Fundusze Europejskie
Wiedza Edukacja Rozwój

Rzeczpospolita Polska

Unia Europejska
Europejski Fundusz Społeczny

POLISH-JAPANESE ACADEMY OF INFORMATION TECHNOLOGY

Computer and Information Sciences

Justyna Pawłowska-Bebel

Adapting Recommender Systems of E-Commerce Platforms to Deal with Cognitive Aging

Doctoral Thesis

Thesis supervisor:
prof. dr hab. Adam Wierzbicki
Auxiliary supervisor:
prof. dr hab. Grzegorz Sędek

2023

Acknowledgments

I would like to express my gratitude to Adam Wierzbicki, my research supervisor, for his unwavering support and patient guidance throughout my research journey. In particular, I am grateful for his guidance in exploring beyond the strictly technical aspects of research problems to include social, psychological, and even philosophical dimensions.

I would also like to thank Grzegorz Sędek, Klara Rydzewska, and Radosław Nielek for their continuous support during my PhD program. The majority of the research presented in this thesis was a result of the collaborative work of our SECON research group. Whenever the term "we" is used to present predictions or explain findings in this dissertation, it is to acknowledge the joint effort of the research group.

Lastly, I would like to acknowledge the National Science Centre, Poland, for supporting the research and preparation of this doctoral dissertation through the grant 2018/29/B/HS6/02604 awarded to Grzegorz Sędek.

Contents

1 Introduction **6**

1.1 Multi-disciplinary problem of self-induced bias 6

1.1.1 Motivation 6

1.1.2 Research Objectives 8

1.1.3 Approach 9

1.1.4 Research Contributions 9

1.2 Organization of the Thesis 10

2 Related work **13**

2.1 Recommender Systems 13

2.1.1 Content-Based Recommendation Algorithms 13

2.1.2 Collaborative Filtering Algorithms 14

2.1.3 Hybrid Recommendation Algorithms 15

2.1.4 Challenges in implementing recommender systems 16

2.1.5 Recommender systems for users with cognitive limitations . . . 17

2.2 Cognitive aging 19

2.2.1 Working memory limitations 20

2.2.2 Executive control 20

2.2.3 Intellectual helplessness 21

2.2.4 Decision-making 22

2.2.5 Adapting online services to deal with cognitive aging 23

2.3 Fairness and bias in algorithms 24

2.3.1 Fairness 25

2.3.2 Bias 27

3 Multi-Attribute choice on e-commerce platforms **30**

3.1 Product comparison on e-commerce platforms 30

3.1.1 Academic research on comparison shopping 31

3.1.2 Best practices in designing Comparison Tables - UX perspective 33

3.2 The problem of the multiattribute choice 35
3.2.1 Age-related differences in the multi-attribute tasks 36
3.2.2 Strategies in multi-objective decision making 37
3.3 Experiment for studying aging effects on multi-attribute choice 39
3.3.1 Experimental setting 39
3.3.2 Experiment 1 41
3.3.3 Experiment 2 46
3.3.4 Conclusions from the experiments' results 60

4 Aging Agent Model **62**
4.1 Model description 62
4.1.1 Model overview 62
4.1.2 Experimental determination of model parameters 64
4.1.3 Model design - agents 66
4.1.4 Individual preferences and preference function 67
4.1.5 Strategies in multi-attribute choice problems 67
4.1.6 Working memory 68
4.1.7 Affective Heuristics 68
4.1.8 Consideration sets 70
4.1.9 Enviromental parameters 70
4.2 Simulation description 71
4.2.1 Simulation Flow 71
4.2.2 Modeling decision support algorithms 73
4.2.3 Model validation 74
4.3 Results of the simulation - baseline 75
4.3.1 Working memory effect 77
4.3.2 Impact of affective heuristics 78

5 Effectiveness of different types of Recommender Systems **79**
5.1 Content-Based 79
5.1.1 Algorithm specification 79
5.2 Collaborative Filtering 80
5.2.1 Algorithm specification 80
5.3 Results 81
5.4 Quantifying Self-Induced Bias 82

5.4.1 Self-Induced Bias Calculation 82
5.4.2 Self-Induced Bias of CB and CF algorithms 84

6 Proposal of New Recommendation System 86
6.1 Preferences Twin 86
6.2 Decision competency-based 87
6.3 Collaborative Filtering With Working Memory Threshold 88
6.4 Results 89

7 Conclusions 93
7.1 Summary of Contributions 93
7.2 Limitations 94
7.2.1 Aging Agent Model Limitations 94
7.2.2 Recommender System Implementation Limitations 95
7.2.3 Barriers in implementing the proposed algorithms in commercial setting 95
7.2.4 Barriers in measuring self-induced bias in commercial setting . . 96
7.3 Future Work 96

Adapting Recommender Systems of E-Commerce Platforms to Deal with Cognitive Aging

Justyna Pawłowska-Bebel

March 1, 2023

Abstract

As the global population continues to age, with the number of people aged 60 and over projected to double by 2050, it is worth noting that older adults are becoming increasingly tech-savvy and are actively engaging with e-commerce platforms. However, with the multitude of options available on these platforms, it can be difficult for users, especially older adults, to make informed decisions about which products to select.

Recommender systems are tools designed to support users' decision-making by offering suggestions for items that may be of interest to them. Recommender systems, while widely used, suffer from a significant weakness in that they heavily rely on a user's past decisions, assuming that their purchases, interactions, or high ratings accurately reflect their preferences and needs. However, human decision-making is often influenced by heuristics and cognitive biases, leading to suboptimal choices. This is especially true for older users, who may be more susceptible to the effects of cognitive aging. This thesis presents the results of two experiments supporting the hypothesis that a customer's age can impact their ability to make optimal choices in e-commerce systems.

The thesis aims to analyse the negative impact of users' sub-optimal decisions on the quality of recommender systems' proposals, resulting in a phenomenon elucidated in the thesis as a "self-induced bias". This work explores the concept of self-induced bias in the context of existing research on recommendation algorithms' fairness and bias. The proposed framework offers a means of quantifying the magnitude of self-induced bias and is applied to popular recommendation algorithms to analyze their impact.

To measure the impact of cognitive limitations on decision-making quality, a simulator of the e-commerce purchasing process supported by a recommender system was developed with a focus on modeling clients with cognitive limitations. Several versions of classical recommender systems were implemented, and their effectiveness was measured. Using the simulated environment, the magnitude of self-induced bias among older users of traditional recommender systems was compared to customers who did not use any system. The results provide insights into how cognitive limitations can impact users' decision-making quality.

Lastly, three new algorithms with an aim of reducing the self-induced bias among older users were designed and tested in the simulated model based on the theoretical research on cognitive aging. They proved to be effective in providing better recommendations both for younger and older users. The new recommendation algorithms also have a lower self-induced bias among older users than recommendation algorithms that disregard users' cognitive limitations.

Streszczenie

Ponieważ globalna populacja starzeje się, a liczba osób w wieku 60 lat i starszych ma się podwoić do 2050 r., osoby starsze są coraz częstszymi użytkownikami Internetu i platform e-commerce. Platformy te oferują konsumentom szeroki wachlarz opcji i wielu użytkownikom trudno jest podjąć decyzję, które produkty wybrać.

Systemy rekomendacyjne to narzędzia mające na celu wspieranie użytkowników w podejmowaniu decyzji poprzez oferowanie sugestii produktów lub usług które mogą ich zainteresować. Znaczącą słabością systemów rekomendacyjnych jest to, że opierają się one na wcześniejszych decyzjach użytkownika, zakładając, że zakup, interakcja lub wysoka ocena przedmiotu przez użytkownika oznacza, że przedmiot jest pożądany i korzystny dla użytkownika. W rzeczywistości na ludzkie decyzje często wpływają heurystyki i błędy poznawcze, które mogą prowadzić do podejmowania przez nich nieoptymalnych decyzji. Jest to szczególnie istotne w przypadku starszych użytkowników, na których decyzje mogą negatywnie wpływać skutki procesu starzenia się kognitywnego. W niniejszej pracy zaprezentowano wyniki eksperymentów potwierdzających hipotezę, że wiek konsumenta może wpływać na jego zdolność do dokonywania optymalnych wyborów na platformach e-commerce.

Celem niniejszej pracy jest analiza negatywnego wpływu suboptymalnych decyzji użytkowników na jakość propozycji systemów rekomendacyjnych, skutkujących zjawiskiem zdefiniowanym w pracy jako „błąd wywołany decyzjami użytkownika”. Pojęcie to zostało opracowane w kontekście istniejących badań na temat błędów algorytmów rekomendacyjnych. Zaproponowano ramy kwantyfikowania błędu wywołanego decyzjami użytkownika, które wykorzystano potem do analizy tego efektu dla popularnych algorytmów rekomendacyjnych.

Do przeprowadzenia eksperymentów opracowano symulator procesu zakupowego na platformie e-commerce wspieranej przez system rekomendacyjny z uwzględnieniem modelowania ograniczeń poznawczych użytkowników. Następnie zaimplementowano kilka wersji klasycznych systemów rekomendacyjnych, aby zmierzyć, w jaki sposób ograniczenia poznawcze agentów wpływają na jakość ich decyzji. Na podstawie wyników uzyskanych na symulowanej platformie e-commerce zmierzono wielkość błędu wywołanego decyzjami użytkownika starszych użytkowników tradycyjnych systemów rekomendacyjnych w porównaniu z klientem niekorzystającym z żadnego systemu.

Na koniec, w oparciu o badania nad starzeniem się funkcji poznawczych, zaprojektowano i przetestowano w symulowanym środowisku trzy nowe algorytmy mające na celu zmniejszenie błędu wywołanego decyzjami użytkownika wśród starszych użytkowników. Wykazano ich skuteczność w dostarczaniu lepszych rekomendacji zarówno dla młodszych, jak i starszych użytkowników. Nowe algorytmy rekomendacji mają również mniejszy wskaźnik błędu wywołanego decyzjami użytkownika wśród starszych użytkowników niż algorytmy rekomendacji które nie biorą pod uwagę ograniczeń poznawczych użytkowników.

CHAPTER 1

Introduction

1.1 Multi-disciplinary problem of self-induced bias

1.1.1 Motivation

As the world's population continues to grow, so too does the number of older persons. According to the World Health Organization (WHO), by 2030, 1 in 6 people in the world will be aged 60 years or over. This demographic shift will lead to an increase in the proportion of the population aged 60 years and over from 1 billion in 2020 to 1.4 billion. By 2050, the number of people aged 60 years and older will double to 2.1 billion. The number of persons aged 80 years or older is also expected to triple between 2020 and 2050, reaching 426 million. This trend has significant implications for healthcare, social services, and other areas that cater to the needs of older individuals [1].

In developed countries, the share of older adults in the population rises and are also well connected to the Internet - in the US, 73% of citizens aged over 65 had an Internet connection in 2019 [2]. The older population has become increasingly prominent as users of e-commerce, particularly during the COVID-19 pandemic, which saw a surge in online shopping. Digital marketplaces and e-commerce platforms reported an increase in transactions ranging from 5% to over 100% in 2020 compared to the previous year [3].

E-commerce platforms offer an abundance of choices to users, with large web retailers having practically unlimited capacity to stock items. To assist users in making informed decisions, recommender systems have been developed. Before their development, individuals had limited options for obtaining advice on products or services, either relying on the experience of someone they knew or conducting research through guidebooks, specialist press, or subject matter experts. However, these methods were limited by either a lack of personalization or reliance on a single individual's experience. Recommender systems address both issues by combining the wisdom of crowds with the ability to tailor recommendations to an individual's preferences, providing a more personalized and effective decision support tool.

This solution, despite being efficient and flexible, suffers from one major weakness. Recommender systems rely on the user's past decisions, assuming that a purchase, interaction, or high rating of an item indicates that the item was desirable and beneficial for the user. However, this assumption may not always hold true. Human decision-making can be influenced by various factors, such as cognitive biases and heuristics, which can lead to suboptimal decision-making. As a result, the recommendations provided by these systems may not always align with the user's true preferences or needs.

Based on what we have learned about human decision making, human decisions are not always optimal. On the contrary, research indicates that an all-knowing, all-analyzing and utility-maximising homo oeconomicus is a creature of legends. Real humans make decisions based on complex sets of rules, often heuristics that lead to systematic errors and cognitive biases [4]. In addition, age-related changes in cognitive processes can negatively affect the decision-making abilities of older consumers, leading to suboptimal choices. As such, this thesis aims to investigate the hypothesis that age can have an impact on the ability of customers to make optimal decisions within e-commerce systems.

This thesis also proposes the concept of self-induced biases in recommendation algorithms. **Self-induced is connected to the recommendation system's users' limited or deteriorating ability to make optimal decisions**. The thesis takes into account the internal limitations that users may face in making decisions due to their limited cognitive abilities, which are extensively discussed in the literature review section. Specifically, the study analyzes how suboptimal decisions made by older adults are utilized as a training set for creating recommendations, potentially resulting in a higher likelihood of making further poor decisions. The goal of this study is to propose appropriately designed recommendation systems that can interrupt this cycle instead of perpetuating or exacerbating it.

The study of self-induced bias due to cognitive limitations requires an **interdisciplinary approach**. This thesis incorporates psychological research on the topic of cognitive aging, a simulation-based approach used in social informatics, and the engineering approach needed for understanding the state-of-art recommendation techniques in order to propound the method for identifying, measuring, and reducing self-induced bias.

1.1.2 Research Objectives

To summarize, this thesis has three main research objectives. The first objective is to evaluate the decision quality of older and younger consumers in e-commerce, testing the hypothesis that cognitive limitations of older customers lead to less optimal choices compared to younger customers.

The second objective is to investigate the impact of suboptimal consumer decisions on recommendation algorithms, with the hypothesis that training on suboptimal user choices can further worsen decision quality.

The third objective is to propose a framework to measure self-induced bias and place it within the context of existing research on the fairness and bias of recommendation algorithms. The hypothesis is that traditional recommender systems can magnify the self-induced bias of a system with older users.

The last research objective is to propose methods for combating self-induced bias in recommendation systems for e-commerce platforms. Hypothesis 4 is that recommendation algorithms taking into account the cognitive limitations of certain users can improve their users' decision quality and reduce self-induced bias.

1.1.3 Approach

To investigate the impact of cognitive limitations on decision-making in e-commerce, a psychological experiment was conducted with human participants. The experiment involved comparing the decision-making ability of older and younger users when selecting a washing machine model.

Using the results of the experiment, a cognitive model of an agent making purchasing decisions in an online shop was developed. Multiple simulations were then run to obtain training sets for recommender systems.

Various classical recommender systems were implemented to determine the extent to which cognitive limitations impact decision quality. The magnitude of self-induced bias of older users of traditional recommender systems was measured and compared to customers who did not use any system.

To reduce self-induced bias among older users, three new algorithms were developed based on the theoretical research on cognitive aging. The effectiveness of these algorithms was evaluated in a simulated e-commerce environment.

1.1.4 Research Contributions

Contribution to Informatics

This thesis makes the following contributions to the current state of the art:

1. Definition of new research problem concerning recommendation system trained on sub-optimal user choices: the problem of self-induced bias (a special case of data bias). Proposal of a method to quantify self-induced bias using simulation.
2. Simulator of e-commerce purchasing process supported by recommender system for clients that have cognitive limitations (code available).
3. Proposal of three recommendation algorithms that counteract self-induced bias and improve decisions of customers with cognitive limitations, outperforming optimally configured Collaborative Filtering and Content-based algorithms.

Contribution to Psychology

The key contribution of this research to the field of psychology is a simulator of consumer decisions taking into account cognitive limitations and heuristics. As shown in Section 2.1.5, although there are available simulators of recommender systems in e-commerce applications, the cognitive aspects of the simulated consumers are beyond their scope.

