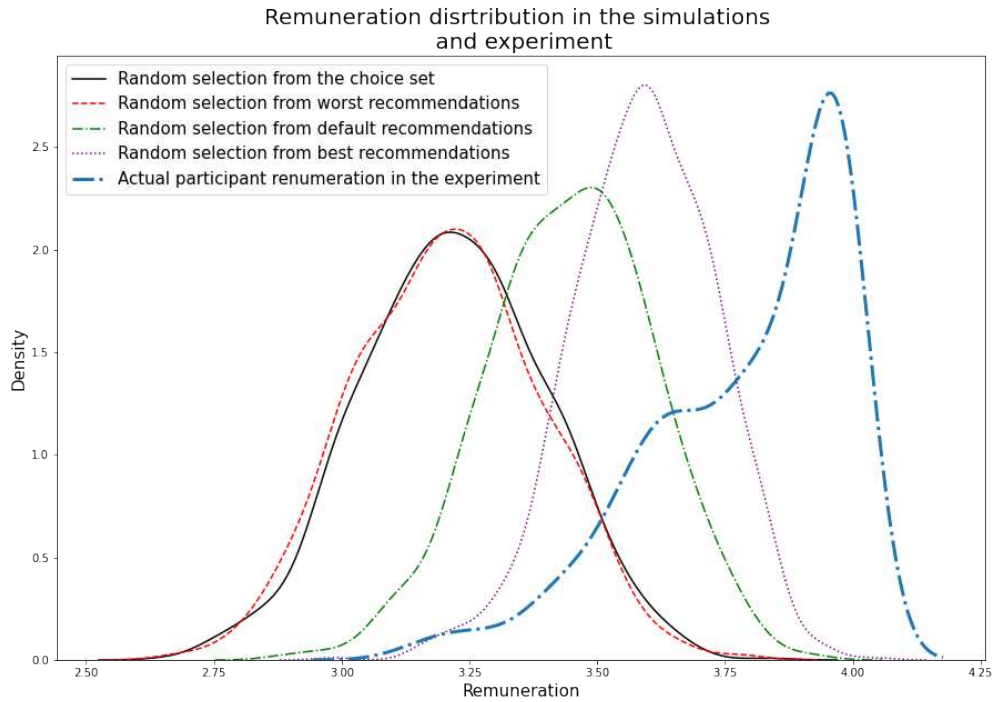


Figure 16: Remuneration distribution in the simulations and the actual experiment.

to the US population. 43% of the respondents were female while having 4 non-binary individuals. 52% of the respondents had a bachelor degree and 10% also had a master degree. The results show that the group of participants is diverse in gender and education level and is not abnormal.

Before proceeding to the main results, it is important to assess the effectiveness of the recommender system. I have previously shown their hypothetical effectiveness in the figure 15. Figure 17 shows the performance observed of the recommendations given different scenarios. The recommendations which were generated using only nonuseful features as expected had the least utility, while recommendations comprising of only useful features topped the remaining scenarios. Recall that there were four treatment groups. Participants who could alter recommendation generation on average scored higher than those who could not. On the contrary, the sorting was found not to improve the overall performance of the participants. Table 16 shows descriptive information about the groups and their performances.

One advantage of having an external experiment setup was that it allowed me to mea-

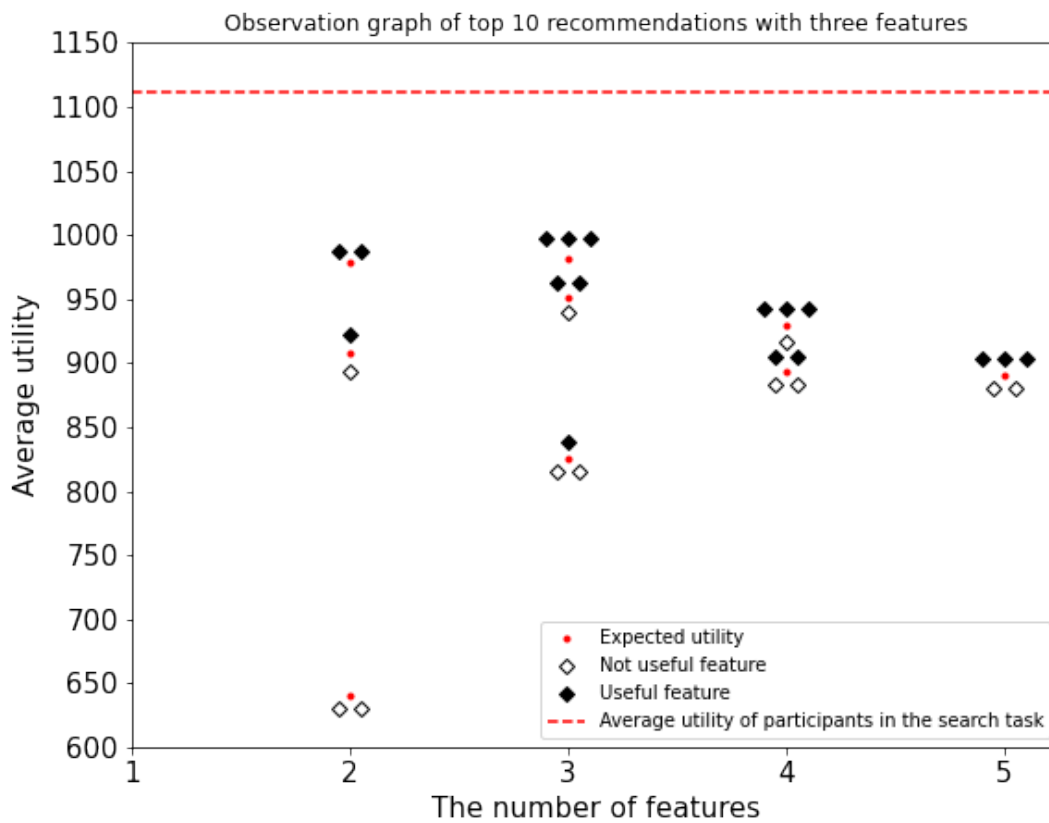


Figure 17: The observed utilities of the recommendations given different scenarios. Diamond symbols represent the features affecting the algorithm. Their positions are arbitrary. Dashed line represents the average utility of participants in the search task.

Treatment group	Hiding	Sorting	Count	Mean bonus	Mean completion time
1	No	No	100	1.78	257
2	No	Yes	100	1.74	260
3	Yes	No	99	1.81	239
4	Yes	Yes	101	1.81	246

Table 16: Overview of treatment groups and reward and time spent (in seconds) on the search task.

sure all experiment-related activity of the participants. They ranged from selection and deselection of a particular option to sorting and hiding of particular features. From the final choice of the participants, I identified which options were selected from the recommendation table and which ones were selected from the overall choice set. To test the effects of sorting and hiding, I have conducted t-tests comparing treatment groups on three metrics: completion time of the search task, amount of performance-based bonus and the share of the options which were chosen from the recommendation set. Despite its statistical insignificance, by analyzing the share of the options selected from recommendations in the participant’s submission, it was concluded that users who could change the way the recommendations are generated were more likely to select the final options from the recommendations, when they could not sort. Furthermore, participants who could hide finished the search task on average faster than the other group. Sorting alone did not lead to increased performance among participants. However, the combination of sorting and hiding did have a positive and significant effect on user performance. Table 17 shows the results of the comparisons between different treatment groups.

Effect	Treatments	Mean bonus		Mean completion time in seconds			Selection share from recommendations		
		Effect size	t-stat	p-value	Effect size	t-stat	p-value	Effect size	t-stat
Sorting	2 vs 1	-0.04	-1.23	0.90	3	0.36	-0.01	-0.61	0.54
Sorting	4 vs 3	-0.01	-0.01	0.50	7	0.86	-0.03	-0.11	0.91
Hiding	3 vs 1	0.03	1.32	0.10	-18	-2.23	0.01	0.06	0.95
Hiding	4 vs 2	0.07	2.35	0.01	-14	-1.54	-0.24	-0.68	0.50

Table 17: Difference tests among treatment groups.

Another interesting activity to analyze was the sorting and hiding activity that users have performed. Although only two treatment groups could remove features from the recommendation algorithm or sort by them, 70% of the time they removed “non-useful” features. The average participants used the sorting feature 13 times until they made their final choice, approximately 80% of the time sorting by “useful” features. Table 18 shows further information on the usage of sorting and hiding per feature.

Feature	Hidden		Sorted	
	count	share	count	share
F1	65	0.15	151	0.16
F2	154	0.35	135	0.15
F3	47	0.11	331	0.36
F4	146	0.33	67	0.07
F5	24	0.06	232	0.25

Table 18: Sorting and hiding statistics per feature.

Main results

To analyze the relationship defined by the research questions, I have used Structural Equation Modeling. Factor loadings and Cronbach alphas are presented in table 19. Although the goodness-of-fit chi-square test yields a significant p-value, the large size of the sample justifies using other statistics to test for goodness of fit (Schermelleh-Engel et al., 2003). In the literature, for such cases, other goodness-of-fit measures were proposed (Schreiber et al., 2006). Two of them, the comparative fit index and the Tucker-Lewis index for the goodness-of-fit analysis, both are 0.97 which is considered a good fit (Schumacker and Lomax, 2004). 75% of the variation in BI is explained by the model and five of six hypotheses are confirmed. The results of SEM are presented in figure 18.

All three hypotheses corresponding to the original relationships of the TAM model were confirmed (H4, H5, H6). Users perceived the easy-to-use RS to be more useful and intend to use them more. Furthermore, perceived usefulness had a strong positive effect on behavioral intentions, suggesting that users were more likely to adopt an RS if they find it useful to meet their needs.

Users found RS systems with user control to be more useful. A percentage point increase in user control led to almost a half percentage point increase in perceived usefulness.

Factor	Item	Loading	Cronbach's alpha
Perceived Usefulness	PU1	0.97	0.98
	PU2	0.97	
	PU3	0.96	
	PU4	0.97	
Perceived Ease of Use	PE1	0.76	0.85
	PE2	0.79	
	PE3	0.65	
	PE4	0.90	
User Control	UC1	0.91	0.92
	UC2	0.96	
	UC3	0.90	
	UC4	0.73	
Behavioral Intention	BI1	0.98	0.98
	BI2	0.97	

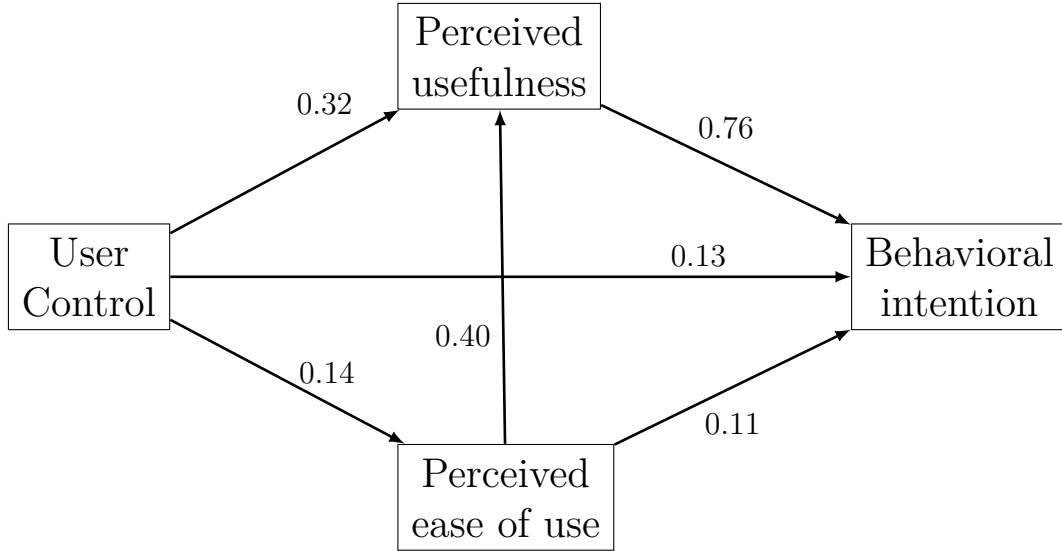
Table 19: Factor loadings of items and Cronbach's alpha of the main factors. Refer to Appendix A for more information about factors and corresponding questions.