The thesis also makes a significant contribution by expanding the scope of research on e-commerce beyond the limited perspective of user interface design, which is presently the primary area of focus in this field, as described in the section 2.2.5.

Lastly, as demonstrated in sections 2.1.5, psychologically informed recommender systems have not yet addressed the variations in a cognitive capacity that exist among individuals. The purpose of the research outlined in this thesis is to motivate the creation of a connection between computer science and psychology and to pinpoint crucial areas of collaborative investigation in the realm of recommendation algorithms.

1.2 Organization of the Thesis

This thesis is organized as follows:

1. **Chapter 2** expounds on the theoretical background concerning different recommendation algorithms and user modeling techniques, followed by a brief summary of psychological knowledge on processes related to cognitive aging. The final section of the chapter provides a review of the existing literature on bias and fairness in algorithms.
2. **Chapter 3** discusses the issue of making choices with multiple attributes when purchasing items online. It provides an overview of the psychological and user experience (UX) theories underlying comparison shopping, followed by experimental findings examining how aging affects decision-making in such scenarios.
3. **Chapter 4** explicates the development of multi-agent simulation for studying the effects of cognitive limitations of agents in the e-commerce space and their interaction with recommender systems.

4. **Chapter 5** - I elucidate the results of using the classical recommendation algorithms in the simulation and their impact on the self-induced bias. The chapter also propounds a method of quantifying the self-induced bias.
5. **Chapter 6** presents three novel recommendation algorithms that consider the cognitive limitations of older users. The efficacy of these designs in decreasing self-induced bias is demonstrated through simulations.
6. **Chapter 7** summarises the findings as well as their limitations. Potential future work directions are proposed as well.

The material and results included in this thesis are partially based on the following publications:

1. Pawlowska, J., Rydzewska, K., Wierzbicki, A., (2023). Using Cognitive Models to Understand and Counteract the Effect of Self-Induced Bias on Recommendation Algorithms. In Journal of Artificial Intelligence and Soft Computing, 13(2). [5]
2. Rydzewska, K., Pawłowska, J., Nielek, R., Wierzbicki, A., Sedek, G. (2021). Cognitive Limitations of Older E-Commerce Customers in Product Comparison Tasks. In Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part III 18 (pp. 646-656). Springer International Publishing. [6]
3. Nielek, R., Pawlowska, J., Rydzewska, K., Wierzbicki, A. (2021). Adapting Algorithms on the Web to Deal With Cognitive Aging. Multiple Pathways of Cognitive Aging: Motivational and Contextual Influences, 368. [7]
4. Pawlowska, J., Nielek, R., Wierzbicki, A. (2019). Lost in online stores? Agent-based modeling of cognitive limitations of elderly online consumers. In Social, Cultural, and Behavioral Modeling: 12th International Conference, SBP-BRiMS 2019, Washington, DC, USA, July 9–12, 2019, Proceedings 12 (pp. 204-213). Springer International Publishing. [8]
5. Rydzewska, K., Pawłowska, J., Nielek, R., Wierzbicki, A., Sedek, G., Aging Consumer Performance in E-commerce Choice Task: The Role of Working Memory and Decision Strategies in Multi-Attribute Product Comparisons. [1]

[1] Article currently under preparation

Publication	Chap. 2	Chap. 3	Chap. 4	Chap. 5	Chap. 6	Chap. 7
1. (Pawlowska et al., 2023)	X	X	X	X	X	X
2. (Rydzewska et al., 2021)		X	X			X
3. (Nielek et al., 2021)	X	X				X
4. (Pawlowska et al., 2019)			X	X		X
5. (Rydzewska et al., under prep)	X	X	X			X

Table 1.1: The material included in this thesis is partially based on publications, as indicated for all chapters

CHAPTER 2

Related work

2.1 Recommender Systems

2.1.1 Content-Based Recommendation Algorithms

Content-Based (CB) Algorithms[1] are based on product features. Such algorithms compare possible options to those a user chose in the past and recommend most similar items [9]. Content-based algorithms rely solely on the preferences and behavior of the current user and do not rely on data from other users. This results in a higher level of transparency because the system can explain which specific features of the recommended item led it to be included in the list of recommendations. Additionally, content-based algorithms are less susceptible to the "cold start" problem that can occur when a new item is added. Unlike recommendation algorithms that rely on the ratings of multiple users, which may not have enough information to generate recommendations for unrated items, content-based algorithms can recommend items that have not yet been rated by any user.

The content-based approach to recommender systems has several drawbacks. The algorithm's focus on finding the most similar items hinders its ability to recommend anything new. Tailoring the recommended items too close to a user's profile confines them to a 'bubble' of the same choices. Additionally, this method has limited abilities in analysing content, as the algorithms may struggle to capture all significant aspects

[1]The python code for the Content-Based and Collaborative Filtering algorithms was based on the article https://www.kaggle.com/code/gspmoreira/recommender-systems-in-python-101/notebook

of the content, particularly those that are not easily defined, such as aesthetic qualities and domain ontologies. Finally, content-based systems are vulnerable to the new user problem. Without sufficient data on a user's history, they cannot offer reliable recommendations. Over the past two decades, researchers have developed various strategies to address these issues and provide more value for users [7].

2.1.2 Collaborative Filtering Algorithms

Collaborative Filtering (CF) techniques combine the active user's own experience and preferences with the experiences of other users. Among collaborative recommendation approaches, methods based on nearest neighbors are the most popular [10]. This approach relies on the assumption that if user u rated several items similarly to user v, then it's highly likely that user u will rate a new item in the same way as user v did. A collaborative filtering algorithm overcomes most of the issues associated with the Content-Based method. As it uses the experiences of multiple users, the recommendation system achieves high accuracy in recommending new items to the active user. It is easy to understand and implement and doesn't require costly training phases that need to be frequently carried out in large commercial applications. Adding new users doesn't require re-training of the system, and it can justify its recommendations by providing a list of close neighbors based on which the selection was made. Lastly, this approach is more versatile than the Content-Based approach as it doesn't need recommended objects to be labeled or have sufficient content. This approach is not impervious to some drawbacks. Firstly, it suffers from limited coverage because it only considers users who have rated common items and ignores those who have similar preferences but have not rated the same items. This makes it less effective in cases where there are few or no common ratings between users. Additionally, CF algorithms are sensitive to sparse data and may not be able to provide accurate recommendations when there are not enough ratings available for a significant number of items. Furthermore, CF algorithms are not useful for new users or items added to the system that have no ratings at all, which is commonly referred to as the cold-start problem. Many extensions of collaborative filtering algorithms have been proposed [11]. As a case in point, extending the collaborative filtering approach to take into consideration the contextual information, such as when, where and with whom a movie is seen, has been shown to outperform the pure collaborative filtering method in the case of movie recommendation [12].

2.1.3 Hybrid Recommendation Algorithms

Hybrid recommendation algorithms aim to enhance the performance of individual recommendation techniques by combining two or more of them. This approach is particularly useful in addressing the typical challenges associated with individual techniques and achieving better overall performance. Most commonly, CF is complemented by some other techniques in an attempt to avoid the cold-start problem.

Based on the algorithm used for combining different recommendation components, Hybrid Recommendation Systems can be classified into seven classes[13]:

1. **Weighted** - Scores (or votes) of several recommendation techniques are combined using weighted additive formula.
2. **Switching** - System switches between recommendation techniques depending on the current situation.
3. **Mixed** - different recommenders provide their recommendation presented at the same time. This class is based on merging and presentation of multiple rated list into single rated list
4. **Feature combination** - This algorithm treats collaborative information as additional feature data associated with each example and use CB techniques over the augmented data set.
5. **Cascade** - Cascade hybrid involves a staged process where every recommender assigns some priority, and assigns priority accordingly; lower priority recommenders play a tiebreaker role over higher priority.
6. **Feature augmentation** - One technique is employed to produce a rating or classification of an item; this information is then incorporated into the processing of the next recommendation technique.
7. **Meta-level** - The model learned by one recommender is used as input to another one

2.1.4 Challenges in implementing recommender systems

Various techniques used in a recommender system experience some of the hurdles that may be described [14]in terms of basic problems as:

1. **Sparsity Problem** While working on large item sets from e-commerce and shopping Web sites, the user–item metric unavoidably becomes sparse which results in a limitation to many recommender systems. The main reason behind data sparsity is that most users do not rate most of the items and the available ratings are usually sparse. Few values of ratings in user–item metrics lead to worsening accuracy of predictions. Collaborative filtering faces this problem because it depends on the rating matrix in most cases To deal with data sparsity problem, many techniques have proposed dimensionality reduction like singular value decomposition [15], probabilistic matrix factorization [16], [17], and hybrid techniques.

2. **Cold Start problem** The cold start problem arises when a new user or item is introduced into the system. In CF-type recommender systems, it takes a significant amount of time to gather enough data on user preferences and item ratings, which are crucial for generating accurate recommendations. However, content-based methods do not rely on previous rating information from other users and can provide recommendations for new items.

3. **Scalability** Scalability refers to a system's ability to efficiently handle a large volume of data. Some recommender system algorithms face challenges in handling exponentially growing computation demands, resulting in increased costs, system errors, and inaccurate recommendations. To address these challenges, approximation mechanisms have been proposed to speed up the recommendation process. However, while these methods can improve system performance, they often come at the cost of reduced accuracy.

4. **Over Specialization Problem** In some cases, users are limited to recommendations that closely match their known or defined preferences, which is known as the over-specialization problem. This can hinder users from discovering new items and other available options. To address this issue and improve the user's experience with the recommender system, diversification has become a major topic of research in the field of recommender systems. Multiple methods have been developed to tackle this problem, in addition to solving the over-fitting issue [18].

2.1.5 Recommender systems for users with cognitive limitations

A comprehensive review of psychologically informed recommender systems was carried out in [19]. Some portion of the recommender systems listed in the review can be classified as systems addressing the issue of cognitive limitations. The earliest adoptions of the concept include the user modelling via stereotypes [20], which although more complex is still being proposed as a solution for problems like cold-start scenarios in CF [21].

Cognitive models of memory and its limitation are being used in recommender systems in different ways. A recommender system where a cognitive model of human long-term memory is used to resemble the manner in which a human expert makes recommendations are proposed in [22]. This system can model various human memory effects such as the fan effect [2]. Bollen et al. [23] exploit positivity effects from human memory theory in order to investigate temporal dynamics of ratings across recommender systems. The models of losing memory over time - Ebbinghaus forgetting curve [24] is used to account for shifts in user interests in works like [25], [26], [27] and [28]. ACT-R is a notable cognitive architecture used in designing and testing recommender systems. The architecture is designed to reflect how different cognitive domains work together [3] [29]. ACT-R model was used in the context of recommender systems mainly to model the activation equation of human memory. In this model, the probability that a piece of information (i.e., a memory unit) will be activated to achieve a processing goal is predicated on its usefulness in the current context along with a human's prior exposure to this information. Such a model has been used in the study of recommender systems in works: [30], [31], [32], [33], [34] and [35].

However, the **differences in cognitive capacity among individuals are yet to be addressed even in psychologically informed recommender systems**. For example, Mozer and Lindsey [36], who follow a hybrid approach that integrates collaborative filtering and computational models of forgetting, such as a variant of the above-described ACT-R activation equation assume the same memory retention patterns for older and young users.

[2] recognition times for a concept increases as more information is available about the concept

[3] short for adaptive control of thought-rational

Simulations of Recommendation Systems

In order to evaluate and test new solutions for recommendation systems, running experiments in a live environment with actual human users can be expensive and risky, given that it may negatively impact the user experience. Additionally, not all researchers have access to real-world recommender systems. Simulations have been proposed as an alternative. This approach involves simulating user feedback or decisions based on historical user data that has been logged.

Most research on simulations of recommender systems is focused on predicting the relevance of items to individual users and based on this information, simulating the user interaction with the recommender system (e.g., a click, a rating, dwell time, an order, etc). Initially, simulators were designed for specific datasets and specific recommendation tasks, like news article recommendation [37] or real-life retail platforms and their customers [38], which makes them difficult to use for more general research. As a solution, several simulation platforms like RecoGym [39], RecSim [40], MARS-Gym [41] and PyRecGym [42] have been developed, allowing for a varying degree of flexibility in modelling products, users, system and the interaction. Bernardi and colleagues [43] summarised the specifications of the simulators from the industry perspective.

The existence of user biases causes a challenge for offline testing of recommender systems. This issue has been addressed in [44] with a method for debiasing the data; however, our contribution to the field of recommender system simulators is the incorporation of user bias and cognitive limitations into the user model. This is premised on state-of-the-art research in psychology and marketing science. By simulating these factors, we aim to create a more realistic and accurate simulation environment for testing new recommender system solutions.

Agent-based model of e-commerce customer using Recommendation System

Our first study of self-induced bias in recommender systems [8] was an attempt to model the behavior of e-commerce platform users (agents) with an emphasis on reflecting the decision-making characteristics of older adults. An agent-based model has been utilized to evaluate the impact of cognitive limitations of e-commerce customers on the efficiency of collaborative filtering and content-based recommender systems.

System feature	RecoGym [39]	RecSim [40]	MARS-Gym [41]	PyRecGym [42]
Generalized Recommendation Task	NO	YES	NO	YES
User Feedback Flexibility	NO	YES	NO	YES
Marketplace Simulation	NO	NO	NO	YES
Modular Environments Support	YES	YES	NO	NO
External real/benchmark dataset	NO	NO	YES	YES

Table 2.1: Comparison of Recommender Systems Simulation Platforms, after [43]

The findings of the study indicated that recommender systems, such as CF and the naive popularity voting model, that rely on a larger group of agents for recommendations are more effective in situations where a portion of the population makes suboptimal choices due to age-related cognitive limitations. However, the limitations of the simulation model used in the study were noted, specifically in regard to the simplified modeling of human decision-making and the use of limited product sets. In this thesis, an improved simulation model of consumer decision-making is introduced, which considers cognitive limitations and is based on empirical data on user preferences in online shopping. The new model incorporates more comprehensive and sophisticated consumer decision-making heuristics and cognitive characteristics.

2.2 Cognitive aging

Aging is associated with negative changes in basic fluid cognitive abilities. Three popular aging theories explaining age-related changes are the Inhibition, Resources, and Speed Theories [45]. The theory of inhibition suggests that older adults experience a decline in cognitive performance due to a reduced ability to suppress irrelevant information. Similarly, according to the resources theory, older adults have a limited amount of cognitive resources available to allocate to a given cognitive task, compared to younger adults [46]. Finally, speed theory indicates that aging is associated with a decline in the speed with which information processing can be performed [47].

2.2.1 Working memory limitations

Working memory is defined as the preservation of information while simultaneously processing the same or other information[48]. It is widely agreed upon that older adults experience impairments in their working memory, which is the ability to temporarily store and manipulate information. However, there is still debate about the specific mechanisms underlying these impairments, particularly in the context of complex everyday tasks such as decision-making, problem-solving, and planning. These tasks require the integration and reorganization of information from multiple sources, which may be more challenging for older adults due to working memory deficits [49]. Some results suggest that different areas are activated in young and old adults, particularly within the prefrontal cortex, thus indicating that younger and older adults are performing these tasks differently[50]. An understanding of age-related neurophysiological changes may help to account for these differences.

2.2.2 Executive control

Executive control is a multifaceted concept that encompasses various processes necessary for organizing, planning, coordinating, implementing, and evaluating our non-routine activities [49]. It is a crucial component of various cognitive processes, involving the management, organization, and assessment of non-routine activities. Its importance lies in directing attention, inhibiting irrelevant or distracting information in working memory, devising encoding and retrieval strategies, and facilitating problem-solving, decision-making, and other goal-oriented tasks. This control is particularly crucial when dealing with new tasks that do not have a set of established process

2.2.3 Intellectual helplessness

The issue of intellectual helplessness is one of many manifestations of learned helplessness. People who experience prolonged exposure to difficult, uncontrollable situations, which they can neither avoid nor alleviate, will 'learn' to accept the situation as a given [51]. This can result in learned helplessness, where the individual stops trying to correct or avoid negative situations because they believe their efforts won't make a difference. In situations where there is an overload of information, much of which is false or misleading, it becomes difficult and mentally taxing to determine what information is accurate or reliable. This can substantially increase cognitive costs. If the cognitive cost is high enough, it may lead to cognitive exhaustion, low motivation, and anxiety when audiences process a high volume of news and information, stymying their ability to evaluate the information's veracity. As a result, these factors can cause individuals to become less careful when assessing the truthfulness of messages and more prone to relying on peripheral cues rather than the actual arguments presented in the information they come across. This phenomenon was widely observed during the COVID-19 outbreak [52].

The concept of intellectual helplessness is theoretically derived from the informational model of learned helplessness. This informational model ([53], [54][55]) strongly differentiates the modes of cognitive activity of people in controllable and uncontrollable task situations.

When situations are controllable, such as in a school setting or with proper training, individuals can progressively use their cognitive abilities to solve increasingly challenging tasks. However, in uncontrollable situations, such as unsolvable problems, despite intensive mental effort, individuals may not make any progress in understanding the problem. Even if they formulate tentative predictions, they may be unable to distinguish between good and poor options. This informational model suggests that prolonged cognitive exertion without any progress results in a negative psychological state called "cognitive exhaustion." This state can seriously impede analytical thinking and is particularly disruptive in tasks that require complex information processing for problem-solving and decision-making ([53], [54][55]).

Results of the experimental study of Sedek et al. [54] revealed that intellectual helplessness could reduce the pre-decisional time of older adults. This is because they do not anticipate achieving high accuracy and therefore, are unwilling to devote more time to solving tasks. Furthermore, the study found that cognitive exhaustion caused participants to approach intellectually complex tasks with less analytical thinking. This led to the oversimplification of strategies in multi-attribute tasks, resulting in lower decision-making accuracy in the compensatory environment.

2.2.4 Decision-making

While the decline in fluid cognitive abilities impact negatively performance in some decision-making tasks, there are also tasks where older adults perform better than their younger counterparts. Bruine de Bruin et al. [56] analyse the adult age differences in decision-making competence and concluded that while there is a negative relationship between age and performance on tasks that were mediated by fluid cognitive ability (Resistance to Framing, Applying Decision Rules), there is no age-related relationship or a positive age-related relationship on some tasks. The authors proposed that in the second set of tasks (Consistency in Risk Perception, Recognizing Age-group Social Norms, Under/Overconfidence, and Resistance to Sunk Costs), older adults may compensate for declining fluid cognitive abilities with experience. They acknowledged that it is challenging to measure experience-related abilities, such as crystallized cognitive abilities (such as specialized knowledge and decision strategies) and emotion-related abilities (such as enhanced processing of affective information), that may improve with age.

The same research showed that although not all decision-making tasks showed age-related declines in performance, older adults perceived themselves as worse decision-makers, which in turn may lead to lower motivation. It is worth noting, that older adults typically assess the task demands as requiring more effort than younger adults [57]. The selective engagement theory proposes that older adults, due to their cognitive limitations, are more selective in the information they engage with during decision-making. As a result, they prioritize obtaining sufficient information over maximizing their decision outcome. In contrast, younger adults are more likely to engage in systematic and maximizing decision strategies and value informational gain more highly. Additionally, the mental costs associated with searching for information increase with worsening cognitive abilities, further reinforcing the tendency of older adults to engage in selective engagement [58] the magnitude of this effect can be less prominent in the real-life

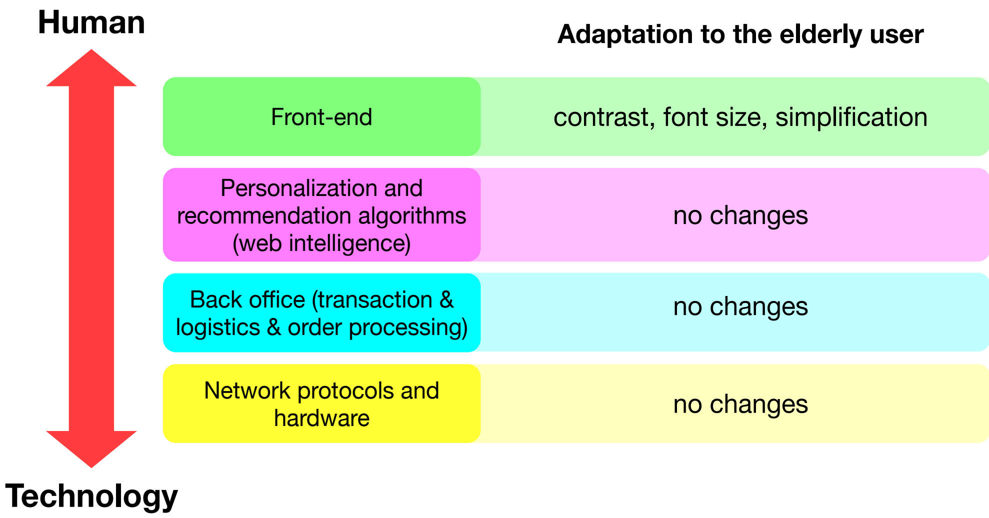

Figure 2.1: High-level overview of the components for the majority of online services (e.g. e-banking, online shopping websites etc.) Source: [7]

situations in comparison to the experimental setting as the older adults may be less willing to maximise effort in mentally taxing tasks. Both age-related declines in cognitive resources and motivational changes may interplay contributing to the reduction in information search among older adults.

2.2.5 Adapting online services to deal with cognitive aging

A brief look at Figure 2.1 reveals that only the very top layer of online services is designed, or at least can adapt to the requirements and expectations of elderly users. Researchers and practitioners involved in designing and studying user interfaces were among the first to recognize the need for special attention and dedicated solutions for older adults in the online environment. Morris [59] published one of the earliest works on this subject over 25 years ago and it has been followed by thousands of other studies, frameworks, and guidelines (see [60] [61], [62], [63]). A recent systematic literature review done by Nurgalieva et al. [64] shows that in many cases, these guidelines are contradictory. It is worth noticing attempts such as the CHI 2019 workshop "HCI and Aging: Beyond Accessibility" [65] to push interface design guidelines beyond font size or contrast, and to stress the need to avoid reinforcing negative stereotypes by design.

2.3 Fairness and bias in algorithms

The aim of this section is to place the self-induced bias in the context of existing work on algorithms' fairness and bias. While more details on the computation of the self-induced bias will be presented in Section 5.4, for comparison with other types of bias it is enough to state that self-induced bias occurs when users provide a recommendation algorithm with suboptimal decision data due to their limited decision-making abilities. To illustrate, John, an older consumer, wants to purchase a new fridge, but is unable to understand the technical details and does not want to spend much time considering available options. While he could easily find the same quality much cheaper, he eventually buys an expensive model of celebrity-endorsed brand X. The e-commerce site he used for the purchase remembers his choice, and the site recommends an expensive microwave from brand X when a month later John comes to replace a broken microwave.

The terms fairness, bias and discrimination can have different meanings in some of the cited works. I would like to start this section by clarifying the meaning of these terms in this dissertation.

Although fairness can be a multifaceted notion explained through various perspectives such as economic, philosophical, anthropological, and mathematical models, it can be generally defined as a system that does not discriminate against individuals or groups, treating them with equal respect and without any bias towards others.

The term bias can have two different interpretations. The first one, as defined by the Oxford Advanced Learner's Dictionary, refers to a non-objective or prejudiced attitude towards a particular group or perspective. The second meaning is more statistical, referring to the systematic deviation of estimators, models, measurements, or data from their intended ideal target. In this dissertation, the focus is on the latter definition of bias, particularly in relation to algorithms.

The fundamental contrast between the technical definition of "bias" and the notion of "fairness" is that bias is inherently non-normative, whereas fairness involves a moral evaluation. Although most fairness issues in the realm of algorithms stem from biases present in the system, its data, or its evaluation, it is helpful to differentiate between the technical reality and the moral, ethical, or legal implications.

2.3.1 Fairness

Social psychology identifies three kinds of fairness [66]: distributive fairness, procedural fairness, and retributive fairness. **Distributive fairness** is usually related to the problem of distribution of **goods, resources or costs**; its goal is to find an arrangement that is perceived as fair by the agents concerned. **Procedural fairness** focuses on the perceived fairness of **procedures** leading to outcomes, while **retributive fairness** is concerned with **rule violation** and the severity of **sanctions** for norm-breaking behavior. One way to conceptualize distributive fairness is as a particular form of procedural fairness. When a fair distribution of goods or costs is achievable, a fair procedure would mandate that all parties receive an equitable share. In situations where a fair solution cannot be reached in advance or agreed upon, procedural fairness is still relevant. While retributive fairness may be important for cases where artificial intelligence would potentially support the legal system, in the context of machine learning the main concerns are procedural and distributive fairness. Sometimes, these two types of fairness are also referred to as [67]: fairness by treatment, which happens when the protected attributes are not used explicitly by a predictor, and parity by impact in which the outcomes should be balanced across groups.

Fairness and discrimination in the context of protected attributes

Throughout this section, we will utilise the summary of fair algorithm concepts from Kusner et al [68]. In order to determine if a group is subject to discrimination, there must be a shared characteristic among its members, such as age, ethnicity, or gender. This shared characteristic is denoted as the set of **protected attributes**. The decision of whether an attribute is considered protected or not is a fundamental component of any fairness framework. Additionally, the conceptual framework requires that there are **observable attributes** of an individual as well as relevant **latent attributes** that are not directly observed. The outcome to be predicted based on the observable attributes and historical outcomes may contain historical biases.

To illustrate the different fairness concepts, considering a recommender system for choosing the best candidates for an analytical role: **protected attribute** can be the applicant's **gender**, **observable attributes** all attributes included in the candidate's **cv**,

latent attributes crucial for candidate's fit for the role are the candidate's **problem-solving skills**. The goal of the machine learning algorithm is to predict the candidate's performance on the job.

Different approaches are used to measure, whether the algorithm used to solve the problem was fair. These include fairness through unawareness [69], individual fairness [70], demographic parity/disparate impact [71], and equality of opportunity [72]. For the sake of simplicity, we often assume the protected attribute is encoded as a binary attribute, as in the recruitment example the candidate either belongs to the protected class or not, but this can be generalized.

Fairness Through Unawareness (FTU). An algorithm is fair so long as any protected attributes are **not explicitly used** in the decision-making process. Although the removal of the applicant's gender from the algorithm's input information seems like a simple solution, it has a major drawback. The information contained in the elements of observable attributes may include discriminatory information similar to the protected attribute, which may or may not be detectable by a human, but is captured by the algorithm. Therefore, expert knowledge is required to assess the relationship between the protected attribute and observable attributes, as demonstrated in the research on individual fairness:

Individual Fairness (IF). An algorithm is deemed fair if it gives similar forecasts to similar individuals. As elucidated in [70], the metric of similarity must be carefully chosen, requiring an understanding of the domain at hand beyond black-box statistical modeling.

Demographic Parity (DP). A predictor satisfies demographic parity if the probability of predicting a certain outcome for an individual belonging to the protected class is the same as for an individual not belonging to that class. In the given example, to achieve demographic parity, removing the sex of the applicant from the information used as input to the algorithm should result in an equal probability of selection for both men and women. This approach disregards the remaining attributes, contained in the observable set: if in a hypothetical pool of candidates, men have on average 5 years of experience less than women, the same probability of being accepted for the job is not desired in a well-calibrated algorithm.

Equality of Opportunity (EO). A predictor satisfies equality of opportunity if the probability of **correctly** predicting a certain outcome for an individual belonging to the protected class is the same as for an individual not belonging to that class. This eliminates the weakness of the fairness defined as DP: a man with a true potential to be a good employee has the same probability of being chosen by the recommender system as a woman who has the potential to be a good employee, however, the probability of male and female candidates accepted does not have to be equal.

Counterfactual Fairness In the extant literature, the description of counterfactual fairness is [68] for algorithm to be fair, the protected characteristic should not be a cause of prediction in any individual instance. In other words, changing the class while holding things that are not causally dependent on it constant will not change the distribution of prediction. The intuition underpinning the concept of counterfactual fairness is that a decision is fair towards an individual if it is the same in the existing circumstances as in a counterfactual world where the individual belonged to a different demographic group.

2.3.2 Bias

As noted in a comprehensive review of the literature on bias in recommended systems by Chen and colleagues [73], recent years have witnessed a surge of research endeavours on recommendation biases. However, they also observed that the term "bias" is used inconsistently across different papers, and it can refer to various phenomena. Researchers have identified different types of bias, including selection and conformity biases in **explicit feedback data**, exposure bias and position bias in **implicit feedback data**, inductive bias in models, and popularity bias and unfairness in **results**.

While the self-induced bias as defined in this thesis matches the criteria of a general data bias [74] [4], in the context of recommender systems biases it does not fit into any of the already defined categories:

Selection Bias happens as users are free to choose which items to rate so that the observed ratings are not a representative sample of all ratings. Put differently, the rating data is often missing not at random. This bias focuses on the issue that certain

[4] systematic distortion in the sampled data that compromises its representativeness

items are not chosen to be rated. In the majority of cases, only the particularly good or particularly bad items are reviewed. The self-induced bias places the emphasis on the choices actually made by the user.

Conformity Bias happens as users tend to rate similarly to the others in a group, even if doing so goes against their own judgment, making the rating values not always signify the user's actual preference. Self-induced bias is unrelated to any group or peer pressure.

Exposure Bias happens as users are only exposed to a part of specific items so that unobserved interactions do not always represent negative preference. Exposure bias occurs as users are only exposed to a part of items so unobserved interactive data does not always mean a negative signal. Exposure bias will mislead both the model training and evaluation. It is true that in the specific case of elderly users in e-commerce, certain items might be excluded from their scope of interest for reasons unrelated to their preferences, which fits into the category of exposure bias. However, the self-induced bias is a different problem that can manifest, apart from 1) lack of interaction with some items, also in 2) positive interactions with objectively non-desirable items, and 3) negative interactions with objectively desirable items (eg. in our model rejecting item based on non-favourite brand). Hence although in theory the methods proposed for case 1) should help to solve part of the problems connected with self-induced bias, they definitely cannot be treated as a solution to the general issue.

Position Bias happens as users tend to interact with items in a higher position of the recommendation list regardless of the items' actual relevance so the interacted items might not be highly relevant. **Inductive bias** denotes the assumptions made by the model to better learn the target function and to generalize beyond training data. **Popularity Bias** Popular items are recommended even more frequently than their popularity would warrant. Position, Inductive, and Popularity biases are not related to self-induced bias.

Unfairness happens when the system systematically and unfairly discriminates against certain individuals or groups of individuals in favor of others [51]. In the case of older users, we explicated in the thesis that self-induced bias indeed leads to recommended systems being effectively discriminatory against users with cognitive limitations in a sense, that it consistently proposes to them the less optimal items than the other groups. Notably, this phenomenon occurs although the classical recommender models we implemented do not consider the protected attribute (in this case - users' age) in the observable or latent data used for training the algorithm. Moreover, including the sensitive attribute in a way we propose in our new systems, leads to minimised unfairness. In addition, the fact that self-induced bias cannot be corrected by the methods proposed typically for reducing unfairness (like **rebalancing** [75] the data to add more observations for underrepresented groups, **regularisation** or **adversarial learning** in order to eliminate any direct or indirect information about sensitive attributes) suggest that this problem is different from what is usually described in the literature. Furthermore, the methods for assuring **counterfactual fairness** [68] are not applicable, given the suboptimal recommendations are attributed to causally dependent characteristics of older users, such as lower working memory or different decision-making strategy.