Therefore, H1 was confirmed. This is not surprising, as this result is similar to previous research conducted in specific domain settings (Bostandjiev et al., 2012; Millecamp et al., 2018).

User control was also an important construct to directly explain the behavioral intention to use an RS. For every one percentage point increase in user control, participants showed more than half the percentage point more intentions to use an RS. Therefore, H3 was also confirmed.

However, contrary to my expectations, H2 was not confirmed. Participants did not perceive the RS with control to be less easy to use. This result needs further discussion. The theoretical foundations for H2 have been based on established research streams that established a connection between information overload and consumer decision-making (Jacoby et al., 1974; Chen et al., 2009). Although increased information leads to better informed decisions, research has established a trade-off between information richness and decision quality (Jacoby et al., 1974; Malhotra, 1982). Furthermore, information overload has been found to negatively affect the acceptance of the internet tools (Shih, 2004).

A potential explanation for such a counterintuitive outcome on the result of H2 testing is the type of control mechanisms employed in the experiment. Recall that I have defined the scope of control mechanism as a mechanism which “has immediate effect on recom-



Hypothesis	Expected sign	Effect size	t-stat	p-value
H1	+	0.480	128.2	0.000
H2	-	0.115	75.3	0.995
H3	+	0.568	142.3	0.000
H4	+	0.773	284.2	0.000
H5	+	0.647	140.9	0.000
H6	+	0.608	134.9	0.000

Figure 18: Main results (above), summary of hypotheses and total effect sizes.

mendations, is easily understandable and does not prevent the user from interacting with the system”. The previous literature has mentioned a number of user control mechanisms that can be employed. They range from simple critique-based to more weight scale and visualization-based complex control mechanisms (Jannach et al., 2017; Jin et al., 2018). Considering the theoretical foundations of H2, one needs to further investigate this outcome. Recall that users had ten minutes to solve the information search task and they were given two control mechanisms. Disentangling the effect of both control mechanisms on completion time is one possible avenue to pursue. One way to do this is to look at the completion time of the search task. T-tests were conducted where different groups were compared in terms of completion time. The results showed that hiding allowed participants to finish their task earlier, while sorting did not lead to any significant reduction in completion time. These results lead to the conclusion that the critiquing (hiding) and sorting control mechanisms have different, rather than strictly contradicting, effects on

task completion time. On the one hand, the hiding mechanism showed a significant negative effect on task completion time, indicating that users who could hide items were able to complete the task more quickly and easily. This suggests that the hiding mechanism has effectively streamlined the decision-making process for participants while allowing them to focus on more relevant options and eliminate unwanted or non-useful aspects of the algorithm. This could have had a positive impact on perceived ease of use. On the other hand, the sorting mechanism having an albeit insignificant but positive effect on task completion time implies that its impact on perceived ease of use is less clear and may vary. While the effects of these two control mechanisms are not necessarily contradicting each other, they do highlight that the two control mechanisms have differing impacts on users' experiences.

Another explanation could be that engaging with control features helped participants learn more about the recommender system's functionalities and underlying logic. Control mechanisms provided participants with the ability to interact with the system and receive feedback from it. This process could facilitate the learning process and help participants develop a mental model of functionality, giving them a perception of increased ease of use (Norman, 2013). As participants became more familiar with a given system, they might have found it easier to navigate and interact with its features. This could have led to a higher perceived ease of use, as users felt more confident and comfortable using the system (Venkatesh, 2000).

5.8 Conclusion

Implications

The growing speed of digitization of our lives and the presence of AI means even more applications of recommender systems. Although well-formulated and diverse recommendations are crucial for the success of companies and the growing importance of human-AI interaction, it is important to understand the challenges in the implementation of control mechanisms (Dietvorst et al., 2018; Song et al., 2019). There are several important implications for system designers that arise from this study.

Firstly, system designers must take into account that, although user control might in some cases lead to users believing those systems are easier to use, the relationship between

them are more complex and nuanced. Hence, they must not sacrifice the simplicity of the systems by adding more control. They must be aware that choice context may have a big effect on users' attitudes towards control mechanisms. Users who know exactly what they are looking for will feel more comfortable when given more granular control over the recommendation generation. Casual users, on the other hand, might not want to have to do everything and rather leave that to algorithms. Real-world applications of such approach could be a system which offers simple control to new users and gradually increases the complexity of control mechanisms as the user becomes more proficient with the system. Also, more granular control will require a steeper learning curve for the users, and considering that algorithm alteration results are sometimes not immediately visible, it might take some time for users to understand the results of their actions. Another angle to look at this would be from a decision-paralysis perspective. When users face a multitude of options to choose from, in terms of which control mechanism to use, they might feel confused and not decide at all (Schwartz, 2004).

Secondly, the fine trade-off between the level of control and the usefulness of the results must be taken into account as well. Studies have concluded that users appreciate serendipitous results now and then, but are still sensitive to the relevance of results to their preference profiles (Kotkov et al., 2016). Having greater control over recommendation generation might allow users to shift the weights of their profiles in the system too much. This will result in RS either generating very similar items to those the user has consumed, or completely throw-off the system (Mantovani et al., 2019). Too much control might also enable algorithm hijacking, intentionally or unintentionally, leading to suggestions of irrelevant items for users (Xing et al., 2013).

Thirdly, control is more useful when users understand how it works. Although a detailed explanation of complex control tools is not in the interest of the business because it can reveal proprietary knowledge to competitors (Lubit, 2001), nor the users because they have limited ability to understand complex systems (Kahneman and Tversky, 1979a), people generally use algorithmic tools more when they have some understanding of the underlying mechanisms (Guidotti et al., 2018). The current experimental study was free of contextual factors for generalizability reasons. However, RS in real-world scenarios are contextualized, which means, regardless of user control method, during the process con-

sumer might reveal some sensitive information about themselves to the system. It can be anything from their political views, financial or social status, geographical information, and so on. Too granular control makes it harder to manage user consent for system designers (Bélanger and Crossler, 2011). With various control options and granular settings, users may find it challenging to understand the extent of their data being collected and how it is used, potentially leading to privacy concerns. In those cases, it is important to inform users about how their data are gathered, how it affects recommendations, and how one can change that. For example, Netflix gives a basic understanding of its recommendation algorithm, which is easy to understand to any user type, be it novice or expert (net). Furthermore, giving too granular feedback to the recommendation algorithm may result in the user leaving a fine detailed digital footprint, which could make anonymization harder and open doors to various malicious attacks (Sweeney, 2002).

Summary, limitations and future directions

I have investigated the effect of user control on recommender systems' acceptance in a domain-independent experimental setting. TAM was used as a theoretical framework. This study confirmed the relationships of the original TAM model, indicating that users perceive easy-to-use RS as more useful and intend to use them more. Also, usefulness had a strong positive effect on the behavioral intentions to use an RS, suggesting that users were more likely to use an RS if they found it useful in meeting their needs. Furthermore, user control has been found to be an essential construct in directly explaining the intention of the behavior to use an RS.

However, contrary to expectations, the participants did not perceive RS with control to be less easy to use. Possible explanations include the type of control mechanisms employed in the experiment and their varying impacts on users' experiences. The hiding mechanism showed a significant negative effect on task completion time, suggesting a positive impact on perceived ease of use, while the sorting mechanism's effect was more ambiguous. Engaging with control features may have helped participants learn more about the recommender system's functionalities and underlying logic, hence contributing to increased perceived ease of use.

One limitation of the study is the usage of a single utility function although I have provided the reasons granting me the certainty of using such an approach. One way to

address this limitation can be by introducing randomly generated utility functions. For example, having utility defined as $U = F \cdot W$ where $W_{i,j} \sim U(-1, 1)$. This would allow to capture more diverse preferences and reduce the differences between the experimental and real-world choice scenarios even further (Vesänen, 2007). Secondly, it would introduce a preference trade-off between different features. It would make the choice setup more complex and hence further resemble the actual choices we face on a daily basis.

Another limitation of the study arises from the fact that I have used two of many control mechanisms. Future research may incorporate a broader range of control mechanisms and study their effect both individually and in combination with each other to investigate the effect of user control on RS acceptance. By doing so, one may identify the differences between various control methods from the user's perspective and better understand the aspects that affect user acceptance by disentangling their effects.

The findings indicating the varying effects of control methods open doors for future research of recommender systems with dynamic control mechanisms. Dynamic user controls have the potential to provide superior decision support by presenting control options that align with users' current context or requirements. For instance, a travel booking website's recommender system could furnish users with more granular control options when searching for accommodations in popular tourist destinations with a multitude of choices to which the user has previously been, but more simplistic options in case of a first visit. Another example can be a music streaming platform that could extend different control ways depending on users' ongoing activities, such as providing streamlined controls for users listening to music while working and more intricate controls for users actively curating a playlist for a social gathering. Lastly, a streaming service's recommender system might present basic control options for users seeking a quick movie recommendation, while supplying more comprehensive options for users interested in delving into a specific genre or director's oeuvre. Balancing these aspects well might be crucial to increase sales and drive greater engagement.

Recent technological progress in face of LLMs (large language models) creates another exciting opportunity for future research in the context of user control in recommender systems. LLM based agents that leverage billions of parameters and terabytes of human generated knowledge enable human-like natural language interactions between users and

the system, and can easily become an instrumental technology in enhancing user control, increasing trust, and satisfaction. The key benefits of incorporating LLMs in RS through conversational agents is their ability to understand human input better, take user input by offering a more intuitive and transparent way to express their preferences and requirements. They can also be used to give explanations for recommendations, thus enhancing transparency.

Considering the rapid technological developments of the recent years, I may state with confidence that user control in recommender systems is about to take a huge leap forward and open even more research directions.

6 Conclusion

Part of the journey is the end.

The journey of my dissertation began with the enthusiastic aim of exploring the nature and nuanced influences of context effects on decision making and their applications on recommender system design. This voyage comprises four studies, each contributing to a richer understanding of these effects and jointly offering a new perspective on the complex interplay of context effects.