To summarize, the differences in its origins and nature together with the inability of applying solutions that are effective for other bias types make self-induced bias distinct research problems, although the self-induced bias is a type of data bias and holds some similarities to certain known subtypes of biases.

CHAPTER 3

Multi-Attribute choice on e-commerce platforms

3.1 Product comparison on e-commerce platforms

The developed markets are bounded to provide the customer with an abundant number of choices. This is especially true for online stores: while traditional retail stores can stock thousands of items, the capacity of large web retailers is practically unlimited. This abundance, however, can lead to choice overload: the feeling of being overwhelmed by an individual when he or she is confronted with excessive choices. Based on Herbert Simon's bounded rationality theory [76], we describe choice overload as a mental state in which the amount of choice information that needs to be processed exceeds the committed cognitive capacity of the decision-maker.

Choice overload is unpleasant for the customer and unwelcome for businesses. While having a personal choice, in general, has positive affective and motivational consequences for the individual, too many options decreases his or her satisfaction from the choice and motivation [77]. The consumer may defer the decision to buy, or resign from the purchase in general. Another option is overly relying on heuristics and biases while making the choice [78].

3.1.1 Academic research on comparison shopping

E-commerce platforms provide different tools to help consumers who are overwhelmed by too many options. In the literature, aids that operate within one e-commerce store suggest **what** to buy and **whom to buy from**. The former is referred to as **product brokering** and the latter - **merchant brokering** [79]. The product brokering aides can be further divided into two broader categories [80]: recommendation agents and the comparison matrix. The recommendation techniques that can provide the decision-maker with a pre-selected set of options are described in more detail in chapter 2.

Consumers can make in-depth comparisons among alternatives using the comparison matrix (CM). The CM allows consumers to organize attribute information about multiple products in an alternatives attributes matrix and to have alternatives sorted by any attribute. The matrix can be operating within a single particular e-commerce store (e.g. euro.com.pl in Poland) or across multiple online stores (e.g. ceneo.pl).

The research [81] on the manner in which the comparison matrix impacts the decision-making process of the customer demonstrated that the usage of the tool leads to:

1. an increase in the number of alternatives for which detailed product information is viewed,
2. a reduction in the number of alternatives considered seriously for purchase,
3. a larger share of nondominated alternatives in the set of alternatives considered seriously for purchase,
4. an increased probability of a nondominated alternative being selected for purchase,
5. a reduced probability of switching to another alternative (after making the initial purchase decision),
6. however the impact of the use of the comparison matrix on a degree of confidence in purchase decisions was not proven significant

While the mentioned study demonstrated the overall positive impact of the comparison matrix on the customer's decision quality, a further study [80] demonstrated that under certain conditions such aids can negatively affect decision-making abilities.

The study showed that decision quality decreased abruptly when the **the number of choices** increased from 12 to 24 or more. In addition, decision quality decreased significantly when the **the number of attributes** increased from five attributes to 10 in all contexts. Choice overload was comparatively more affected by increasing attributes than by increasing choices. Thus, the application designers should be aware of the decision overload effect and limit the options and attributes in comparison tools carefully.

Additionally, the study tested whether short-listing and sorting tools available on most e-commerce platforms can further affect decision quality. Short-listing tools allowed subjects to select their preferred products from the list by removing unwanted choices while sorting tools allowed them to sort and rank each attribute. The results suggest that the sorting tool materially improved the decision quality. Another finding was that short-listing counterintuitively increased the decision time. Moreover, the use of both tools leads to lower customer satisfaction from the choice.

Another study [82] analysed how the perceived usefulness of comparison shopping tools depends on consumers' comparison shopping proneness, which in turn is influenced by consumers' online decision-making styles. The authors use the concept of **comparison shopping**: "process whereby a consumer gathers as much information as possible about particular products and services for comparison before purchasing them". Due to the relative ease of collecting and processing information in online commerce compared to traditional retailers, customers in online stores are more likely to seek out alternatives across retail venues, product types, and brands based on their need for variety [83]. The authors described online decision-making style with characteristics such as Perfectionism, Brand Consciousness, Price-Value Consciousness, Recreational Shopping Consciousness, Impulsive/Careless Shopping, Confused by Over-Choice Shopping, Habitual, and Brand Loyal Shopping, Incentive/Bargain Consciousness, and Empowered Shopping.

The study highlighted a number of interesting findings. As expected, the customer with high Perfectionism traits tended to both, engage in comparison shopping activities and value the aid of online comparison shopping tools. Similarly, both Price-Value and Incentive conscious consumers are willing to engage in comparison shopping to satisfy their needs and appreciate the digital decision support systems. Not surprisingly, Comparison shopping proneness influences the perceived usefulness of comparison shopping tools, suggesting that consumers who are prone to engage in comparison shopping see these tools as supporting their shopping goals. Interestingly, even the consumers overwhelmed by an overabundance of choice and **not eager to engage in comparison shopping** as it would create even more information overload, **perceive comparison shopping tools as useful**. The same relationship holds for empowerment-looking customers: while not enthusiastic about comparison shopping in general, they have a positive stance toward comparison shopping tools. Another group not expressing interest in the concept of comparison shopping, but finding some comparison shopping tools useful are recreational shoppers. The negative stance toward both the concept of comparison shopping as well as the perceived utility of the decision support tools was expressed by people characterised by impulsive, brand-conscious, and habitual/brand-loyal shopping styles. Overall, as per the findings, comparison shopping tools are perceived as useful by consumers with a variety of online decision-making styles but there are still groups of customers who would not appreciate such aids.

3.1.2 Best practices in designing Comparison Tables - UX perspective

The best practices advised by UX practitioners are closely in line with the above-mentioned academic research. The Nielsen authors [84] recognise for example that too many alternatives lead to a cognitive overload and force customers to reduce the choice to a single attribute, hence engaging in **non-compensatory decision making**. They point out that a well-designed user interface allows the customer to select and compare multiple important features of a product, enabling them to apply **compensatory decision-making strategies**.

The author differentiates static and dynamic comparison tables. **Static** tables are used when the number of offerings is 5 or less and they are popular for comparing, for example, different subscription options, pricing packages, or different versions of a product from the same producer. **Dynamic** tables allow users to select which items they want to see in the comparison table.

The best practices for designing the comparison tables mentioned by the author are:

1. Use comparison tables for up to 5 items,
2. Be consistent in presenting the attributes' value,
3. Support scannability with a clear table format
4. Use sticky column headers
5. Chose meaningful attributes
6. Give users control
7. Simplify the comparison for mobile

The rules, particularly those concerned with the number of items, consistent presentation of attribute values, and focusing only on a selected set of meaningful attributes are very consistent with research such as [81] and [80]. The author stresses the role of comparison tables in reducing users' information overload. In the ending note, the author reminds the reader of the principle of putting the customer's goals first and aiding him in comparing and choosing the best option, instead of persuading or up-selling the most profitable products. This highlights one of the potential issues with using comparison tables: they can be constructed in a way that manipulates the user to choose a certain option over the other.

Another article [85] repeats the basic principles listed above and adds a lot of detailed technical design solutions, like improving the visibility and layout of some features, making the objects clickable so navigating the page and adding products to the shopping cart is easy, etc. Moreover, the advice to comparison tables designers is as follows:

1. Offer capability for users to mouse hover over technical phrasing/jargon to see a definition or description,
2. If there are many specs to compare, they should be organized, grouped, and easy to scan.
3. Prominently includes rating and number of ratings for each product on the comparison page.

The first two principles underscore the explanatory and educational functions the comparison tables can have given the right features are implemented. In the last example, the review serves as a social proof and a meta-attribute summarizing the quality of all the features. There is a potential drawback of such an approach - it creates the risk that the choice, reduced again to a single value, becomes a non-compensatory task.

A lack or missing feature should be clearly indicated in other articles, using simplified symbols rather than text if possible. [86], making the access and use of the comparison tables immediately obvious [87], including visual aids to highlight differences/similarities and using color-coding for better comparison [88]. Each publication focuses on reducing the cognitive load of the user and prioritizing user convenience, resulting in better service, higher perceived trustworthiness, and more purchases.

3.2 The problem of the multiattribute choice

Under multiple criteria decision-making, a person must evaluate multiple, sometimes conflicting criteria. Problems of this kind can be found in almost any field of human activity, but the growing scale and complexity of the e-commerce space create both a new demand for further research of this problem as well as the possibility to implement new online tools to assist people.

In the extant literature [89], the complexity of the multi-attribute problem is attributed to factors such as (factors most important in e-commerce in bold):

1. **complex and often incomplete, uncertain or conflicting knowledge of how to define and achieve the goals**
2. loosely defined alternatives,
3. **a large number of parameters that influence the decision**,
4. **a large number of alternatives**,
5. the presence of several decision-making groups with different objectives
6. time constraints imposed upon the process

Considering, that older users are becoming more and more active participants in the e-commerce business and the multi-attribute choice is a typical problem they face while shopping for more complex and costly goods, the age-related worsening in such tasks are particularly important to research and remediate.

3.2.1 Age-related differences in the multi-attribute tasks

According to many studies, older adults tend to perform worse than younger adults in cognitively demanding experimental decision-making tasks. A multi-attribute choice task, which requires a more complicated strategy and thorough information processing, exemplifies this phenomenon [90, 91].

It is difficult for older adults to apply decision rules for selecting an alternative from a list of options [92]. In multi-attribute choice tasks older adults have considerable problems with learning the value of cues [93] and options [94].

Motivation plays a salient role in the older adult's approach to decision-making. Bruine de Bruin et al. [56] analysing the adult age differences in decision-making competence showed that although not all decision-making tasks showed age-related declines in performance, older adults **perceived** themselves as **worse decision makers**, which in turn may lead to lower motivation. Combined with the fact that the mental costs of searching increase with aging, older adults place a greater emphasis on achieving a satisfactory outcome than on gaining information. Compared to younger adults, older adults are often quite likely to use maximizing decision strategies and search for information only for as long as it is necessary. According to the selective engagement theory [58] the magnitude of this effect can be less prominent in the real-life situation than in an experimental setting as the older adults may be less willing to maximise effort in mentally taxing tasks. Among older adults, both cognitive declines and motivational changes may contribute to a reduction in information search.

3.2.2 Strategies in multi-objective decision making

There is a number of well-researched strategies that can be applied to solve a multi-attribute choice problem.

Take The Best (TTB) is one of the simplest of the strategies. It is based on one, most important information. That means making a decision by looking at the cue with the highest validity and its respective values. If the most predictive cue (one with the highest validity) is not discriminative, then TTB involves looking at the cue with the second-highest validity and so on.

TTB is a non-compensatory strategy. **Compensatory strategies**, as the name suggests, allow for compensation, where two cues favoring one alternative can counterbalance another cue favoring another alternative. The concept of designing principles, based on which the person or an institution should make a choice in order to satisfy more than a single objective has been researched by economists since Pareto's [95] elucidated his archived evolution strategy (PAES) for finding optimal solution values.

One of the less complex compensatory strategies is **TALLY** [96], where cue weights are ignored and cue values are added for each alternative. The alternative with the largest sum is the one that is being selected.

Weighted Additive (WADD) is a more information-intensive strategy. It entails the multiplication of cue values by cue validities and the addition of them for each alternative. The alternative with the largest sum is selected.

A review of more complex strategies leveraged in the multi-attribute choice was presented in [97]. **The Global Criterion Method** transforms plural problem optimization into a single problem optimization by minimising the distance between multiple reference points and viable destination areas. Global Criterion Method does not need any preference information from the Decision Maker, i.e. weights and priority ranking are not used for goal functions. Instead, the endeavour is to minimize the distance between chosen solution and the ideal solution.

When cardinal information about the preferences of the decision maker is available, the **Utility Function** can be constructed and solved using - as a case in point - one of the Methods for Cardinal Preference of Attribute over Linear Assignment [98]. Another illustration of a multi-objective decision-making problem with cardinal preferences is the **Bounded Objective Function Method** or **ε- constraint Method** [99] which minimizes the single most important objective function. In contrast, all other objective functions are used to form additional constraints.

With the **lexicographic method**, preferences are imposed by ordering the objective functions based on their importance or significance, as opposed to by assigning weights [100]. Following the order of importance, the optimization process is done individually on each objective. After optimizing the most important objective (the first objective), if only one solution is returned then the solution is the optimal solution. In contrast, optimization will continue on the second objective and with new constraints on the solution derived from the first objective. This cycle continues till the last objective is achieved.

In the **Goal Programming**, the decision-maker determines the aspiration level of the objective function. Optimizing objective function with aspiration level is seen as the goal to be achieved [101].

An example of this method is the **Reference Point Approach** mentioned in [102]. For each feature i of the product, every agent has a desired aspiration level qa_i (which would be good to achieve) and reservation level qr_i (which if not achieved, causes a sharp drop of utility). Then for each product an order-consistent utility function is aggregated from partial utility functions calculated for each feature. This solution has a number of features desired in a model supporting human decision-making, like capturing nonlinearity of preferences including a preference for balanced solutions characterizing most human real-life problems, holistic perception of criteria and intuitiveness of setting the desired and unaccepted reference points instead of somewhat abstract weights.

Multi-objective evolutionary algorithm (MOEA) [103] is a stochastic optimization technique. As is the case with other optimization algorithms, MOEAs help find optimal Pareto solutions for specific problems. However, they differ from population-based approaches. Most existing MOEAs use the concept of domination in their actions, but some do not. The optimization mechanism of the MOEA is very similar to evolutionary algorithms, with the exception of the dominance relationship. At each iteration, the objective value is calculated for each individual before being used to determine the relationship of dominance in the population to select a potentially better solution to produce the hereditary population. This population may be used in conjunction with parent populations to produce populations for the next generation.

Furthermore, the existence of objective space may give MOEA the flexibility to apply certain conventional support techniques including niching.

The presented here list of strategies to solve multi-objective decision-making problems is not exhaustive due to the breadth of the topic. Moreover, not all the strategies can be used by humans, some approaches due to computational complexity are designed for implementation as computer algorithms. For the experiments that this chapter elaborates on, only three approaches were taken into consideration: TTB, TALLY, and WADD.

3.3 Experiment for studying aging effects on multi-attribute choice

3.3.1 Experimental setting

The design of our two experiments aimed to replicate the product comparison process typical in an e-commerce environment. We needed an objective ground truth - a choice in each product comparison we use as a correct one in order to investigate our research question. This was achieved by instructing experiment participants to opt for the best washing machine out of three available products taking into account **multiple features of varying importance**. Each feature's importance was specified in the experiment using a simple metaphor - a number of plus signs associated with each feature, ranging from 1 (least important) to 6 (most important).

Feature	Importance	PRODUCT 1	PRODUCT 2	PRODUCT 3
energy label	++++++	A	A	**A+**
water usage	+++++	65	65	**55**
noise level	++++	**65**	70	70
quick wash	+++	**yes**	**yes**	no
capacity	++	4	**6**	4
max spin speed	+	1000	1000	**1200**

Figure 3.1: Experiment screen showing the single multi-attribute choice task. On the left are the washing machine features and their importance expressed as a number of plus signs. On the right are the three products and their features.

Figure 4.5 displays the main screen of a single experiment task. The experiment interface was designed to resemble a feature comparison table available on many e-commerce platforms where users can select a number of products for comparison, which are subsequently shown in a tabular form.