The first study embarked on a path towards understanding context within choice sets, an idea derived from economics and decision-making literature that views context as a coexistence of alternatives within a menu. The power of a computational model was employed on real-world choice data, uncovering complex patterns and relationships that affirmed the significant role of context effects. The results illustrated the efficacy of the model’s potential as an effective tool for understanding the complex dynamics of context effect in multiattribute and multidimensional setting.

The second study continued this exploration of context effects further by adding a new layer of complexity. Its aim was to unveil three context effects types that previously investigated in experimental setting, but this time achieving it in multidimensional setting and provide a methodology to quantify them. A flexible framework was introduced, which proved to be capable of calculating those effects in different settings, both observational and experimental, and handling diverse choice sets. Its adaptability and scalability make it a robust tool for large-scale applications, both online and offline. The implications of this framework are immense and include the potential for more precise design of consideration sets and using them to enhance recommender systems. This is particularly valuable for online platforms offering an overwhelming range of options considering the proliferation of technologies that project decision making to virtual settings.

The third study built on the findings of the previous one and delved further into the realm of context effects while offering a practical solution to the persistent “cold-start” problem plaguing recommender systems. This novel, context-driven, two-stage decision heuristics approach has further underscored the vital role of context effects steering user

decisions, by integrating heuristics and context effects into the design of recommender systems.

The final study made a crucial shift to the context that arises outside of the set of choices, defined within the literature on recommender systems as information related to the consumer, such as their preferences, time, and day of choice. Here, the focus has narrowed and turned to the user preferences, namely, interaction between user control mechanisms and the acceptance of recommender systems. The study suggested that user-centric control mechanisms could potentially lead to elevated user engagement and acceptance, indicating the positive externalities of those mechanisms in a new era of recommender systems that shift towards the “help me help you” approach.

Altogether, these studies point towards an innovative blueprint for future recommender systems that harness the power of context effects both within and outside the choice sets and provide not only “accurate”, but also “useful, novel, surprising” recommendation. However, as in every work before it, this work is also not without limitations.

Albeit the computational model proved to bring its two cents to the table, its contributions lacked generalizability because it was the only model employed. It provides an exciting look at future research which could develop an approach utilizing other computational models that the current decision-making literature is rich of.

The approach taken in the second study mainly accommodated context effects as a result of ordinal relationships, overlooking the potential of cardinal measurements. Future research could tackle this by factoring in the magnitude of the relationship, adding depth to the analysis. Also, an interesting avenue could be research toward quantifying context effects outside of the “trinity”.

Despite the effectiveness of the two-stage heuristic approach developed in the third study, its direct application to modern systems may not be straightforward due to the traditional emphasis on predictive power. Although this emphasis has already started to gradually change, enhancing predictive capabilities of this approach could definitely make it more attractive for integration into recommender systems.

Lastly, the fourth study has looked at the universe of user control mechanisms only through a small prism, employing two of them. However, there is a vast realm of other control mechanisms awaiting exploration in future research, calling for studies of more

complex experimental settings, while maintaining domain neutrality. Notwithstanding the fact that it seems challenging, I consider this achievable.

In conclusion, this dissertation serves as an illuminating guidepost in a place where understanding our own is limited. Before this work, the study of context effects was restricted to simplistic choice sets and experimental data, and so was limited our understanding of user control tools. This study not only extends our knowledge of context effects to new multidimensional realms, but also achieves isolation of external context via experimental setting when investigating user control mechanisms.

The potential unveiled in this dissertation is substantial, albeit not fully quantifiable at this juncture. The findings of this work provide the foundation for potential advancements in the application of context in recommender systems design and operations, which could alter users' engagement within modern digital landscape. The unique approach to studying context effects and the capability to disentangle them in multiattribute, multi-dimensional settings put forth in this dissertation fosters new insights and perspectives in the complex, dynamic decision-making process in us, the humans. As the renowned psychologist and Nobel laureate Daniel Kahneman once said, "Our comforting conviction that the world makes sense rests on a secure foundation: our almost unlimited ability to ignore our ignorance." This dissertation, through its explorations and findings, nudges our understanding a little closer to recognizing the nuanced complexity of decision making, especially in relation to context effects, thus revealing a bit more of the world that we so often ignore.

References

- Netflix help center. <https://help.netflix.com/en/node/100639>. Accessed: March 31, 2023.
- Emad Abu-Shanab, John M Pearson, and Andrew J Setterstrom. Internet banking and customers' acceptance in jordan: the unified model's perspective. *Communications of the Association for Information Systems*, 26(1):493–524, 2010.
- Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749, 2005.
- Gediminas Adomavicius and Alexander Tuzhilin. Context-aware recommender systems. In *Recommender systems handbook*, pages 217–253. Springer, 2011.
- Herman Aguinis, Isabel Villamor, and Ravi S Ramani. Mturk research: Review and recommendations. *Journal of Management*, 47(1):237–263, 2020.
- Michal Aharon, Natalie Aizenberg, Edward Bortnikov, Ronny Lempel, Roi Adadi, Tomer Benyamini, Liron Levin, Ran Roth, and Ohad Serfaty. Off-set: one-pass factorization of feature sets for online recommendation in persistent cold start settings. In *Proceedings of the 7th ACM Conference on Recommender Systems*, pages 375–378, 2013.
- Kwasi Amoako-Gyampah and Abdus F Salam. An extension of the technology acceptance model in an erp implementation environment. *Information & management*, 41(6):731–745, 2004.
- Annie I Antón, Julia B Earp, and Jessica D Young. How internet users' privacy concerns have evolved since 2002. *IEEE Security & Privacy*, 8(1):21–27, 2010.
- Anocha Aribarg, Thomas Otter, Daniel Zantedeschi, Greg M Allenby, Taylor Bentley, David J Curry, Marc Dotson, Ty Henderson, Elisabeth Honka, Rajeev Kohli, et al. Advancing non-compensatory choice models in marketing. *Customer Needs and Solutions*, 5(1-2):82–92, 2018.

- Zakaria Babutsidze, William Rand, Emil Mirzayev, Ismael Rafai, Nobuyuki Hanaki, Thierry Delahaye, and Rodrigo Acuna-Agost. Asymmetric dominance in airfare choice. In *6th International Choice Modelling Conference*. ICMC, 2019.
- Seyed Ali Bahrainian, Andrea Iacovella, Fabio Crestani, and Mounia Lalmas. A deep hybrid model for recommendation systems. *arXiv preprint arXiv:2009.09748*, 2020.
- Richard R Batsell and John C Polking. A new class of market share models. *Marketing Science*, 4(3):177–198, 1985.
- Gordon M Becker, Morris H DeGroot, and Jacob Marschak. Measuring utility by a single-response sequential method. *Behavioral science*, 9(3):226–232, 1964.
- France Bélanger and Robert E Crossler. Privacy in the digital age: A review of information privacy research in information systems. *MIS quarterly*, 35(4):1017–1038, 2011.
- Moshe E. Ben-Akiva and Steven R. Lerman. *Discrete Choice Analysis: Theory and Application to Predict Travel Demand*. MIT Press, Cambridge, MA, 1985.
- Moshe E Ben-Akiva, Steven R Lerman, and Steven R Lerman. *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press, 1985.
- Nicolas AJ Berkowitsch, Benjamin Scheibehenne, and Jörg Rieskamp. Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology: General*, 143(3):1331, 2014.
- James R Bettman. Memory factors in consumer choice: A review. *Journal of Marketing*, 43(2):37–53, 1979.
- James R Bettman, Eric J Johnson, and John W Payne. A componential analysis of cognitive effort in choice. *Organizational behavior and human decision processes*, 45(1):111–139, 1990.
- James R Bettman, Mary Frances Luce, and John W Payne. Constructive consumer choice processes. *Journal of consumer research*, 25(3):187–217, 1998.
- Sudeep Bhatia. Associations and the accumulation of preference. *Psychological review*, 120(3):522, 2013.

- Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. Recommender systems survey. *Knowledge-based systems*, 46:109–132, 2013.
- Svetlin Bostandjiev, John O’Donovan, and Tobias Höllerer. Tasteweights: a visual interactive hybrid recommender system. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 35–42, 2012.
- Amanda Bower and Laura Balzano. Preference modeling with context-dependent salient features. In *Proceedings of the 37th International Conference on International Conference on Machine Learning*, volume 119, pages 1067–1077, 2020.
- John S Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. *arXiv preprint arXiv:1301.7363*, 2013.
- Gordon DA Brown and William J Matthews. Decision by sampling and memory distinctiveness: Range effects from rank-based models of judgment and choice. *Frontiers in Psychology*, 2:299, 2011.
- Robin Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12:331–370, 2002.
- Andrew Burton-Jones and Geoffrey S Hubona. The mediation of external variables in the technology acceptance model. *Information & management*, 43(6):706–717, 2006.
- Jerome R Busemeyer and Adele Diederich. Survey of decision field theory. *Mathematical social sciences*, 43(3):345–370, 2002.
- Jerome R Busemeyer and James T Townsend. Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*, 100(3):432, 1993.
- Jerome R Busemeyer, Sebastian Gluth, Jörg Rieskamp, and Brandon M Turner. Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. *Trends in cognitive sciences*, 23(3):251–263, 2019.

- Siarhei Bykau, F Koutrika, and Yannis Velegarakis. Coping with the persistent cold-start problem. *Personalized Access, Profile Management, and Context Awareness in Databases*, 2013.
- Andrea M Cataldo and Andrew L Cohen. The comparison process as an account of variation in the attraction, compromise, and similarity effects. *Psychonomic Bulletin & Review*, 26(3):934–942, 2019.
- Yu Chen and Pearl Pu. Designing emotion awareness interface for group recommender systems. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, pages 347–348, 2014.
- Yu-Chen Chen, Rong-An Shang, and Chen-Yu Kao. The effects of information overload on consumers’ subjective state towards buying decision in the internet shopping environment. *Electronic Commerce Research and Applications*, 8(1):48–58, January 2009. ISSN 15674223. doi: 10.1016/j.elerap.2008.09.001.
- Alan DJ Cooke and Barbara A Mellers. Multiattribute judgment: Attribute spacing influences single attributes. *Journal of Experimental Psychology: Human Perception and Performance*, 24(2):496, 1998.
- Clyde H Coombs. Mathematical models in psychological scaling. *Journal of the American Statistical Association*, 46(256):480–489, 1951.
- Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. Performance of recommender algorithms on top-n recommendation tasks. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 39–46, 2010.
- John W Creswell. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications, 2013.
- Imran S Currim, Robert J Meyer, and Nhan T Le. Disaggregate tree-structured modeling of consumer choice data. *Journal of Marketing Research*, 25(3):253–265, 1988.
- Abhinandan S Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web*, page 271–280. ACM, 2007.

- Fred D Davis. *A technology acceptance model for empirically testing new end-user information systems: Theory and results*. PhD thesis, Massachusetts Institute of Technology, 1985.
- Fred D. Davis. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3):319–340, 1989. ISSN 0276-7783. doi: 10.2307/249008.
- Juan de Dios Ortúzar and Luis G Willumsen. *Modelling transport*. John Wiley & sons, 2011.
- Anind K Dey. Understanding and using context. *Personal and ubiquitous computing*, 5: 4–7, 2001.
- Ravi Dhar, Stephen M. Nowlis, and Steven J. Sherman. Trying hard or hardly trying: An analysis of context effects in choice. *Journal of Consumer Psychology*, 9(4):189–200, 2000. ISSN 1057-7408. doi: https://doi.org/10.1207/S15327663JCP0904_1. URL <https://www.sciencedirect.com/science/article/pii/S1057740800703378>.
- Adele Diederich. Mdft account of decision making under time pressure. *Psychonomic Bulletin & Review*, 10(1):157–166, 2003.
- Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3):1155–1170, 2018. doi: 10.1287/mnsc.2016.2643. URL <https://doi.org/10.1287/mnsc.2016.2643>.
- Min Ding, John R Hauser, Songting Dong, Daria Dzyabura, Zhilin Yang, SU Chenting, and Steven P Gaskin. Unstructured direct elicitation of decision rules. *Journal of Marketing Research*, 48(1):116–127, 2011.
- Jeffrey P. Dotson, John R. Howell, Jeff D. Brazell, Thomas Otter, Peter J. Lenk, Steve MacEachern, and Greg M. Allenby. A probit model with structured covariance for similarity effects and source of volume calculations. *Journal of Marketing Research*, 55(1):35–47, 2018. doi: 10.1509/jmr.13.0240. URL <https://doi.org/10.1509/jmr.13.0240>.

- Katharina Dowling, Daniel Guhl, Daniel Klapper, Martin Spann, Lucas Stich, and Narine Yegoryan. Behavioral biases in marketing. *Journal of the Academy of Marketing Science*, 48(3):449–477, 2020. doi: 10.1007/s11747-019-00699-x. URL <https://doi.org/10.1007/s11747-019-00699-x>.
- Ioannis Evangelidis, Jonathan Levav, and Itamar Simonson. The asymmetric impact of context on advantaged versus disadvantaged options. *Journal of Marketing Research*, 55(2):239–253, 2018. doi: 10.1509/jmr.14.0483. URL <https://doi.org/10.1509/jmr.14.0483>.
- Nathan J Evans, William R Holmes, and Jennifer S Trueblood. Response-time data provide critical constraints on dynamic models of multi-alternative, multi-attribute choice. *Psychonomic Bulletin & Review*, 26:901–933, 2019.
- Mauricio S Featherman and Paul A Pavlou. Predicting e-services adoption: a perceived risk facets perspective. *International journal of human-computer studies*, 59(4):451–474, 2003.
- Fishburn. Exceptional paper-lexicographic orders, utilities and decision rules: a survey. *Management Science*, 20(11):1442–1471, 1974a.
- Peter C Fishburn. Exceptional paper—lexicographic orders, utilities and decision rules: A survey. *Management science*, 20(11):1442–1471, 1974b.
- Shane Frederick, Leonard Lee, and Ernest Baskin. The limits of attraction. *Journal of Marketing Research*, 51(4):487–507, 2014. doi: 10.1509/jmr.12.0061. URL <https://doi.org/10.1509/jmr.12.0061>.
- Brendan J. Frey and Delbert Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972–976, 2007. doi: 10.1126/science.1136800. URL <https://www.science.org/doi/abs/10.1126/science.1136800>.
- Xavier Gabaix and David Laibson. A boundedly rational decision algorithm. *American Economic Review*, 90(2):433–438, 2000.

- Chongming Gao, Wenqiang Lei, Xiangnan He, Maarten de Rijke, and Tat-Seng Chua. Advances and challenges in conversational recommender systems: A survey. *AI Open*, 2:100–126, 2021.
- Steven Gaskin, Theodoros Evgeniou, Daniel Bailiff, and John Hauser. Two-stage models: Identifying non-compensatory heuristics for the consideration set then adaptive polyhedral methods within the consideration set. In *Proceedings of the Sawtooth Software Conference*, volume 13, pages 67–83. Citeseer, 2007.
- David Gefen, Elena Karahanna, and Detmar W Straub. Trust and tam in online shopping: An integrated model. *MIS quarterly*, pages 51–90, 2003.
- Dennis H Gensch. A two-stage disaggregate attribute choice model. *Marketing Science*, 6(3):223–239, 1987.
- Dennis H Gensch and Wilfred W Recker. The multinomial, multiattribute logit choice model. *Journal of Marketing Research*, 16(1):124–132, 1979.
- Nejc Geržinič, Sander van Cranenburgh, Oded Cats, Emily Lancsar, and Caspar Chorus. Estimating decision rule differences between ‘best’ and ‘worst’ choices in a sequential best worst discrete choice experiment. *Journal of Choice Modelling*, 41:100307, 2021. ISSN 1755-5345. doi: <https://doi.org/10.1016/j.jocm.2021.100307>. URL <https://www.sciencedirect.com/science/article/pii/S1755534521000403>.
- Timothy J Gilbride and Greg M Allenby. A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science*, 23(3):391–406, 2004.
- Qi Gu, Jian Cao, Yafeng Zhao, and Yudong Tan. Addressing the Cold-Start Problem in Personalized Flight Ticket Recommendation. *IEEE Access*, 7:67178–67189, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2918210.
- Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5):1–42, 2018.
- Ido Guy, Naama Zwerdling, David Carmel, Inbal Ronen, Erel Uziel, Sivan Yogev, and Shila Ofek-Koifman. Personalized recommendation of social software items based on

- social relations. In *Proceedings of the third ACM conference on Recommender systems*, pages 53–60, 2009.
- Gerald Häubl and Valerie Trifts. Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing science*, 19(1):4–21, 2000.
- John R Hauser. Testing the accuracy, usefulness, and significance of probabilistic choice models: An information-theoretic approach. *Operations Research*, 26(3):406–421, 1978.
- John R Hauser. Consideration-set heuristics. *Journal of Business Research*, 67(8):1688–1699, 2014.
- John R Hauser and Birger Wernerfelt. An evaluation cost model of consideration sets. *Journal of consumer research*, 16(4):393–408, 1990.
- John R Hauser, Min Ding, and Steven P Gaskin. Non-compensatory (and compensatory) models of consideration-set decisions. In *Proceedings of the Sawtooth Software Conference*, volume 14, pages 207–232, 2009.
- Chen He, Denis Parra, and Katrien Verbert. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications*, 56:9–27, 2016.
- Jonathan L Herlocker, Joseph A Konstan, and John Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 241–250, 2000.
- Kaisa Herne. Decoy alternatives in policy choices: Asymmetric domination and compromise effects. *European Journal of Political Economy*, 13(3):575–589, 1997.
- Yoshinori Hijikata, Yuki Kai, and Shogo Nishida. The relation between user intervention and user satisfaction for information recommendation. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, pages 2002–2007, 2012.
- Yoshinori Hijikata, Yuki Kai, and Shogo Nishida. A Study of User Intervention and User Satisfaction in Recommender Systems. *Journal of Information Processing*, 22(4):669–678, 2014. ISSN 1882-6652. doi: 10.2197/ipsjjip.22.669.

- Mireille Hildebrandt. The issue of proxies and choice architectures. why eu law matters for recommender systems. *Frontiers in Artificial Intelligence*, page 73, 2022.
- Robin M Hogarth and Howard Kunreuther. Risk, ambiguity, and insurance. *Journal of risk and uncertainty*, 2:5–35, 1989.
- Joel L Horowitz and Jordan J Louviere. What is the role of consideration sets in choice modeling? *international Journal of Research in Marketing*, 12(1):39–54, 1995.
- Jared M Hotaling and Jörg Rieskamp. A quantitative test of computational models of multialternative context effects. *Decision*, 6(3):201, 2019.
- Andrew Howes, Paul A Warren, George Farmer, Wael El-Deredy, and Richard L Lewis. Why contextual preference reversals maximize expected value. *Psychological review*, 123(4):368, 2016.
- Joel Huber and Christopher Puto. Market Boundaries and Product Choice: Illustrating Attraction and Substitution Effects. *Journal of Consumer Research*, 10(1):31–44, 06 1983. ISSN 0093-5301. doi: 10.1086/208943. URL <https://doi.org/10.1086/208943>.
- Joel Huber, John W. Payne, and Christopher Puto. Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research*, 9(1):90–98, 06 1982. ISSN 0093-5301. doi: 10.1086/208899. URL <https://doi.org/10.1086/208899>.
- Kurt Jacobson, Vidhya Murali, Edward Newett, Brian Whitman, and Romain Yon. Music personalization at spotify. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 373–373, 2016.
- Jacob Jacoby, Donald E Speller, and Carol A Kohn. Brand choice behavior as a function of information load. *Journal of marketing research*, 11(1):63–69, 1974.
- Anthony Jameson and Eric Schwarzkopf. Pros and cons of controllability: An empirical study. In *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 193–202. Springer, 2002.

- Dietmar Jannach, Sidra Naveed, and Michael Jugovac. User control in recommender systems: Overview and interaction challenges. In *E-Commerce and Web Technologies: 17th International Conference, EC-Web 2016, Porto, Portugal, September 5-8, 2016, Revised Selected Papers 17*, pages 21–33. Springer, 2017.
- Dietmar Jannach, Michael Jugovac, and Ingrid Nunes. Explanations and user control in recommender systems. In *Proceedings of the 23rd International Workshop on Personalization and Recommendation on the Web and Beyond*, pages 31–31, 2019.
- Yucheng Jin, Nava Tintarev, and Katrien Verbert. Effects of personal characteristics on music recommender systems with different levels of controllability. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pages 13–21, Vancouver British Columbia Canada, September 2018. ACM. ISBN 978-1-4503-5901-6. doi: 10.1145/3240323.3240358.
- Eric J Johnson, Robert J Meyer, and Sanjoy Ghose. When choice models fail: Compensatory models in negatively correlated environments. *Journal of Marketing Research*, 26(3):255–270, 1989.
- Daniel Kahneman and Amos Tversky. On the interpretation of intuitive probability: A reply to jonathan cohen. 1979a.
- Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979b. ISSN 00129682, 14680262. URL <http://www.jstor.org/stable/1914185>.
- Manohar U Kalwani and Chi Kin Yim. Consumer price and promotion expectations: An experimental study. *Journal of marketing Research*, 29(1):90–100, 1992.
- Wagner A. Kamakura and Rajendra K. Srivastava. Predicting choice shares under conditions of brand interdependence. *Journal of Marketing Research*, 21(4):420–434, 1984. doi: 10.1177/002224378402100407. URL <https://doi.org/10.1177/002224378402100407>.
- Marius Kaminskis and Derek Bridge. Diversity, serendipity, novelty, and coverage: a

- survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(1):1–42, 2016.
- Boonchai Kijsanayotin, Supasit Pannarunothai, and Stuart M Speedie. Factors influencing health information technology adoption in thailand’s community health centers: applying the utaut model. *International journal of medical informatics*, 78(6):404–416, 2009.
- Jun B Kim, Paulo Albuquerque, and Bart J Bronnenberg. The probit choice model under sequential search with an application to online retailing. *Management Science*, 63(11):3911–3929, 2017.
- Sunghan Kim and Lynn Hasher. The attraction effect in decision making: Superior performance by older adults. *The Quarterly Journal of Experimental Psychology Section A*, 58(1):120–133, 2005.
- Julia Kiseleva, Alexander Tuzhilin, Jaap Kamps, Melanie JI Mueller, Lucas Bernardi, Chad Davis, Ivan Kovacek, Mats Stafseng Einarsen, and Djoerd Hiemstra. Beyond movie recommendations: Solving the continuous cold start problem in e-commerce recommendations. *arXiv preprint arXiv:1607.07904*, 2016.
- Ran Kivetz, Oded Netzer, and V. Srinivasan. Alternative models for capturing the compromise effect. *Journal of Marketing Research*, 41(3):237–257, 2004. doi: 10.1509/jmkr.41.3.237.35990. URL <https://doi.org/10.1509/jmkr.41.3.237.35990>.
- Daniel Kluver and Joseph A Konstan. Evaluating recommender behavior for new users. In *Proceedings of the 8th ACM Conference on Recommender Systems*, pages 121–128, 2014.
- Jack L Knetsch. The endowment effect and evidence of nonreversible indifference curves. *The american Economic review*, 79(5):1277–1284, 1989.
- Bart P Knijnenburg, Niels JM Reijmer, and Martijn C Willemsen. Each to his own: how different users call for different interaction methods in recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 141–148, 2011.