On the screens preceding the task, participants were provided with a succinct description of the meaning of each feature along with its relative significance for an average consumer expressed as a number of plus signs, ranging from 1 plus sign for the least important feature (max spin speed) to 6 for the most important one (energy label). The participants were explicitly instructed to follow the preferences of an average consumer. The list of features is presented in descending order of importance on the left side of the screen, as depicted in Figure 4.5. On the right side, there are three washing machines and their feature values. When a feature value was better for one or two washing machines, such as a lower noise level preferred by most customers, it was printed in bold.

We designed multiple sets of tasks, three products each. In contrast to the multi-attribute choice task experiment explained in the literature, which presented attribute values as a dichotomous yes/no, our experiments calibrated feature values to closely match those of washing machines currently available on e-commerce platforms. There were three difficulty levels in the tasks. Initially, 'simple', washing machine feature values were selected to model a non-compensatory choice: all 3 strategies, TTB, TALLY, and WADD strategies would yield the same correct answer so even a user fluent in using only the simplest of the strategies would get all the answers right. The remaining 15 tasks were constructed to model a compensatory choice, so only using a compensatory strategy would make it possible to answer correctly. In the case of compensatory choice tasks that were marked as ’moderately difficult’ both TALLY and WADD strategies yielded the same correct answer. In the tasks marked as ’difficult’, only the WADD strategy allowed to choose the correct answer.

3.3.2 Experiment 1

Participants

One hundred and thirty-five people participated in the study, including 49 younger adults (26 women; age range 20-30; mean age = 25.65, SD = 2.96; mean years of education = 14.76, SD = 2.45), 45 middle-aged adults (24 women; age range 42-52; mean age = 46.38, SD = 3.02; mean years of education = 14.98, SD = 2.55) and 41 older adults (23 women; age range 65-76; mean age = 69.27, SD = 3.00; mean years of education = 15.41, SD = 2.94). Participation in the study took between 1 and 2 hours, and participants were remunerated for their participation. The study was approved by the Ethics Committee of the SWPS University of Social Sciences and Humanities in Warsaw.

Questionnaires

Participants completed the Scales of Helplessness and Anxiety about Contracting an Infectious Disease (Rydzewska Sedek, in preparation). Each of these scales was composed of ten items with answers on a 5-point Likert Scale ranging from 1 = “never” to 5 = “always”. An exemplary item from the Scale of Helplessness was "I feel helpless in the face of the possibility of contracting an infectious disease" and a sample from the Scale of Anxiety was "I am concerned when someone around me sneezes without covering their mouth". These scales’ reliabilities were impressively high: Cronbach’s alpha = .92 and Cronbach’s alpha = .93, respectively for the Scale of Helplessness and the Scale of Anxiety of Contracting an Infectious Disease.

Finally, participants also concluded Scales of Subjective Numeracy [104] and Objective Numeracy [105]. The scale of Subjective Numeracy comprised seven questions measuring subjective views on cognitive abilities concerning fractions and percentages, and preferences for displaying numeric information.

The Objective Numeracy Scale consisted of three open-ended questions designed to measure numerical skills, along with a warm-up question. The Subjective Numeracy Scale showed satisfactory reliability (Cronbach's alpha = .83), while the Objective Numeracy Scale, despite consisting of only three items, also showed satisfactory reliability (Cronbach's alpha = .65).

Tasks sets

We designed 32 task sets of three products each. The values of these features were chosen to realistically reflect washing machines currently available on e-commerce platforms, which is new for a multi-attribute choice task (as the values are usually presented as dichotomous yes/no). For 17 tasks marked as 'simple', washing machine feature values were selected to model a non-compensatory choice. In these tasks, the use of either a simpler TTB or TALLY strategy or the WADD strategy allowed the user to select the same correct answer. The rest of the 15 tasks were constructed to model a compensatory choice. Eight compensatory choice tasks were marked as 'moderately difficult'. A user could determine the correct answer using both the TALLY and WADD strategy. In the remaining 7 tasks, marked as 'difficult', only the WADD strategy made it possible to choose the correct answer.

Results and Discussion

Age-related differences in accuracy of multi-attribute choice tasks

In the first part, we will describe the main results of the impacts of age and parameters of tasks on the accuracy of performance. Subsequently, we will overview the correlation matrix of the main variables. Finally, we will demonstrate the role of cognitive and emotional variables as potential mediators and covariates of the relationship between age and multi-attribute decision-making task performance.

Proportion of correct decisions. The proportion of correct decisions is a key measure of performance in the multi-attribute product choice task. For the record, the participants did not receive feedback concerning the accuracy of their decision-making,

Variable	1	2	3	4	5	6	7	8
1. Age	-							
2. Simple	$-.24^{**}$	-						
3. Moderate	-.11	.01	-					
4. Difficult	.04	.12	$.38^{**}$	-				
5. S_Numeracy	-.10	$.25^{**}$	.13	.06	-			
6. O_Numeracy	-.05	$.39^{**}$	$.19^{*}$	.16	$.53^{**}$	-		
7. Helplessness	$.23^{**}$	$-.23^{**}$	$-.17^{*}$	-.13	.01	$-.19^{**}$	-	
8. Anxiety	-.08	-.12	$-.22^{**}$	-.09	-.01	$-.20^{**}$	$.80^{**}$	-

Table 3.1: Correlation table of dependent and independent variables. Simple = simple decision tasks; Moderate = moderately difficult decision tasks; Difficult = difficult decision tasks; S_Numeracy = Subjective Numeracy Scale; O_Numeracy = Objective Numeracy Scale; Helplessness = Scale of Helplessness of Contracting an Infectious Disease; Anxiety = Scale of Anxiety of Contracting an Infectious Disease. Note: *p < .05, **p < .01.

barring the training tasks. A 3 x 3 (Age [younger adults, middle-aged adults, older adults] x Decision Difficulty [simple, moderately difficult, difficult; within-subject variable]) mixed ANOVA on the proportion of correct decisions yielded one main effect along with an interaction effect. It is important to note that the assumption of equal error variances was not violated, as Levene's test did not show significance across age groups. A main and strong effect was observed about the decision difficulty, $F\ (2, 264) = 92.22$, $MSE = .055$, $p < .001$, $\eta_p^2 = .411$, with the significant differences ($p < .001$, Sidak post-hoc tests) between all decision tasks – the more difficult was the task, the worse the performance. This main effect was qualified by the significant Age x Decision Difficulty interaction, $F\ (2, 264) = 3.47$, $MSE = .055$, $p = .01$, $\eta_p^2 = .050$. As displayed in Figure 3.2, for the simple decision task the performance of younger adults was significantly higher than that of both the middle-aged group ($p = .005$) and the older adults group ($p = .006$). For the moderately difficult decision task, the performance was better in the middle-aged group

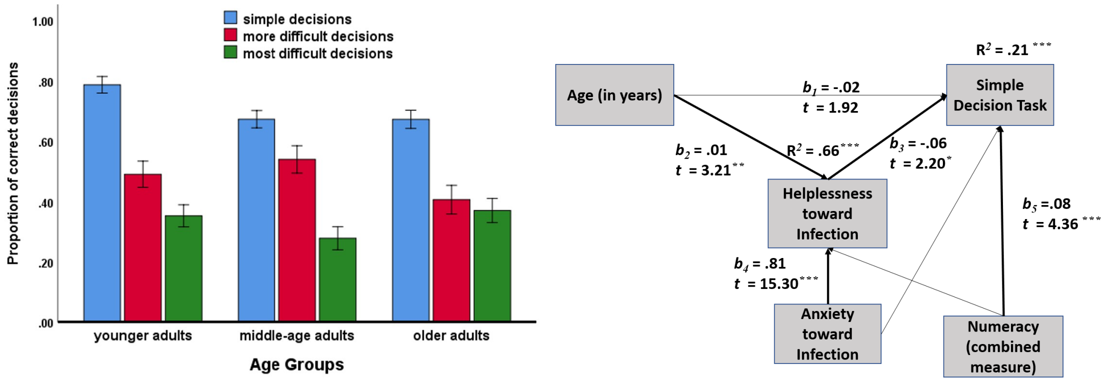

Figure 3.2: Performance of decision task as the function of age and decision difficulty and model of the mediational effect of age via the Scale of Contracting an Infectious Disease on simple decision task. Error bars denote the standard errors of the mean. Entries are unstandardized regression coefficients with t-test values. *$p < .05$, **$p < .01$, ***$p < .001$

in comparison to the older adult group (p = .04). No significant age differences were seen for the most difficult decision tasks. For decision difficulty comparisons within the age groups, significant differences were seen between three levels of difficulty for younger adults ($p < .01$) and middle age adults groups ($p < .05$). However, there was no significant difference in performance between the moderately difficult and the difficult decisions for the older adults group, but significantly better performance ($p < .001$) for the simple decision tasks in comparison to the moderately difficult decisions (see Figure 3.2). Prior findings were confirmed by the discovered age-related limitations in solving simple and moderately difficult multi-attribute product choice tasks, especially in the older adults group [90, 91]. Interestingly, age differences are observed even between younger and middle-aged groups in relatively simple tasks. This indicates that challenges with learning values of cues and options that were previously demonstrated in older adults [93, 94] could be also identified among the middle-aged group.

Mediation analysis for the simple decision task.

The correlation table between the accuracy of the decisions, age, and questionnaire measures (see Table 3.1) indicates that age (in years) significantly and negatively correlates with the performance of simple decisions and age positively correlates with the Helplessness Scale. Furthermore, there were significant (although not systematic) correlations between numeracy scales, Scales of Helplessness, and Anxiety and accuracy in simple and moderately difficult decision tasks. These results align with previous findings that numeracy is a crucial predictor of performance across various decision-making tasks. The observed significant correlations between objective numeracy and both anxiety and helplessness support the anticipated association between low numeracy and emotions that hinder effective decision-making [106]. To the best of our knowledge, the results demonstrating that feelings of helplessness and anxiety about contracting an infectious disease are significant predictors of the multi-attribute choice task performance's accuracy are a novel demonstration (specific to the COVID-19 pandemic) of the dysfunctional role of such negative emotions when making a rational multi-attribute choice in the domain of e-commerce.

In spite of the fact that the correlational analyses above indicated (in an indirect manner) that numeracy scales and the Scale of Helplessness and Anxiety of Contracting an Infectious Disease played an important role in explaining the age-related limitations on accurate decision-making, the Scale of Helplessness of Contracting an Infectious Disease appeared to be the classical mediation variable. Namely, the mediation analysis (applying the Process software) [107] examining the role of Helplessness in Contracting an Infectious Disease (see Figure 3.2), indicates that aging has a significant impact on Helplessness, and Helplessness is a reliable mediator of the relationship between aging and proportion of correct answers in simple decision tasks. A total of 10000 bootstrap samples were used to estimate percentile bootstrap confidence intervals. To integrate these findings, this model is supplemented by two covariates. We created a combined measure of numeracy (averaging the standardized scores for both subjective and objective numeracy scales) due to the highly correlated responses of the subjective and objective numeracy scales (see Table 1). The second covariate was the Scale of Anxiety about Contracting an Infectious Disease. This combined Numeracy scale was a strong, additional predictor of simple decision accuracy. Even though the Scale of Anxiety was strongly related to the Scale of Helplessness, it did not significantly predict the accuracy of simple decision tasks.

In sum, the mediational model with covariates nicely integrates the correlational analysis and demonstrates the interaction between age-related limitations and the role of additional variables (particularly helplessness and numeracy) in explaining the relationship between aging and performance in simple multi-attribute decision-making tasks. However, we are cognisant of the fact that this mediational model, although original, is still preliminary, and elucidating psychological mechanisms underpinning the role of intellectual helplessness as a mediator demands further experimental work (e.g., experimental manipulation of the intensity of a mediator, see [108]).

Conclusions In this experiment, we have considered the question of how the cognitive limitations of older consumers affect their choices when using a popular user interface function in e-commerce systems: product comparison.

We have found significant aging effects, especially for moderately difficult decisions, when the proportion of correct decisions of older adults drops as low as 40%. The pandemic exacerbates feelings of helplessness. This effect can have a severe adverse impact on the training and functioning of recommendation systems. In order to mitigate this effect, we recommend that content-based recommendation systems should not be used for older adults since these are particularly sensitive to individual choices that may be - as shown in this research - 60% incorrect. Instead, it is better to use collaborative or hybrid recommendation systems, in particular specially designed algorithms that would take into account age diversity.

3.3.3 Experiment 2

Participants

In the online study, the preliminary sample of participants recruited consisted of 170 participants. Twenty-one participants were excluded from analyses (10 younger adults, 12 middle-aged adults, and 9 older adults) because they had both very low levels of performance in the simplest decision task (accuracy below 70%) and their decision times were below the general mean. The final sample comprised one hundred forty-nine participants, including 50 younger adults (25 women; age range 19–30; mean age = 25.72, SD = 2.86; mean years of education = 15.16, SD = 2.60), 45 middle-aged adults

(22 women; age range 42–53; mean age = 47.82, SD = 3.37; mean years of education = 16.18, SD = 3.31) and 54 older adults (28 women; age range 65–76; mean age = 69.37, SD = 3.03; mean years of education = 15.43, SD = 3.04). Participation took between 1 and 2 hours, and participants were remunerated for their participation. The study was approved by the Ethics Committee of the SWPS University of Social Sciences and Humanities in Warsaw.

Questionnaires

Measure of Visual Working Memory (VWM) As the measure of VWM, we applied the visual pattern span task successfully applied in the large internet study with participants across adult life-span [109]. A rectangular matrix pattern with white and blue squares was presented for two seconds before being replaced with a blank matrix; the task required us to correctly select the squares that were previously blue. The patterns commenced with a matrix of 3 squares x 2 squares (4 blue squares) and the matrices increased up to a maximum of 5 squares x 5 squares (12 blue squares), with two patterns shown at each level (matrix size). The test concluded when participants were unable to accurately recall blue squares on two consecutive trials of a particular matrix size. Performance was evaluated based on the number of correctly recalled patterns.

Measures of Subjective Numeracy, Need for Cognitive Closure, and The helplessness of Contracting an Infectious Disease Participants completed the Scale of Subjective Numeracy [104]. This scale consisted of seven questions measuring subjective views on cognitive abilities in relation to fractions and percentages, and preferences for displaying numeric information. The reliability of the Subjective Numeracy Scale was high: Cronbach's alpha = .87.

Next, participants completed the Need for Cognitive Closure Scale [110], which included 12 statements that measure epistemic motivation on the inter-individual level. More specifically, the Scale measures a desire for a definite answer to a question, rather than ambiguity, uncertainty, or confusion, regardless of the quality of the answer. The reliability of this scale was the following: Cronbach's alpha = .74.

Finally, participants concluded the Scale of Helplessness of Contracting an Infectious Disease (Rydzewska & Sedek, in preparation). This scale was composed of ten items with answers on a 5-point Likert Scale ranging from 1 = “never” to 5 = “always”. An exemplary item was “I feel helpless in the face of the possibility of contracting an infectious disease”. The reliability of this scale was impressively high: Cronbach’s alpha = .93.

Correlation Measure of Features Similarity Spearman’s rank correlation was calculated between the presented and subjective importance of washing machine features. (more details needed)

Task sets

The 64 task sets, each consisting of three products compared by the participants in the Decision Task were prepared to reflect real-life choices. To assure the washing machines’ parameters are close to the ones observed on the market, we web-scrapped the assortment of one of the popular household appliances e-commerce stores. For 6 key characteristics of the washing machines, we collected data on the statistical distribution of values in the population to determine what can be considered a norm in high-end and low-end products and what is the typical variation of the values. Based on the data, we determined the parameters Quality(Q) and Delta (D). All products in tasks where Q was high had parameter values characteristic for high-end products (e.g. noise level in the range 50-40 db) and products with low Q analogically worse values (e.g. noise level in the range 70-60 db). The D parameters specified the extent to which the difference would be between better and worse parameter values within the same task e.g. in large D, high Q the participant would compare a washing machine generating 50 db with one generating 40 db, while in small D, high Q the parameters would be analogically 50 db and 45 db. The initial 32 task sets were displayed with feedback: after concluding each task the participant was informed about the accuracy of the answers. The second half of the task was not followed with feedback.

Results and Discussion

Proportion of Correct Decisions Part I of the Decision Task focused on the role of feedback. The primary performance measure for the multi-attribute product choice task was the proportion of correct decisions. During this part, participants received feedback on the accuracy of their performance in the choice task, tailored to the strategy required by the given decision task (WADD, TALLY, or TTB). The 32 trials were divided into four phases to assess the impact of continuous feedback on the improvement of performance within this initial part of the decision task.

A 3 x 3 x 4 (Age [younger adults, middle-aged adults, older adults] x Decision Difficulty [simple, moderately difficult, difficult; within-subject variable] x Phase of the task [first, second, third, fourth; within-subject variable]) mixed ANOVA on the proportion of correct decisions yielded three main effects and an interaction effect. There was main effect of age, $F(1,146) = 8.82, MSE = .28, p < .001, \eta_p^2 = .108$. The performance of younger adults (M = .71) was significantly higher than that of both the middle-aged group (M = .62, p = .03, for this and the next comparisons, Sidak post-hoc tests were applied) and older adults group (M = .58, p < .001), no significant differences were observed between middle-age and older adults groups. There was also a main and strong effect of decision difficulty, $F(2,292) = 138.00, MSE = .18, p < .001, \eta_p^2 = .486$, with the significant differences (p < .001,) between all decision tasks – the more difficult was the task, the worse the performance. Finally, the main effect of Phase, $F(3,438) = 47.02, MSE = .07, p < .001, \eta_p^2 = .244$ was also seen. A significant increase was observed in performance in phase 2 in comparison to phase 1 (p < .001), with a lack of further significant progress in phase 3 and phase 4. The last two main effects were qualified by the significant Decision Difficulty x Phase interaction, $F(6,876) = 9.30, MSE = .07, p < .001, \eta_p^2 = .060$ (see Figure 1).

The increase of performance in phase 2 when compared with phase 1 and further lack of improvement was also characteristic for moderately and most difficult tasks. However (see Figure 1), for the simplest decision task there was a linear increase trend across all four phases (p < .001).

In summation, there were reliable differences between age groups (best performance of younger adults) and also between the difficulty of decision tasks (the more difficult task the worse performance) with a lack of interaction between age and difficulty of the tasks. There was a salient improvement in the performance of simplest tasks across all phases of Part I of this study (with constant feedback), however, for more difficult tasks there was only improvement in phase 2 in comparison to phase 1 with a

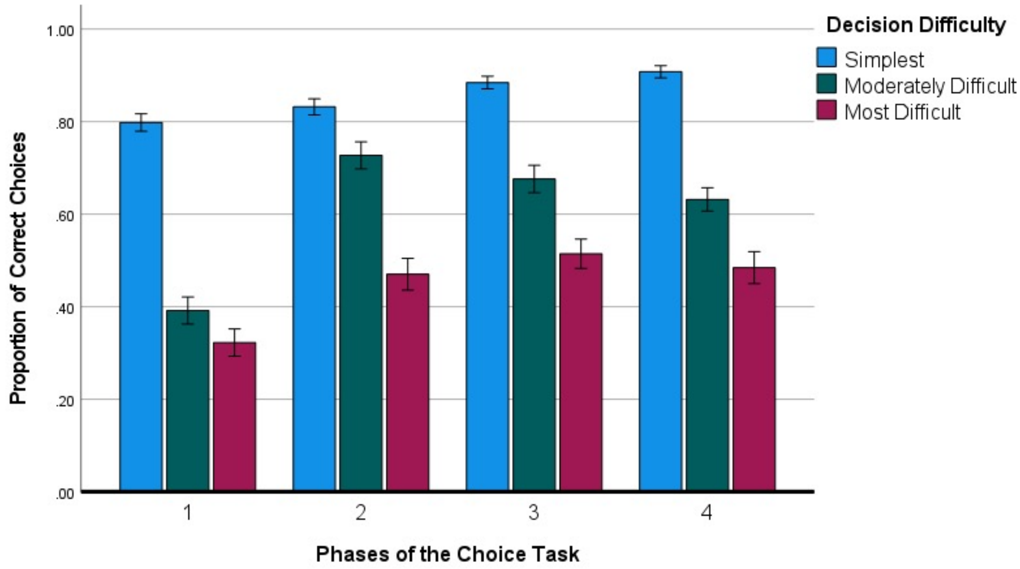

Figure 3.3: Proportion of correct choices in multi-attribute decision task (Part I with feedback) as a function of phases of the task and decision difficulty. Error bars represent standard errors.

lack of further progress in the subsequent three consecutive phases. The pattern of this improvement was similar for all age groups (no significant interaction with age).

To compare Part I and Part II of the Decision Task, participants received feedback regarding the accuracy of their decision-making during 32 trials of decision tasks with varying difficulty in Part I. In Part II, participants performed 32 decision tasks of varied difficulty but without feedback. We investigated the possibility of a transfer effect (whether performance in Part II would be better than Part I due to feedback learning) and the consistency of age and difficulty disparities across both parts of the decision task.

A 3 x 3 x 2 (Age [younger adults, middle-aged adults, older adults] x Decision Difficulty [simple, moderately difficult, difficult; within-subject variable] x Part of Decision Task [I, II; within-subject variable]) mixed ANOVA on the proportion of correct decisions yielded three main effects and an interaction effect. There was again the main effect of age, $F(2,146) = 8.06, MSE = .13, p < .001, \eta_p^2 = .099$. The performance of younger adults was marginally or significantly higher than that of both the middle-aged group (p = .054) and the older adults group (p < .001). Additionally, decreased linear trend of age on decision performance was highly significant, t (146) = 15.89, p < .001.

There was also again a main and strong effect of decision difficulty, $F(2,292) = 228.91, MSE = .07, p < .001, \eta_p^2 = .611$, with the significant differences (p < .001) between all decision tasks – the more difficult was the task, the worse the performance. Figure 2 depicts the pattern of the main effects of age and decision task difficulty.

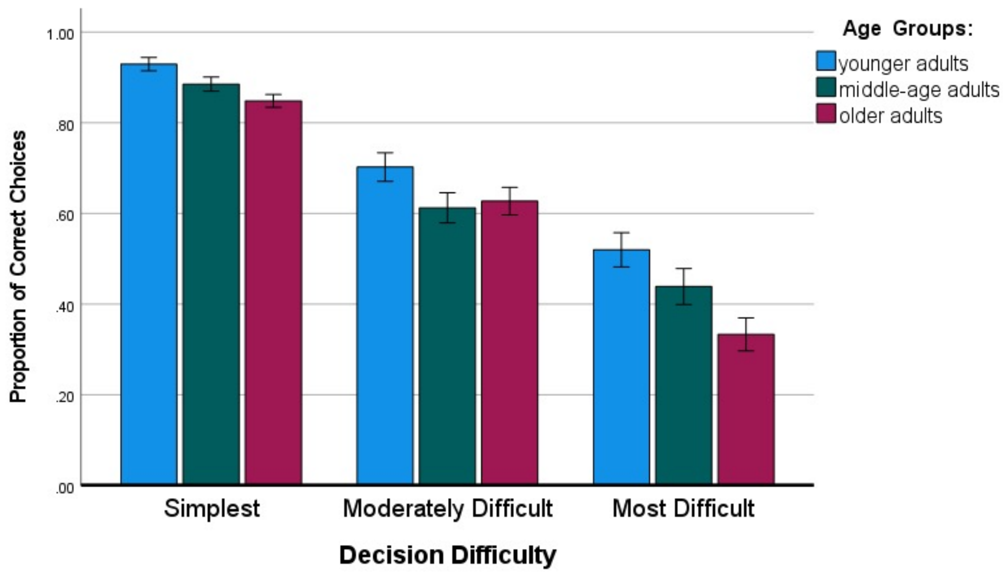

Figure 3.4: Proportion of correct choices in multi-attribute decision task as a function of age groups and decision difficulty. Error bars represent standard errors.

Additionally, linear trends of age on decision tasks of different difficulty showed strong decreased linear trends for simplest decision tasks, t (146) = 3.98, p < .001, and most difficult tasks, t (146) = 3.56, p < .001. However, such a linear trend was only marginally significant for moderately difficult tasks, t (146) = 1.69, p < .09.

Finally, a main effect of part of the decision task, $F(1,146) = 15.03, MSE = .02, p < .001, \eta_p^2 = .092$ was observed. There was a significant increase in performance in Part II of the decision task in comparison to Part I. The last two main effects were qualified by the significant Decision Difficulty x Part of Decision Task, $F(2,292) = 13.59, MSE = .02, p < .001, \eta_p^2 = .085$. There was a significant increase in performance in part II when compared to part 1 (p < .001) for simplest and moderately-difficult decision tasks and a lack of improvement for the most difficult tasks.

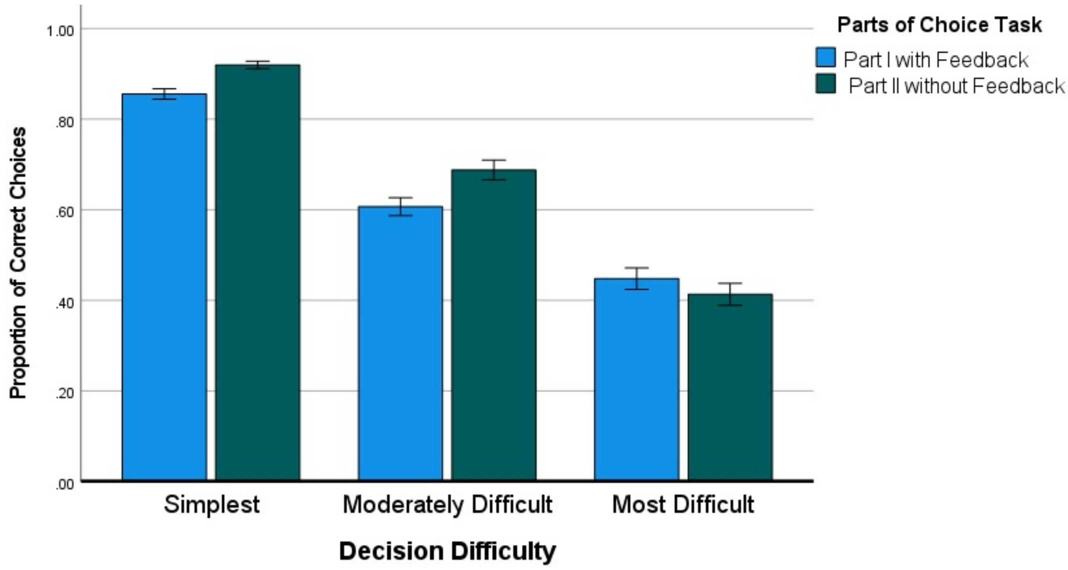

Figure 3.5: Proportion of correct choices in multi-attribute decision task as a function of parts of the choice task and decision difficulty. Error bars represent standard errors.

In summary, the main findings of the two parts of the decision tasks showed consistent differences between age groups (with younger adults performing better and a linear decrease in performance with increasing age) as well as variations in performance based on the difficulty of decision tasks (with worse performance in more challenging tasks). Notably, there was a significant improvement in performance for simpler and moderately difficult tasks in the second part of the study (without feedback), indicating a transfer effect of learning from the first part of the study (with feedback). The improvement pattern was consistent across all age groups, with no significant interaction with age. The observed age-related restrictions in the performance of multi-attribute product choice tasks, particularly in the older adults group, support previous research that typically compares only younger and older adults. Specifically, the significant limitations of older adults (compared to younger adults) in the most challenging multi-attribute choice tasks align with prior laboratory experiments [90, 91]. What is new is that in our online study imitating e-commerce shopping, we demonstrated strong decreased linear trends in both the simplest and most difficult multi-attribute choice tasks across three age groups of adult participants, including the middle-age group.

Moreover, we examined the role of quality (whether washing machine features indicate the high quality of the compared three products or rather moderate quality) and the role of delta parameter (whether washing machine features were very different for the compared three products or rather those features were pretty similar). The results of mixed ANOVA 3 x 2 x 2 (Age x Quality x Delta) across all trials of multi-attribute choice tasks did not yield any significant results.

The Role of Decision Times The prior analyses revealed significant effects of age and task difficulty on the accuracy of the multi-attribute choice tasks. The subsequent analyses aimed to assess whether participants in different age groups were able to perceive the difficulty differences of the decision tasks and allocated more time to choice tasks requiring a more advanced strategy or if they took similar amounts of time and relied on a time-consuming guessing strategy for more complex tasks. As a precautionary measure against outliers, we used medians as measures of central tendency due to a small number of excessively high decision times in an online study. We also thoroughly examined the distribution of decision time indices to potentially exclude any extreme values (none were found).

A 3 x 3 x 2 (Age [younger adults, middle-aged adults, older adults] x Decision Difficulty [simple, moderately difficult, difficult; within-subject variable] x Part of Decision Task [I, II; within-subject variable]) mixed ANOVA on the time of decisions yielded three main effects. There was a main and strong effect of age, $F(2,146) = 29.75, MSE = 161.36, p < .001, \eta_p^2 = .290$. The decision times of older adults were significantly higher than those of both the middle-aged group and the younger adults group (p < .001), no significant differences between younger adults and middle-aged groups were observed. In addition, the performed increased linear trend of age on decision time was highly significant, t (146) = 6.93, p < .001.

There was also a main and strong effect of decision difficulty, $F(2,292) = 61.66, MSE = 24.4, p < .001, \eta_p^2 = .297$, with the significant differences (p < .005) between all decision tasks – the more difficult was the task, the longer the decision time. Figure 4 presents the patterns of the main effects of age and decision task difficulty for times of multi-attribute choices.

Finally, there was also a main and strong effect of part of the decision task, $F(1,146) = 41.70, MSE = 23.39, p < .001, \eta_p^2 = .222$. A significant decrease was seen

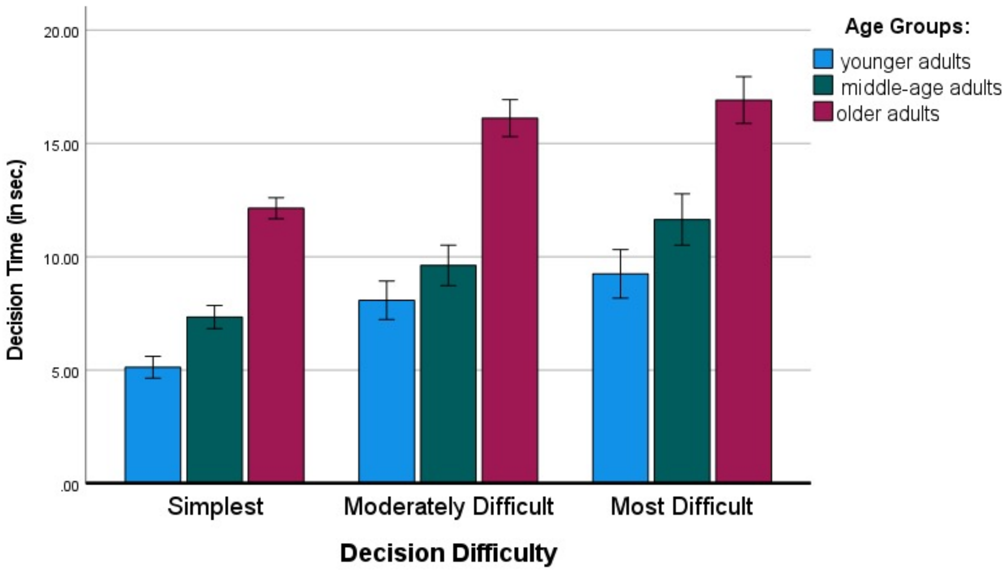

Figure 3.6: Means of decision times in multi-attribute decision task as a function of parts of the choice task and decision difficulty. Error bars represent standard errors.

in times of decisions in Part II in comparison to Part I (M = 11.74 sec. vs. M = 9.65 sec.).