- Bart P. Knijnenburg, Martijn C. Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5):441–504, October 2012. ISSN 0924-1868, 1573-1391. doi: 10.1007/s11257-011-9118-4.
- Sören Köcher, Michael Jugovac, Dietmar Jannach, and Hartmut H. Holzmüller. New hidden persuaders: An investigation of attribute-level anchoring effects of product recommendations. *Journal of Retailing*, 95(1):24–41, 2019. ISSN 0022-4359. doi: <https://doi.org/10.1016/j.jretai.2018.10.004>. URL <https://www.sciencedirect.com/science/article/pii/S0022435918300630>.
- Denis Kotkov, Shuaiqiang Wang, and Jari Veijalainen. A survey of serendipity in recommender systems. *Knowledge-Based Systems*, 111:180–192, November 2016. ISSN 09507051. doi: 10.1016/j.knosys.2016.08.014.
- Abhishek Kumar, Pankaj Kumar, and Shalini Varshney. Movie recommendation models using machine learning. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 491–496, 2018. doi: 10.1109/ICMLA.2018.00086.
- Caroline Lancelot Miltgen, Aleš Popovič, and Tiago Oliveira. Determinants of end-user acceptance of biometrics: Integrating the “Big 3” of technology acceptance with privacy context. *Decision Support Systems*, 56:103–114, December 2013. ISSN 01679236. doi: 10.1016/j.dss.2013.05.010.
- Byung-Kwan Lee and Wei-Na Lee. The effect of information overload on consumer choice quality in an on-line environment. *Psychology & Marketing*, 21(3):159–183, 2004.
- Ya-Ching Lee. An empirical investigation into factors influencing the adoption of an e-learning system. *Online information review*, 30(5):517–541, 2006.
- Alix Lhéritier, Michael Bocamazo, Thierry Delahaye, and Rodrigo Acuna-Agost. Airline itinerary choice modeling using machine learning. *Journal of choice modelling*, 31: 198–209, 2019.
- Sarah Lichtenstein and Paul Slovic. The construction of preference: An overview. *The construction of preference*, 1, 2006.

- Shi Xian Liew, Piers DL Howe, and Daniel R Little. The appropriacy of averaging in the study of context effects. *Psychonomic bulletin & review*, 23:1639–1646, 2016.
- Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. Facing the cold start problem in recommender systems. *Expert Systems with Applications*, 41(4):2065–2073, 2014.
- Chang-Ming Lin, Chun-Yin Wu, Ko-Ying Tseng, Chih-Chiang Ku, and Sheng-Fuu Lin. Applying two-stage differential evolution for energy saving in optimal chiller loading. *Energies*, 12(4):622, 2019.
- Qing Liu and Neeraj Arora. Efficient choice designs for a consider-then-choose model. *Marketing Science*, 30(2):321–338, 2011.
- S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982. doi: 10.1109/TIT.1982.1056489.
- Roy Lubit. Tacit knowledge and knowledge management: The keys to sustainable competitive advantage. *Organizational dynamics*, 2001.
- R. Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. John Wiley & Sons., New York, NY, 1959.
- R Duncan Luce. *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2012.
- Pattie Maes, Robert H Guttman, and Alexandros G Moukas. Agents that buy and sell. *Communications of the ACM*, 42(3):81–ff, 1999.
- Mo Adam Mahmood, Janice M Burn, Leopoldo A Gemoets, and Carmen Jacquez. Variables affecting information technology end-user satisfaction: a meta-analysis of the empirical literature. *International Journal of Human-Computer Studies*, 52(4):751–771, 2000.
- Tariq Mahmood and Francesco Ricci. Improving recommender systems with adaptive conversational strategies. In *Proceedings of the 20th ACM conference on Hypertext and hypermedia*, pages 73–82, 2009.

- Oded Maimon and Lior Rokach. Data mining and knowledge discovery handbook. 2005.
- Naresh K Malhotra. Information load and consumer decision making. *Journal of consumer research*, 8(4):419–430, 1982.
- Ajay K Manrai and Prabhakant Sinha. Elimination-by-cutoffs. *Marketing Science*, 8(2): 133–152, 1989.
- Rafael G Mantovani, André LD Rossi, Edesio Alcobaça, Joaquin Vanschoren, and André CPLF de Carvalho. A meta-learning recommender system for hyperparameter tuning: Predicting when tuning improves svm classifiers. *Information Sciences*, 501: 193–221, 2019.
- Nikola Marangunić and Andrina Granić. Technology acceptance model: a literature review from 1986 to 2013. *Universal access in the information society*, 14:81–95, 2015.
- Daniel McFadden. Economic choices. *American Economic Review*, 91(3):351–378, June 2001. doi: 10.1257/aer.91.3.351. URL <https://www.aeaweb.org/articles?id=10.1257/aer.91.3.351>.
- Daniel McFadden et al. Conditional logit analysis of qualitative choice behavior. 1973.
- Barbara A Mellers and Alan DJ Cooke. Trade-offs depend on attribute range. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5):1055, 1994.
- Martijn Millecamp, Nyi Nyi Htun, Yucheng Jin, and Katrien Verbert. Controlling Spotify Recommendations: Effects of Personal Characteristics on Music Recommender User Interfaces. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, pages 101–109, Singapore Singapore, July 2018. ACM. ISBN 978-1-4503-5589-6. doi: 10.1145/3209219.3209223.
- Peter NC Mohr, Hauke R Heekeren, and Jörg Rieskamp. Attraction effect in risky choice can be explained by subjective distance between choice alternatives. *Scientific reports*, 7(1):1–10, 2017.
- Judith Möller, Damian Trilling, Natali Helberger, and Bram van Es. Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their

- impact on content diversity. *Information, Communication & Society*, 21(7):959–977, July 2018. ISSN 1369-118X, 1468-4462. doi: 10.1080/1369118X.2018.1444076.
- Alejandro Mottini and Rodrigo Acuna-Agost. Deep choice model using pointer networks for airline itinerary prediction. In *Proceedings of the 54th Hawaii International Conference on System Sciences*, volume 23, pages 1575–1583, 2017.
- Takao Noguchi and Neil Stewart. In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition*, 132(1):44–56, 2014.
- Takao Noguchi and Neil Stewart. Multialternative decision by sampling: A model of decision making constrained by process data. *Psychological review*, 125(4):512, 2018.
- Donald A Norman. *The design of everyday things*. Basic books, 2013.
- Javad Omidi and Karim Mazaheri. Differential evolution algorithm for performance optimization of the micro plasma actuator as a microelectromechanical system. *Scientific Reports*, 10(1):18865, 2020.
- Mahamed GH Omran, Andries P Engelbrecht, and Ayed Salman. Bare bones differential evolution. *European Journal of Operational Research*, 196(1):128–139, 2009.
- A. Yeşim Orhun. Optimal product line design when consumers exhibit choice set-dependent preferences. *Marketing Science*, 28(5):868–886, 2009. doi: 10.1287/mksc.1080.0449. URL <https://doi.org/10.1287/mksc.1080.0449>.
- Seung-Taek Park and Wei Chu. Pairwise preference regression for cold-start recommendation. In *Proceedings of the third ACM conference on Recommender systems*, pages 21–28, 2009.
- Marcel Paulssen and Richard P Bagozzi. A self-regulatory model of consideration set formation. *Psychology & Marketing*, 22(10):785–812, 2005.
- John W Payne, James R Bettman, and Eric J Johnson. Adaptive strategy selection in decision making. *Journal of experimental psychology: Learning, Memory, and Cognition*, 14(3):534, 1988.

- Michael J Pazzani and Daniel Billsus. Content-based recommendation systems. In *The Adaptive Web*, page 325–341. Springer, 2007.
- Karlson Pfannschmidt, Pritha Gupta, and Eyke Hüllermeier. Learning choice functions: Concepts and architectures. *arXiv: 1901.10860*, 2019.
- Ivens Portugal, Paulo Alencar, and Donald Cowan. The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, 97:205–227, 2018.
- Pearl Pu, Li Chen, and Rong Hu. Evaluating recommender systems from the user’s perspective: survey of the state of the art. *User Modeling and User-Adapted Interaction*, 22(4):317–355, 2012.
- Girish Punj and Robert Moore. Information search and consideration set formation in a web-based store environment. *Journal of Business Research*, 62(6):644–650, 2009.
- Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor. *Recommender Systems Handbook*. Springer-Verlag, Berlin, Heidelberg, 1st edition, 2010. ISBN 0387858199.
- Francesco Ricci, Lior Rokach, and Bracha Shapira. Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer, 2011.
- Francesco Ricci, Lior Rokach, and Bracha Shapira. Recommender systems: introduction and challenges. *Recommender systems handbook*, pages 1–34, 2015.
- Robert M Roe, Jermone R Busemeyer, and James T Townsend. Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological review*, 108(2):370, 2001.
- Lior Rokach and Oded Maimon. Clustering methods. In *Data mining and knowledge discovery handbook*, pages 321–352. Springer, 2005.
- Robert P. Rooderkerk, Harald J. Van Heerde, and Tammo H.A. Bijmolt. Incorporating context effects into a choice model. *Journal of Marketing Research*, 48(4):767–780, 2011. doi: 10.1509/jmkr.48.4.767. URL <https://doi.org/10.1509/jmkr.48.4.767>.

- Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- Amina SAMIH, Abderrahim GHADI, and Abdelhadi FENNAN. Exmrec2vec: explainable movie recommender system based on word2vec. *International Journal of Advanced Computer Science and Applications*, 12(8), 2021.
- Martin Saveski and Amin Mantrach. Item cold-start recommendations: learning local collective embeddings. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 89–96, 2014.
- J Ben Schafer, Joseph Konstan, and John Riedl. Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce*, pages 158–166, 1999.
- J Ben Schafer, Joseph A Konstan, and John Riedl. Meta-recommendation systems: user-controlled integration of diverse recommendations. In *CIKM '02*, pages 43–51. ACM, 2002.
- Jonathan Schaffer, Tobias Höllerer, and John O'Donovan. Hypothetical recommendation: A study of interactive profile manipulation behavior for recommender systems. In *FLAIRS '15*, pages 507–512, 2015.
- Andrew I Schein, Alexandrin Popescul, Lyle H Ungar, and David M Pennock. Methods and Metrics for Cold-Start Recommendations. page 8, 2002.
- Karin Schermelleh-Engel, Helfried Moosbrugger, Hans Müller, et al. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online*, 8(2):23–74, 2003.
- James B Schreiber, Amaury Nora, Frances K Stage, Elizabeth A Barlow, and Jamie King. Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of educational research*, 99(6):323–338, 2006.
- Randall E Schumacker and Richard G Lomax. *A beginner's guide to structural equation modeling*. psychology press, 2004.

- Barry Schwartz. *The paradox of choice: Why more is less*. HarperCollins Publishers, 2004.
- A Schwarz and WW Chin. Looking forward: toward an understanding of the nature and definition of it acceptance. *Journal of the Association for Information Systems*, 8(4): 244–273, 2007.
- Shibin Sheng, Andrew M Parker, and Kent Nakamoto. Understanding the mechanism and determinants of compromise effects. *Psychology & Marketing*, 22(7):591–609, 2005.
- Hung-Pin Shih. Extended technology acceptance model of internet utilization behavior. *Information & management*, 41(6):719–729, 2004.
- Allan D Shocker, Moshe Ben-Akiva, Bruno Boccara, and Prakash Nedungadi. Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing letters*, 2(3):181–197, 1991.
- Mark S Silver. User perceptions of decision support system restrictiveness: An experiment. *Journal of Management Information Systems*, 5(1):51–65, 1988.
- Herbert A Simon. Rational choice and the structure of the environment. *Psychological review*, 63(2):129, 1956.
- Itamar Simonson. Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, 16(2):158–174, 09 1989. ISSN 0093-5301. doi: 10.1086/209205. URL <https://doi.org/10.1086/209205>.
- Alireza Soltani, Benedetto De Martino, and Colin Camerer. A range-normalization model of context-dependent choice: a new model and evidence. *PLoS computational biology*, 8(7):e1002607, 2012.
- Yicheng Song, Nachiketa Sahoo, and Elie Ofek. When and How to Diversify—A Multicategory Utility Model for Personalized Content Recommendation. *Management Science*, 65(8):3737–3757, August 2019. ISSN 0025-1909, 1526-5501. doi: 10.1287/mnsc.2018.3127.

- Mikhail S Spektor, David Kellen, and Jared M Hotaling. When the good looks bad: An experimental exploration of the repulsion effect. *Psychological science*, 29(8):1309–1320, 2018.
- Mikhail S Spektor, Sebastian Gluth, Laura Fontanesi, and Jörg Rieskamp. How similarity between choice options affects decisions from experience: The accentuation-of-differences model. *Psychological review*, 126(1):52, 2019.
- Neil Stewart, Nick Chater, and Gordon DA Brown. Decision by sampling. *Cognitive psychology*, 53(1):1–26, 2006.
- Rainer Storn and Kenneth Price. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341, 1997.
- Kirsten Swearingen and Rashmi Sinha. Beyond algorithms: An hci perspective on recommender systems. In *ACM SIGIR 2001 workshop on recommender systems*, pages 1–11. Citeseer, 2001.
- Latanya Sweeney. k-anonymity: a model for protecting privacy. *International Journal on uncertainty, fuzziness and knowledge-based systems*, 10(5):557–570, 2002.
- Timothy Teo. Students and teachers’ intention to use technology: Assessing their measurement equivalence and structural invariance. *Journal of Educational Computing Research*, 57(1):201–225, 2019.
- Andrei R Teodorescu and Marius Usher. Disentangling decision models: from independence to competition. *Psychological review*, 120(1):1, 2013.
- Nava Tintarev and Judith Masthoff. Explaining recommendations: Design and evaluation. *Recommender systems handbook*, pages 353–382, 2015.
- Nava Tintarev, Shahin Rostami, and Barry Smyth. Knowing the unknown: Visualising consumption blind-spots in recommender systems. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, pages 1396–1399, Pau France, April 2018. ACM. ISBN 978-1-4503-5191-1. doi: 10.1145/3167132.3167419.

- Kiran Tomlinson and Austin R. Benson. Learning interpretable feature context effects in discrete choice. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, KDD '21*, pages 1582–1592, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383325. doi: 10.1145/3447548.3467250. URL <https://doi.org/10.1145/3447548.3467250>.
- Jennifer S. Trueblood, Scott D. Brown, Andrew Heathcote, and Jerome R. Busemeyer. Not just for consumers: Context effects are fundamental to decision making. *Psychological Science*, 24(6):901–908, 2013. doi: 10.1177/0956797612464241. URL <https://doi.org/10.1177/0956797612464241>. PMID: 23610134.
- Jennifer S Trueblood, Scott D Brown, and Andrew Heathcote. The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological review*, 121(2):179, 2014.
- Jennifer S Trueblood, Scott D Brown, Andrew Heathcote, and Jennifer Trueblood. The fragile nature of contextual preference reversals: Reply to tsetsos, chater, and usher. *Psychological Review*, 2015.
- Amos Tversky. Elimination by aspects: A theory of choice. *Psychological review*, 79(4):281, 1972.
- Amos Tversky and Shmuel Sattath. Preference trees. *Psychological Review*, 86(6):542, 1979.
- Amos Tversky and Itamar Simonson. Context-dependent preferences. *Management Science*, 39(10):1179–1189, 1993. doi: 10.1287/mnsc.39.10.1179. URL <https://doi.org/10.1287/mnsc.39.10.1179>.
- Glen L Urban, Philip L Johnson, and John R Hauser. Testing competitive market structures. *Marketing Science*, 3(2):83–112, 1984.
- Marius Usher and James L McClelland. The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, 108(3):550, 2001.
- Viswanath Venkatesh. Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. *Information*

- Systems Research*, 11(4):342–365, December 2000. ISSN 1047-7047, 1526-5536. doi: 10.1287/isre.11.4.342.11872.
- Viswanath Venkatesh and Hillol Bala. Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, 39(2):273–315, 2008. ISSN 1540-5915. doi: 10.1111/j.1540-5915.2008.00192.x.
- Viswanath Venkatesh, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3): 425–478, 2003. ISSN 02767783. URL <http://www.jstor.org/stable/30036540>.
- Jari Vesanen. What is personalization? a conceptual framework. *European Journal of Marketing*, 41(5/6):409–418, 2007.
- Chong Wang and David M Blei. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, page 448–456. ACM, 2011.
- Wei-quan Wang and Izak Benbasat. Attributions of trust in decision support technologies: A study of recommendation agents for e-commerce. *Journal of Management Information Systems*, 24(4):249–273, 2008.
- Wei-quan Wang and Izak Benbasat. Interactive decision aids for consumer decision making in e-commerce: The influence of perceived strategy restrictiveness. *MIS quarterly*, pages 293–320, 2009.
- Douglas H Wedell. Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(4): 767, 1991.
- Michael D Williams, Yogesh K Dwivedi, Banita Lal, and Andrew Schwarz. Contemporary trends and issues in it adoption and diffusion research. *Journal of Information Technology*, 24(1):1–10, 2009.
- Michael D Williams, Nripendra P Rana, and Yogesh K Dwivedi. The unified theory of acceptance and use of technology (utaut): a literature review. *Journal of enterprise information management*, 28(3):443–488, 2015.

- Raymond K Wong, Victor W Chu, and Tianyong Hao. Online role mining for context-aware mobile service recommendation. *Personal and Ubiquitous Computing*, 18(5): 1029–1046, 2014.
- Chunhua Wu and Koray Cosguner. Profiting from the decoy effect: A case study of an online diamond retailer. *Marketing Science*, 39(5):974–995, 2020. doi: 10.1287/mksc.2020.1231. URL <https://doi.org/10.1287/mksc.2020.1231>.
- Jie Wu, Yijun Wang, and Hsinchun Lin. An empirical study of mobile commerce in rural china. *International Journal of Electronic Commerce*, 11(4):57–88, 2007.
- Bo Xiao and Izak Benbasat. E-commerce product recommendation agents: Use, characteristics, and impact. *MIS quarterly*, pages 137–209, 2007.
- Xinyu Xing, Wei Meng, Dan Doozan, Alex C Snoeren, Nick Feamster, and Wenke Lee. Take this personally: Pollution attacks on personalized services. In *USENIX Security Symposium*, pages 671–686, 2013.
- Michael Yee, Ely Dahan, John R Hauser, and James Orlin. Greedoid-based noncompensatory inference. *Marketing Science*, 26(4):532–549, 2007.
- Reza Yousefi Maragheh, Xin Chen, James Davis, Jason Cho, Sushant Kumar, and Kannan Achan. Choice modeling and assortment optimization in the presence of context effects. *Available at SSRN 3747354*, 2020.
- Zi-Ke Zhang, Chuang Liu, Yi-Cheng Zhang, and Tao Zhou. Solving the cold-start problem in recommender systems with social tags. *EPL (Europhysics Letters)*, 92(2):28002, 2010.

7 Appendices

A

Differential evolution

The Differential Evolution Algorithm developed by Storn (1997) is a heuristic-based optimization technique that is widely used in various scientific fields, including economics. This algorithm is capable of exploring the search space of a problem to find the global optimum. Unlike traditional optimization techniques, the DE algorithm does not require gradient information, making it applicable to nondifferentiable, nonlinear, and multimodal optimization problems.

Differential evolution is a population-based stochastic search algorithm that generates new candidate solutions by combining existing individuals in the population. The algorithm considers the differences between their parameter values, hence the term “differential”. It includes three primary operations: mutation, crossover, and selection.

The algorithm operates as follows:

1. Initialization: Generate an initial population of NP candidate solutions randomly, each solution is a vector in the D -dimensional space

For $i = 1$ to $NP, j = 1$ to D :

$$X[i, j] = X_{min}[j] + rand(0, 1) * (X_{max}[j] - X_{min}[j]).$$

2. Mutation: For each individual vector $X_i (i = 1, 2, \dots, NP)$ in the current population, a mutant vector V_i is generated according to:

$$V[i, j] = X[r1, j] + F * (X[r2, j] - X[r3, j]).$$

Here, $r1, r2,$ and $r3$ are indices randomly chosen from the population, and they are distinct from each other and i . The real number $F \in [0, 2]$ is the scaling factor which controls the amplification of the differential variation $X[r2, j] - X[r3, j]$.

3. Crossover: A trial vector U_i is then generated from the original vector X_i and mutant vector V_i according to the rule:

For $j = 1$ to D :

$$U[i, j] = \begin{cases} V[i, j] & \text{if } \text{rand}(0, 1) \leq CR \text{ or } j = j_{\text{rand}} \\ X[i, j] & \text{otherwise} \end{cases} .$$

Here, $\text{rand}(0, 1)$ is a uniform random number in $[0, 1]$, $CR \in [0, 1]$ is the crossover rate that controls the fraction of parameter values copied from the mutant vector, and j_{rand} is a randomly chosen index from the D dimensions.