To summarize the main findings regarding the time taken for multi-attribute choices, there were significant age-related differences with decision time increasing linearly across age groups, consistent with existing research on cognitive slowing across the lifespan [Salthouse, Verhaeghen, Cerellla, etc]. A novel finding was that participants of all age groups spent more time on the moderately and most difficult tasks, indicating their awareness of the increased complexity and demand for compensatory strategies. However, paradoxically, their performance was worse in these more complex tasks, despite investing more time. This suggests that participants may lack the necessary cognitive skills to solve compensatory strategies effectively, but are motivated to invest more time in decision-making. These results highlight the potential for supporting systems to improve performance in more demanding multi-attribute choice tasks.

The Effect of Age on Decision Accuracy: The Mediating Role of Visual Working Memory and Moderating Role of Helplessness of Contracting an Infectious Disease In the subsequent analyses, we applied measures of accuracy of multi-attribute choices and measures of decision times as appropriate means or medians of the simplest, moderately difficult, and most difficult decision tasks. The correlation tables between the accuracy of the decisions, age, decision time, visual working memory, and questionnaire measures (see Table 1 for correlation for the entire sample and Table 2 for correlations within three age groups indicate that only visual working memory correlated positively and significantly with decision accuracy not only for the whole sample (see Table 1) but these correlations were also significant within each age group (see Table 2). Due to the strong correlation between visual working memory and age, it was expected that visual working memory would serve as a dependable mediator between age and decision accuracy in subsequent modeling. The positive and significant correlation between Helplessness (of Contracting an Infectious Disease) and decision accuracy was observed not only for the entire sample (Table 1) but also within two age groups (younger and middle-aged adults). Additionally, Helplessness was found to have a negative correlation with age, suggesting that it would be a reliable moderator of the relationship between age and decision accuracy in subsequent modeling. The other questionnaire measures were evaluated as possible covariates for the above analyses on mediation and moderation.

The unique and strong correlations in the young adult group between decision accuracy decision time and helplessness (see Table 2) will be analyzed separately in the next section.

To examine the reliability of expected mediation and moderation effects we applied Process software [107] and conceptual model 5 indicates that aging is significantly related ($t = 7.22$, $p < .001$) to visual working memory (VWM) and VWM is a reliable mediator ($t = 4.29$, $p < .001$) of the relationship between aging and proportion of correct answers in decision task. In parallell, this mediating model was enriched by the conditional age x helplessness moderation ($t = 3.00$, $p < .01$) of the direct effect of age on decision-making. (see Figure 3.7).

We applied 10000 bootstrap samples for estimating percentile bootstrap confidence intervals. To integrate these findings, this conceptual model 5 is supplemented by two covariates. Subjective Numeracy was a significant predictor of both VWM (t

Variable	1	2	3	4	5	6	7	8
1. Age	-							
2. Acc_Dec	-.35**	-						
3. Tm_Dec	.50**	.04	-					
4. VWM	-.53**	.50**	-.30**	-				
5. Helplessness	.39**	-.30**	.10	-.33**	-			
6. S Numeracy	-.14	-.30**	-.09	.25**	-.12	-		
7. NFC_4subs	.22**	-.17*	.04	-.22**	.16*	-.01	-	
8. Corr_Features	-.02	.18*	.23**	.13	-.03	-.09	-.01	-

Table 3.2: Correlation table of dependent and independent variables. Acc_Dec = Accuracy of Decision Task; Tm_Dec = Time of Decision Task; VWM = Visual Working Memory Task; Helplessness = Scale of Helplessness of Contracting an Infectious Disease; S Numeracy = Subjective Numeracy Scale; NFC_4subs = 4 subscales of the Need for Cognitive Closure Short Scale; Corr_Features = Spearman's rank correlation between presented and subjective importance of washing machine features. *p < .05, **p < .01.

= 2.82, $p < .01$) and decision accuracy (t = 2.98, $p < .01$). The second covariate was the correlation measure of features similarities which predict the decision accuracy (t = 2.00, $p < .05$). Interestingly, the variance in decision accuracy explained by the mediator, moderator, and covariance variables was relatively high in this integrated model (R2 =.36, $p < .001$) and no significant direct effect of age was seen on decision accuracy.

In sum, the mediational model with moderation effect and covariates nicely integrate the correlational analyses and indicates the interplay between the age-related limitation and the role of additional variables (especially VWM, Helplessness) in explaining the relationship between aging and performance in the multi-attribute choice task. However, we are aware that this mediational model, although original, is still preliminary, and elucidating psychological mechanisms behind the role of VWM as

Age group	Variable	1	2	3	4	5	6	7
Younger adults	1. Acc_Dec	-						
	2. Tm_Dec	.51**	-					
	3. VWM	.50**	.27	-				
	4. Helplessness	-.33*	-.38**	-.04	-			
	5. S Numeracy	.31*	.10	.14	-.03	-		
	6. NFC_4subs	-.18	-.23	-.10	.15	-.14	-	
	7. Corr_Features	.30*	.29*	.15	-.17	-.04	.07	-
Middle-age adults	1. Acc_Dec	-						
	2. Tm_Dec	.03	-					
	3. VWM	.41**	-.31*	-				
	4. Helplessness	-.31*	.14	-.32*	-			
	5. S Numeracy	.30*	.00	.35*	-.03	-		
	6. NFC_4subs	-.11	.07	-.14	.10	.04	-	
	7. Corr_Features	.09	.22	.23	.19	-.07	.12	-
Older adults	1. Acc_Dec	-						
	2. Tm_Dec	.11	-					
	3. VWM	.32*	-.14	-				
	4. Helplessness	.05	-.08	-.16	-			
	5. S Numeracy	.20	-.17	.16	-.15	-		
	6. NFC_4subs	.04	-.03	-.15	.00	.22	-	
	7. Corr_Features	.13	.26	.08	-.09	-.16	-.22	-

Table 3.3: Correlation table of independent variables separately for three age groups. Acc_Dec = Accuracy of Decision Task; Tm_Dec = Time of Decision Task; VWM = Visual Working Memory Task; Helplessness = Scale of Helplessness of Contracting an Infectious Disease; S Numeracy = Subjective Numeracy Scale; NFC_4subs = 4 subscales of the Need for Cognitive Closure Short Scale; Corr_Features = Spearman's rank correlation between presented and subjective importance of washing machine features. *p < .05, **p < .01.

a mediator and intellectual helplessness as a moderator demands further experimental work (e.g., experimental manipulation of the intensity of a mediator and moderator, see [108]).

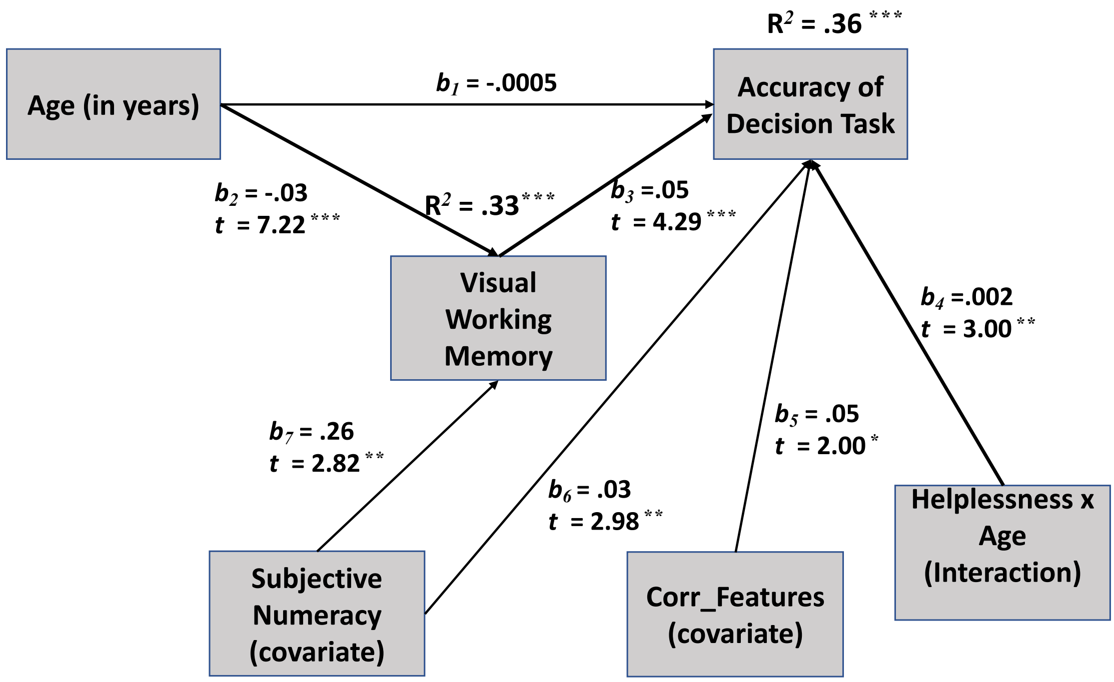

Figure 3.7: Model of the mediational effect of age via the visual working memory on the proportion of correct answers in decision task.

What is more, we noted that only in the young adults group did the decision time positively significantly correlated with decision accuracy and correlated negatively and significantly with Helplessness. To integrate this pattern of findings we examined the model in which the decision time is the mediator of the relationship between helplessness and decision accuracy (Model 4 of Process software, [107]) and with VWM as an independent predictor (see Figure 3.8). Interestingly, this model elucidates the mechanism by which helplessness impairs decision accuracy. Namely, helplessness is a predictor of shortening the decision time ($t = 2.88$, $p < .01$), and shortening the decision time decreases the decision accuracy ($t = 2.59$, $p <.01$). As explained in the theoretical introduction, helplessness state not only impairs the complex cognitive functioning but also abolishes the intrinsic motivation, hence solving complex decision tasks might be aversive in helplessness state and participants might be motivated to decrease the time of decision making. Independently, the visual working memory acts as a cognitive variable increasing the accuracy of decision-making.

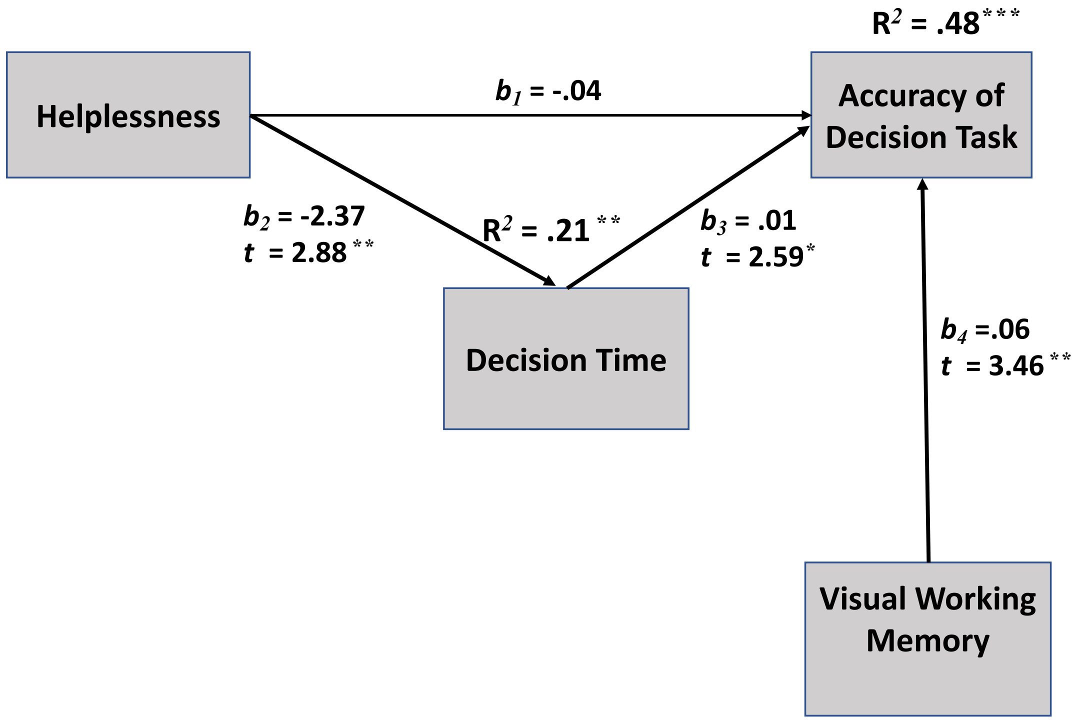

Figure 3.8: Model of the mediational effect of helplessness via the decision time on the proportion of correct answers in decision task.

Conclusions In this experiment, we have further researched the question of how the cognitive limitations of older consumers affect their choices in e-commerce product comparison tasks.

We have confirmed again significant aging effects, where the youngest group outperformed the oldest and middle group in all 3 difficulty-level tasks. The effect was particularly linear for the simplest and the most difficult tasks, while in the moderately difficult tasks, the middle-aged and older adults performed on a similar level. We have observed a strong improvement in performance in the second phase of the experiment which we interpret as a learning effect based on the feedback in the first phase of the experiment. We also observed material differences in decision time across different age groups, correlated with the task difficulty. This shows that although the older group is willing to spend more cognitive resources on the more difficult tasks, the higher effort does not lead to better decision quality.

3.3.4 Conclusions from the experiments' results

The findings of the two experiments showed that in both settings there was a significant aging effect - the percentage of correct choices was lower among the older group of participants. Given the experimental settings were designed in order to closely reflect the typical multi-attribute choice on e-commerce platforms supported by comparison tables, this **confirms Hypothesis 1: Due to the cognitive limitations of older customers, their choices are less optimal than those of younger customers.**

The first experiment demonstrated the mediating role of Helplessness of Contracting an Infectious Disease, while the statistical model applied in the second experiment found it to be a moderator. Both results indicate that older participants were more taxed by this form of intellectual helplessness than the younger ones (see Section2.2.3). These results might potentially be extrapolated to other types of cognitive exhaustion, like for example choice overload, higher difficulty in navigating user interfaces, etc., making older customers more prone to errors.

The comparison of the experimental settings is presented in table 3.4.

Experiment feature	Experiment 1	Experiment 2
Participants	135, 49 younger, 45 middle-aged, 41 older	149, 50 younger, 45 middle-aged, 54 older
Time of the experiment	1-2h	1-2h
Decision-making strategies	WADD, TALLY, TTB	WADD, TALLY, TTB
Consideration sets size	3	3
Tasks sets	32	64
Feedback	NO	YES, first 32 tasks
Performance-related additional reward	NO	YES
Questionnaires	Scales of Helplessness and Anxiety of Contracting an Infectious Disease, Scales of Subjective Numeracy and Objective Numeracy	Scales of Helplessness of Contracting an Infectious Disease, Scale of Subjective Numeracy, Need for Cognitive Closure, Measure of Visual Working Memory

Table 3.4: Comparison of the multi-attribute choice experiments

CHAPTER 4

Aging Agent Model

4.1 Model description

This section outlines the structure and specifications of a simulation model that integrates the impact of cognitive aging on e-commerce customers. The model is a significant addition to this thesis and can be leveraged to examine how cognitive constraints influence consumer buying decisions, contentment, and the efficacy of recommendation algorithms. Every component and parameter of the model is grounded in findings from research in aging psychology, consumer conduct, or our own empirical investigations.