4 Selection: The trial vector U_i competes against the original individual X_i . The one that provides a better objective function value (lower for a minimization problem, higher for a maximization problem) survives into the next generation:

$$X[i, j] = \begin{cases} U[i, j] & \text{if } f(U[i, j]) \leq f(X[i, j]) \\ X[i, j] & \text{otherwise} \end{cases} .$$

Here, $f(\cdot)$ denotes the objective function.

5 Loop: Steps 2 to 4 are repeated until a termination criterion is met (such as a maximum number of generations or a satisfactory fitness level).

B

Comparison pair count calculation for the compromise effect

Given a setting with N vertical attributes, there are a number of ways that a focal option can act as a compromise between two groups of competing options. I discuss this case by case.

Case 1: No dimension is equal between focal and competing options. In this case, one group of options might be better than the focal option in one dimension and worse in $N - 1$ dimensions. The mirror image of this group would be a group that is better than the focal option in the $N - 1$ dimensions and worse in the same dimension. There are $\binom{N}{1}$ of such groups.

Another option within the same option could be a group that is better than the focal option in 2 dimensions and worse in $N - 2$ dimensions. There are $\binom{N}{2}$ such groups.

All in all, there are $N - 1$ such sub-cases. These sub-cases count distinct groups. As we have to pair these groups, we divide the number by two.

Therefore, for this case, the number of comparisons is

$$\frac{1}{2} \sum_{a=1}^{N-1} \binom{N}{a}.$$

Case 2: Only one dimension is equal in all competing options, including the focal option. In this case, we are comparing not N , but $N - 1$ options. Therefore, for each given dimension that is equal between options, we have

$$\frac{1}{2} \sum_{a=1}^{N-2} \binom{N-1}{a}$$

comparisons. However, not only one, but each of the N vertical attributes can be equal across all options. Therefore, the total number of comparisons in this case is

$$\frac{1}{2} \binom{N}{1} \sum_{a=1}^{N-2} \binom{N-1}{a}.$$

Case 3: Multiple dimensions are equal in all competing options. First, we extend the previous case to the situation where two dimensions are equal across all options, resulting

in

$$\frac{1}{2} \binom{N}{2} \sum_{a=1}^{N-3} \binom{N-2}{a}.$$

We iterate the same exercise until (and including) the setup where we have $N - 2$ dimensions equal across all options ²³.

The total of all comparisons will simply be the sum of all these cases, which can be expressed as

$$\Omega = \frac{1}{2} \sum_{b=0}^{N-2} \left[\binom{N}{b} \sum_{a=1}^{N-1-b} \binom{N-b}{a} \right]. \quad (17)$$

Figure ¹⁹ shows how Ω changes with N .

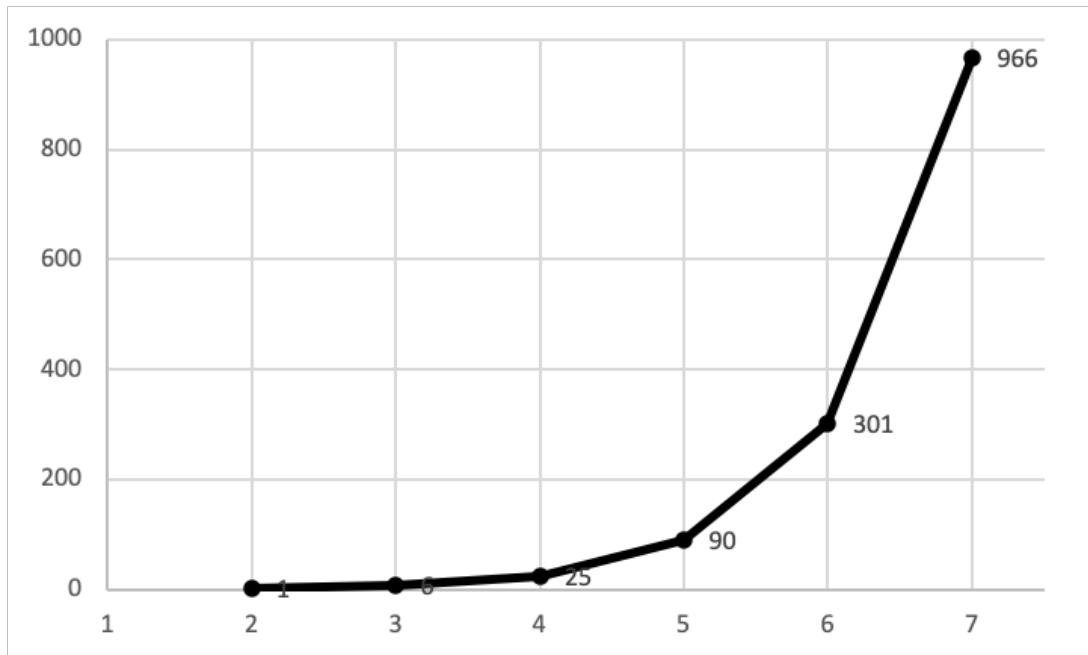


Figure 19: Correspondence between Ω and N .

²³One needs the minimum of two dimensions that can be compared across two comparable groups.

C

Clustering

The goal of clustering is to separate the data into different groups in a way that similar instances belong to the same group, while different instances are assigned to different groups (Maimon and Rokach, 2005). Formally, clustering consists in making a partition $C = C_1, C_2, \dots, C_k$ of some set S in a way that: $S = \cup_{i=1}^k C_i$ and $C_i \cap C_j = \emptyset$ for all $i \neq j$. So, any alternative in the set S belongs to exactly one cluster.

There exist many clustering methods, and each of them have a different way of defining similar items and as a result will group in different ways. Clustering is usually an unsupervised machine learning method in that there are no preconceived labels given to the clusters. This implies that there is no universal way of evaluating the quality of a clustering result. Moreover, most clustering methods require an additional input to determine the number of clusters. Therefore, in order to make sure that our results do not depend on the chosen clustering method, I investigate the two widely used methods, affinity propagation and k-means clustering.

Affinity propagation

The first clustering method, examined, is affinity propagation (Frey and Dueck, 2007). affinity propagation identifies a limited number of “exemplars”, which are identified as the best representative of other objects in the same cluster (“samples”). It calculates the pairwise values that characterize the suitability of one object to be the exemplar of the other. These values are updated in response to the values from other pairs. This updating happens in an iterative manner until convergence, at which point the final exemplars are chosen, and hence the final clustering is identified.

There are two characteristics involved in the process. The responsibility $r(i, k)$ that quantifies how suited k is as an exemplar of the cluster i compared to all other potential exemplars. It is calculated as

$$r(i, k) = s(i, k) - \max [a(i, k') + s(i, k') \text{ for all } k' \neq k], \quad (18)$$

where $s(i, k)$ is the similarity between i and k , measured as the squared negative error.

The second property is availability $a(i, k)$ which measures the extent to which i is an appropriate sample of k , given all the other samples already identified of k . This is calculated as

$$a(i, k) = \min \left[0, r(k, k) + \sum_{i' \text{ s.t. } i' \notin \{i, k\}} r(i', k) \right]. \quad (19)$$

At the start, both r and a are set to zero and the calculations are iterated until full convergence. To eliminate oscillations when updating the values, the damping factor λ is introduced to the iteration process. This facilitates the convergence process and alters the responsibility and availability equations as follows:

$$r_{t+1}(i, k) = \lambda \cdot r_t(i, k) + (1 - \lambda) \cdot r_{t+1}(i, k), \quad (20)$$

$$a_{t+1}(i, k) = \lambda \cdot a_t(i, k) + (1 - \lambda) \cdot a_{t+1}(i, k). \quad (21)$$

The damping factor, $\lambda \in [0; 1]$, affects the number of identified clusters in affinity propagation procedure. After setting its value, the number of clusters in the data are automatically identified. Frey and Dueck (2007) recommend setting $\lambda \in [0.5; 1]$, in order to ensure the convergence in large datasets. I have experimented with the sensitivity of clustering outcomes with respect to damping factor and have found very little differences in the vicinity of the factor between 0.5 and 0.75. Therefore, I conform to the wide usage of $\lambda = 0.5$ (which also increases the convergence speed) for the rest of the paper.

Kmeans

A popular alternative method to identify clusters of comparable objects in data is the k-means algorithm (Lloyd, 1982). It divides a set of N objects of X into K distinct clusters C_j which are described by the means μ_j of the samples within each cluster. Those means are commonly referred to as cluster centroids. The algorithm aims to select centroids that minimize the within cluster sum of squares, i.e

$$\sum_{i=0}^n \min_{\mu_j \in C_j} (|x_i - \mu_j|^2).$$

At its core, k-means is a computationally cheap algorithm. However, it has an important drawback, which is that it explicitly requires the number of clusters to be fixed ex-ante. Different number of clusters result into different clustering outcomes for a given data. This is a disadvantage compared to affinity propagation where the number of clusters is endogenously identified. To solve this problem, there are external measures we can use in order to judge the optimality of clustering outcomes for a given data. One such popular and computationally affordable measure is the silhouette score (Rousseeuw, 1987). The silhouette score is calculated for every object in the set as

$$s = \frac{1}{n} \sum_{i=1}^n \frac{b - a}{\max(a, b)}, \quad (22)$$

where a is the mean distance between this object and all other objects in the same cluster, b is the mean distance between this object and all other objects from the nearest cluster, and n is the number of objects in the set. An important advantage of this method over alternatives is that it is confined to the interval $[-1; 1]$. The higher the silhouette score, the better is the cluster assignment. Then one can compute clusters for every feasible number of clusters, calculate silhouette score for each of the instances and choose the instance (i.e. number of identified clusters) with the maximal silhouette score. This drastically increases computational requirements for the k-means clustering as in this case one has to calculate cluster assignment for a set of potential cluster numbers to choose from.

D

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Price	-0.504*** (0.013)	-0.515*** (0.013)	-0.387*** (0.017)	-0.499*** (0.013)	-0.480*** (0.013)	-0.351*** (0.017)	-0.354*** (0.017)	-0.351*** (0.017)	-0.354*** (0.017)
Trip duration	-0.327*** (0.020)	-0.278*** (0.020)	-0.165*** (0.023)	-0.264*** (0.020)	-0.291*** (0.020)	-0.178*** (0.022)	-0.176*** (0.022)	-0.178*** (0.022)	-0.176*** (0.022)
Number of flights	-0.703*** (0.018)	-0.658*** (0.018)	-0.670*** (0.018)	-0.626*** (0.018)	-0.580*** (0.018)	-0.562*** (0.019)	-0.580*** (0.019)	-0.567*** (0.019)	-0.584*** (0.019)
Number of airlines	-0.473*** (0.023)	-0.474*** (0.023)	-0.430*** (0.024)	-0.475*** (0.023)	-0.461*** (0.023)	-0.420*** (0.023)	-0.431*** (0.023)	-0.418*** (0.024)	-0.430*** (0.023)
Attraction			0.013*** (0.001)		0.011*** (0.001)		0.012*** (0.001)		
Compromise				-0.128*** (0.011)		-0.110*** (0.011)	-0.101*** (0.011)		
Similarity					-0.050*** (0.004)	-0.051*** (0.004)	-0.082*** (0.005)	-0.050*** (0.004)	-0.081*** (0.005)
Attraction within cluster							0.070*** (0.005)		0.070*** (0.005)
Attraction outside cluster							0.009*** (0.001)		0.009*** (0.001)
Compromise within cluster								-0.404*** (0.055)	-0.365*** (0.055)
Compromise outside cluster								-0.092*** (0.013)	-0.088*** (0.013)
Constant included	YES	YES	YES	YES	YES	YES	YES	YES	YES
Horizontal variables as controls	NO	YES	YES	YES	YES	YES	YES	YES	YES
Number of observations	368723	368723	368723	368723	368723	368723	368723	368723	368723
Number of choices	6297	6297	6297	6297	6297	6297	6297	6297	6297
Consistent Akaike information criterion	49786	49124	49004	48924	48924	48643	48537	48643	48537
Log likelihood	-24858	-24465	-24398	-24358	-24358	-24204	-24144	-24197	-24137

Table 20: Outputs from logistic regressions with observational data.

Notes: Standard errors in parentheses. Statistical significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Price	-0.788*** (0.027)	-0.832*** (0.028)	-0.757*** (0.032)	-0.812*** (0.028)	-0.822*** (0.029)	-0.734*** (0.032)	-0.731*** (0.032)	-0.730*** (0.032)	-0.727*** (0.032)
Trip duration	-0.627*** (0.043)	-0.568*** (0.041)	-0.449*** (0.044)	-0.535*** (0.041)	-0.557*** (0.042)	-0.448*** (0.044)	-0.447*** (0.044)	-0.442*** (0.044)	-0.444*** (0.044)
Number of flights	-3.034*** (0.224)	-2.877*** (0.208)	-3.053*** (0.232)	-2.818*** (0.200)	-2.873*** (0.202)	-2.973*** (0.225)	-2.829*** (0.247)	-2.986*** (0.228)	-2.891*** (0.224)
Number of airlines	-0.677*** (0.096)	-0.720*** (0.091)	-0.692*** (0.087)	-0.749*** (0.090)	-0.758*** (0.089)	-0.640*** (0.066)	-0.671*** (0.095)	-0.629*** (0.079)	-0.712*** (0.080)
Attraction			0.009*** (0.002)		0.009*** (0.002)		0.008*** (0.002)		
Compromise				-0.042** (0.019)		-0.058*** (0.020)	-0.056*** (0.017)		
Similarity					-0.008 (0.006)	-0.007 (0.006)	-0.019*** (0.007)	-0.006 (0.006)	-0.016** (0.007)
Attraction within cluster									0.025*** (0.006)
Attraction outside cluster									0.007*** (0.002)
Compromise within cluster									-0.211*** (0.056)
Compromise outside cluster									-0.059** (0.027)
Constant included	YES	YES	YES	YES	YES	YES	YES	YES	YES
Horizontal variables as controls	NO	YES	YES	YES	YES	YES	YES	YES	YES
Number of observations	368723	368723	368723	368723	368723	368723	368723	368723	368723
Number of choices	6297	6297	6297	6297	6297	6297	6297	6297	6297
Consistent Akaike information criterion	34142	33275	33243	33273	33281	33249	33263	33251	33256
Log likelihood	-17043	-16596	-16573	-16588	-16592	-16562	-16562	-16556	-16552

Table 21: Outputs from midex Logit regressions with observational data.

Note: Standard errors in parentheses. Statistical significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Price	-0.225*** (0.006)	-0.232*** (0.006)	-0.201*** (0.008)	-0.227*** (0.006)	-0.231*** (0.006)	-0.200*** (0.008)	-0.203*** (0.008)	-0.200*** (0.008)	-0.203*** (0.008)
Trip duration	-0.136*** (0.009)	-0.116*** (0.009)	-0.085*** (0.010)	-0.112*** (0.009)	-0.116*** (0.009)	-0.087*** (0.010)	-0.090*** (0.010)	-0.086*** (0.010)	-0.090*** (0.010)
Number of flights	-0.342*** (0.008)	-0.322*** (0.008)	-0.324*** (0.008)	-0.309*** (0.008)	-0.320*** (0.008)	-0.308*** (0.009)	-0.310*** (0.008)	-0.310*** (0.009)	-0.312*** (0.009)
Number of airlines	-0.208*** (0.010)	-0.209*** (0.010)	-0.193*** (0.010)	-0.210*** (0.010)	-0.209*** (0.010)	-0.196*** (0.011)	-0.198*** (0.011)	-0.196*** (0.011)	-0.198*** (0.011)
Attraction			0.003*** (0.001)		0.003*** (0.001)		0.003*** (0.001)		
Compromise				-0.038*** (0.003)		-0.037*** (0.003)			
Similarity					-0.001 (0.000)	-0.001* (0.000)	-0.002*** (0.001)	-0.000 (0.001)	-0.002*** (0.001)
Attraction within cluster							0.005***		0.005***
Attraction outside cluster						(0.001)		(0.001)	0.002***
Compromise within cluster						(0.001)		(0.001)	-0.057***
Compromise outside cluster									(0.008)
Constant included	YES	YES	YES	YES	YES	YES	YES	YES	YES
Horizontal variables as controls	NO	YES	YES	YES	YES	YES	YES	YES	YES
Number of observations	368723	368723	368723	368723	368723	368723	368723	368723	368723
Number of choices	6297	6297	6297	6297	6297	6297	6297	6297	6297
Consistent Akaike information criterion	49598	48910	48886	48770	48922	48765	48765	48787	48789
Log likelihood	-24764	-24358	-24339	-24281	-24357	-24265	-24258	-24269	-24263

Table 22: Outputs from choice model with K-means clustering.

Note: Standard errors in parentheses. Statistical significance levels: * * $p < 0.01$, * * * $p < 0.05$, * * * * $p < 0.1$.

E

Variable	Model 1	Model 2	Model 3	Model 5	Models 6-9
Price	-0.397*** (0.015)	-0.410*** (0.015)	-0.411*** (0.015)	-0.384*** (0.016)	-0.375*** (0.016)
Car ride duration	-0.120*** (0.005)	-0.125*** (0.005)	-0.126*** (0.005)	-0.125*** (0.005)	-0.119*** (0.006)
Public transport ride duration	-0.082*** (0.004)	-0.083*** (0.005)	-0.083*** (0.005)	-0.073*** (0.005)	-0.070*** (0.005)
Public transport wait time	-0.085*** (0.005)	-0.089*** (0.005)	-0.089*** (0.005)	-0.076*** (0.005)	-0.072*** (0.006)
1[Public transport is train]	0.307*** (0.074)	0.289*** (0.076)	0.288*** (0.076)	0.239*** (0.077)	0.240*** (0.077)
Attraction		-0.044 (0.158)			0.384** (0.176)
Similarity				-0.216*** (0.043)	-0.263*** (0.048)
Constant included	YES	YES	YES	YES	YES
Control variables included	NO	YES	YES	YES	YES
Number of observations	6180	6180	6180	6180	6180
Number of choices	1296	1236	1236	1236	1236
Number of subjects	108	103	103	103	103
Consistent Akaike information criterion	5201	5100	5108	5084	5087
Log likelihood	-2535	-2383	-2383	-2371	-2369

Table 23: Outputs from logistic regressions with experimental data.

Note: Standard errors in parentheses. Statistical significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Model	Model 1	Model 2	Model 3	Model 5	Models 6-9
Price	-0.222*** (0.008)	-0.229*** (0.008)	-0.229*** (0.008)	-0.215*** (0.009)	-0.209*** (0.009)
Car ride duration	-0.068*** (0.003)	-0.070*** (0.003)	-0.070*** (0.003)	-0.069*** (0.003)	-0.066*** (0.003)
Public transport ride duration	-0.046*** (0.002)	-0.047*** (0.003)	-0.047*** (0.003)	-0.042*** (0.003)	-0.040*** (0.003)
Public transport wait time	-0.048*** (0.003)	-0.050*** (0.003)	-0.050*** (0.003)	-0.043*** (0.003)	-0.041*** (0.003)
1[Public transport is train]	0.189*** (0.042)	0.179*** (0.043)	0.179*** (0.043)	0.146*** (0.044)	0.140*** (0.044)
Attraction		-0.008 (0.089)			0.223*** (0.098)
Similarity				-0.123*** (0.025)	-0.149*** (0.027)
Constant included	YES	YES	YES	YES	YES
Control variables included	NO	YES	YES	YES	YES
Number of observations	6180	6180	6180	6180	6180
Number of choices	1296	1236	1236	1236	1236
Number of subjects	108	103	103	103	103
Consistent Akaike information criterion	6074	5726	5734	5709	5712
Log likelihood	-2534	-2383	-2383	-2371	-2369

Table 24: Outputs from fixed effect Probit regressions with experimental data.

Note: Standard errors in parentheses. Statistical significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

F

Question	Variable	Construct	Answer scale	Adapted from
Using this recommendation system improves my performance during this task	Perceived usefulness	PU1	7 point Likert	Davis (1989), Venkatesh (2000)
Using this recommendation system increases my productivity in performing this task	Perceived usefulness	PU2	7 point Likert	Davis (1989), Venkatesh (2000)
Using this recommendation system enhances my effectiveness in performing this task	Perceived usefulness	PU3	7 point Likert	Davis (1989), Venkatesh (2000)
I find the recommendation system to be useful in performing this task	Perceived usefulness	PU4	7 point Likert	Venkatesh (2000)
My interaction with the recommendation system is clear and understandable	Perceived ease of use	PE1	7 point Likert	Venkatesh (2000)
Interaction with the recommendation system does not require a lot of mental effort	Perceived ease of use	PE2	7 point Likert	Venkatesh (2000)
I find the recommendation system easy to use	Perceived ease of use	PE3	7 point Likert	Venkatesh (2000)
I find the recommendation system easy to do what I wanted it to do	Perceived ease of use	PE4	7 point Likert	Venkatesh (2000)
I have control over how the recommendations are generated	User control	UC1	7 point Likert	Venkatesh (2000)
I have resources necessary to control how the recommendations are generated	User control	UC2	7 point Likert	Venkatesh (2000)
I have knowledge necessary to control how the recommendations are generated	User control	UC3	7 point Likert	Venkatesh (2000)
Given the resources opportunity and the knowledge it takes to control how the recommendations are generated it would be easy for me to do it	User control	UC4	7 point Likert	Venkatesh (2000)
Assuming I had access to this recommendation system I intend to use it for similar tasks	Behavioral intention	BI1	7 point Likert	Venkatesh (2000)
Given I had access to the recommendation system I predict that I would use it for similar tasks	Behavioral intention	BI2	7 point Likert	Venkatesh (2000)

Table 25: Constructs.