4.1.1 Model overview

The model, build using Python 3.6 language[1], comprises three salient elements: Agents, Items, and Recommender Systems. The purpose of the model is to allow an exploration of the benefits and limitations of various designs of recommender systems. Agents are modelled based on older consumers' decision-making characteristics explained in psychological and cognitive research, as well as data collected in our experiment (see Section 2.1.5). Items are real-life products webscrapped from a popular e-commerce platform. Recommender systems will be a target for detailed and realistic modeling.

[1]The source code is available in the repository: https://github.com/Justyna-P/Cognitive-ageing-model

Model feature	Agent-based model 1.0 [8]	Agent-based model 2.0	ACT-R [111]
Individual product preferences	YES, randomly assigned	YES, sourced from experiment	YES
Decision-making strategy	NO	YES	YES
Working memory	YES	YES	YES, complex model
Emotion-based heuristics	YES, chosen by author	YES, research-based calibration	NO
Consideration sets size	NO	YES	NO
Product sets	Randomly generated	Based on real products	NO

Table 4.1: Comparison of simulation models of cognitive processes involved in decision making

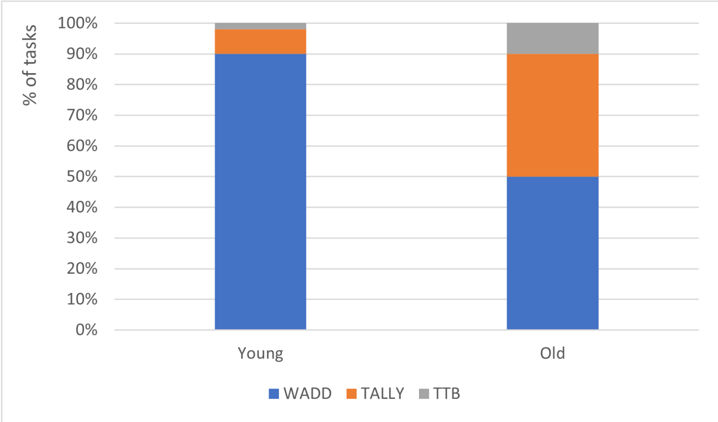

Figure 4.1: Distribution of decision strategies in multi-attribute choice problems among older and younger adults. Source: [91].

We also propose and test three new recommender system designs aimed specifically at dealing with self-induced bias.

The model is a more advanced version of the model presented in our previous research [8]. Table 4.1 compares the first version of the model with the current version used for this thesis. For reference purposes, we compared the model with a well-established unifying cognitive architecture - ACT-R. Given that this architecture is designed primarily to model cognitive processing on a lower level, it has improved precision in simulating memory retrieval. For the purpose of modelling customer decisions, we developed our own simplified architecture that includes higher-level processes, the role of emotions, as well as heuristics.

4.1.2 Experimental determination of model parameters

To retrieve empirical data for our model, we conducted an online experiment that investigated the behavior of older users in product comparison tasks which results are described in 3.3.3. Besides studying the age effect on the multi-attribute choice tasks, the experiment was used to obtain the data for better calibration of the simulation model.

The participants were redirected from a research participant recruitment agency portal to the experiment website. At the study's commencement, participants filled out a

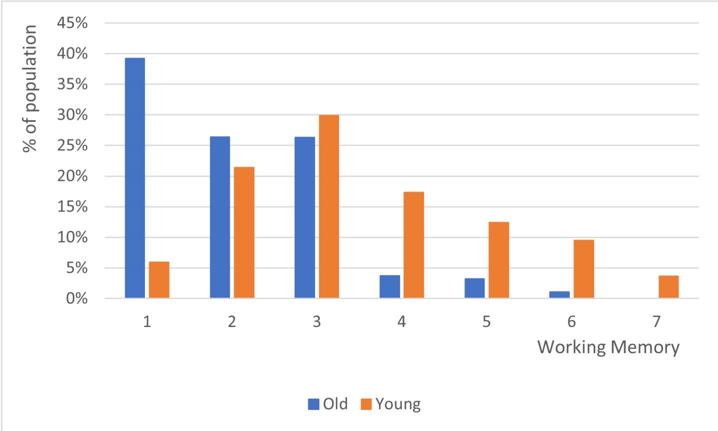

Figure 4.2: Distribution of working memory among older and younger adults. Source: [48].

personal questionnaire that asked for their ages. **There were 76 users in the youngest group (aged 19-30), 80 in the middle age group (42-53), and 87 in the oldest group (65-76)**.

Participants were requested to provide answers to inquiries concerning their individual inclinations with respect to product characteristics (the items used in the research were washing machines), which were subsequently utilized to fine-tune the simulation model employed in this thesis. In the initial survey, participants evaluated the significance of attributes utilized in the product comparison duties by assigning each feature a rating of 1 to 6 pluses. This information constitutes the groundwork for the product preference distribution utilized in our simulation model.

In another questionnaire, participants were asked how many different products they would review before deciding which one to purchase. This data is also used in the simulational model.

Later, participants solved a visual pattern memory task successfully applied in the large internet study with participants across adult life-span [109] to measure working memory capacity. The users exhibited a pattern on a rectangular matrix, which they were supposed to subsequently recreate on an empty matrix. As the number of patterns correctly recalled, performance was scored.

attribute	A0	A1	A2	A3	A4	A5	A6
importance for agent	6	4	5	6	4	2	4
item's$_1$ normalised attribute value	1	0	0.5	1	0	0.2	0

Table 4.2: Example of an Agent's preferences and normalized product attribute values

Our research is premised on a multi-agent simulation model described in the next section. The data collected in our empirical experiments were used to determine some key model parameters. **Each participant of the experiment was represented by a single Agent belonging to the same age group as the real counterpart**[2]. The weights for the Agent's preferences function were derived directly from the user ratings of attribute importance obtained from the first questionnaire. The size of consideration sets for each Agent was obtained from the questionnaire about the number of considered products, which provided the model with this important parameter.

4.1.3 Model design - agents

Every agent in our model has the following properties:

1. individual product preferences - obtained from our own experiments 4.1.2
2. a decision-making strategy - randomly chosen from distribution of decision-making strategies from [91]
3. working memory size - randomly chosen from distribution of age-dependent working memory size from [48]
4. susceptibility to biases (heuristics): brand or negative reviews based on research: [112], [113]
5. number of considered products (size of the consideration set)- - obtained from our own experiments 4.1.2

[2] The youngest and middle groups both were modelled as 'young' agents

4.1.4 Individual preferences and preference function

Every agent has a unique set of preferences, expressed as weights of each attribute of an item. The 'real' utility of an item for the agent is a weighted sum of the item's normalised attribute values weighted by the agent's preferences. To illustrate, for the item displayed in table 4.2, the utility for the agent would be $6*1+4*0+5*0.5+6*1+4*0+2*0.2+4*0=14.9$. However, we assume that the agents do not make decisions perfectly owing to cognitive limitations, hence they use simplified decision-making strategies instead of performing the full utility computation.

4.1.5 Strategies in multi-attribute choice problems

Every agent in our simulator has a decision-making strategy. Recall from Section 3.2.2 that three decision-making strategies have been examined in studies on cognitive aging, namely Weighted Additive (WADD), TALLY, and Take The Best (TTB). These strategies vary in complexity, with WADD being the most complex and TTB being the simplest. The WADD approach is considered optimal but involves computing the total weights of attributes that make a product superior in comparison to others. TALLY involves counting the number of these attributes for each product, while TTB only requires comparing values of the most significant attributes (unless there is a tie, in which case the next attribute in the order of importance is considered).

Agents can be assigned decision strategies randomly from the distribution shown in Figure 4.1. The age-related proportions of decision strategies utilized for multi-attribute choice problems (described in 3.2.2) were sourced from the experiment of Mata and colleagues [91]. When WADD was identified as the most optimal strategy for a task, young participants could use it in 90% of tasks, while older participants used it in only 50% of tasks. TALLY was applied in 8% of the tasks in the younger group and in 40% of the older group, while TTB was in 2% and 10%, respectively.

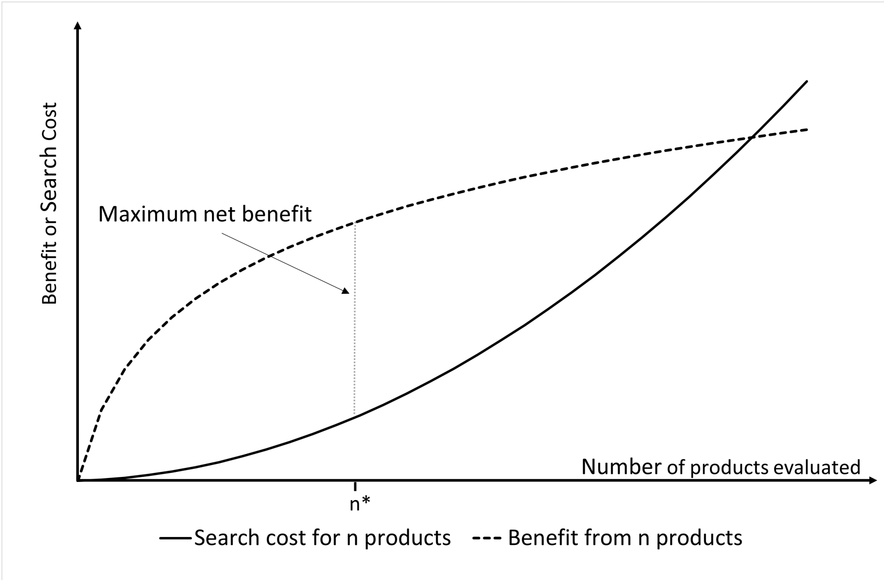

Figure 4.3: Rationality of the consideration sets. Source: [114].

4.1.6 Working memory

The working memory concept (see 2.2.1) and its decrease with age play a key role in modelling fluid intelligence and its impact on the decision-making process. The working memory size parameter specifies how many attributes can be remembered and compared by the Agent at the same time In the model of the Agent. Multi-attribute choice decisions require assessing many attributes at once. A distribution of average working memory in younger (20-30) and older (60-70) adults, derived from [48], is displayed in Figure 4.2. Agents in our simulation were assigned working memory parameter values randomly from this distribution.

4.1.7 Affective Heuristics

The model so far was built under the assumption that the cognitive capacity of Agents may be limited, but their decisions are not influenced by emotions. A vast body of research demonstrates that emotions have a strong impact on our decision-making process. Affective heuristics cause a tendency to deviate from a logical choice in a systematic way, observed repeatedly in many individuals. Older and younger adults

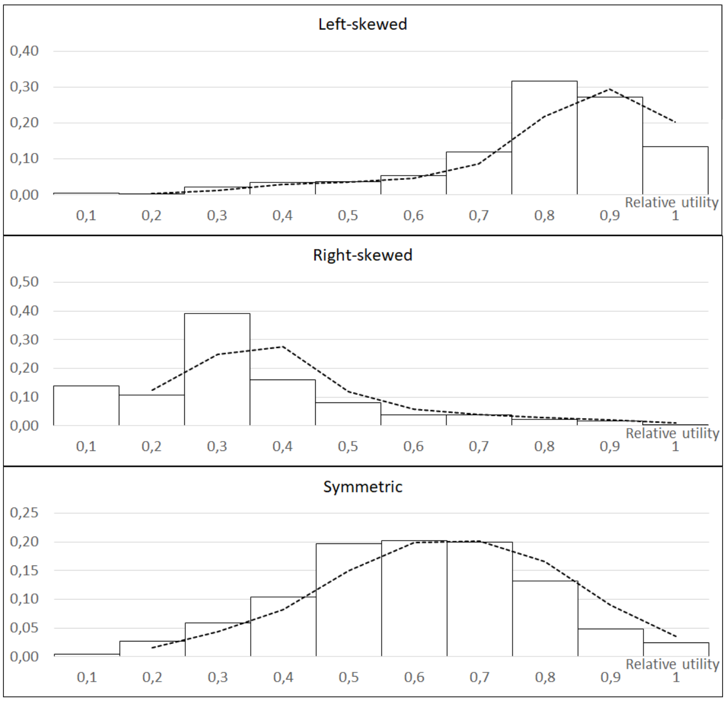

Figure 4.4: Distribution density of product's relative utility in 3 product sets

may be impacted by affective heuristics to a varying degree. The first affective heuristic incorporated in the model is an overly strong reliance on affect-rich reviews. Research conducted by von Helversen and colleagues [112] shows that when comparing two products, one of which has a significantly better quality and average consumer rating, 18% of younger adults and 31% of older adults choose the worse product when the better one had a single highly negative, vivid, and affect-rich review. At the same time, 12% of young adults chose the worse product if it has a single enthusiastic review, which was not observed among older participants. To reflect this in the model, we added a 'sensitive to negative review' attribute to Agents, and randomly choose 18% of younger and 31% of older Agents for which the parameter was set to 1 (for the rest of the Agents the parameter was 0). While assessing the products with negative reviews, an Agent subtracts 100 from the perceived utility, implying that the product will never be chosen over one without such a review.

A second affective heuristic present with different frequencies among older and younger customers is excessive brand loyalty. Research-based on car purchase data [113] has revealed that even among customers who are dissatisfied with their previous experience with a car from a given brand (satisfaction score 3/10 and below), 38% of younger and 50% of older customers still went ahead and purchased the product from the same brand. We added a 'brand sensitivity' attribute to our Agent model and assigned a positive value for 38% younger and 50% older Agents. The brand to which each agent is loyal was assigned randomly among 13 different washing machine brands present in the dataset proportionally to the number of models within each brand. The Agent adds 100 to the perceived utility while assessing the product belonging to Agent's favorite brand, meaning that the product will always be chosen over one not belonging to the brand.

4.1.8 Consideration sets

A developed market offers a variety of products to consumers. When taking into consideration durable, complex products such as washing machines, computers, or cars, there are hundreds of available models. When consumers face a large number of alternative products, as is increasingly common in today's retail and web-based shopping environments, they generally reduce the full set of products to a smaller, more-manageable consideration set which they evaluate further[115]. This concept has a strong foundation in both experimental research and marketing theory and can be explained through a basic model that considers the diminishing returns of additional search efforts and the increasing cost associated with the search (see 4.3). Experiments[116] have shown that the average size of the consideration set is usually small – 9.3 for mobile phones, 7.8 for handheld GPSs (4.8 standard deviation), and nearly 10 for cars. The questionnaire collected during the experiment described in Section 2.1.5 resulted in similar numbers.

4.1.9 Enviromental parameters

There are 268 products, scrapped from one of the most popular e-commerce sites. The items are presented to the user in a random order. Each user has one product fitting his preferences best, in comparison to the others; the product's utility for this user is the largest she can possibly achieve. This utility is used as a benchmark against which other products are measured.

Age group	average number of products considered
1. (20-30 years)	19.63
2. (42-52 years)	13.45
3. (65-76 years)	10.45

Table 4.3: Consideration set sizes in various age groups.

As Figure 4.4 shows, there are three different sets of products. First, left-skewed is based on the original products web-scrapped from an e-commerce page. The relative utility distribution is mainly concentrated around higher values, indicating that producers are competing with each other to provide products that are best suited to the needs of the customers. To assess the performance of recommender systems in less developed markets, we also generated product sets where most of the products are not very suitable for the customers' needs, with only a few exceptions. The third product set falls somewhere in between these two extremes, with a majority of products having moderate utility, and only a few items being either very well or very poorly suited for the customers.

4.2 Simulation description

4.2.1 Simulation Flow

An agent enters the e-commerce platform with the aim to purchase a product. After entering their request into the search engine, they are presented with a set of items (P1, P2, ...Pk). The agent filters the items using minimum or maximum values of product attributes and then browses the remaining items one by one. At the commencement of the "shopping session", each agent chooses features that will be considered while making their decision. This is necessary due to limited working memory capacity; most agents are not able to simultaneously compare all the product features, so an agent with working memory capacity of 3 compares and remembers values of only 3 features most

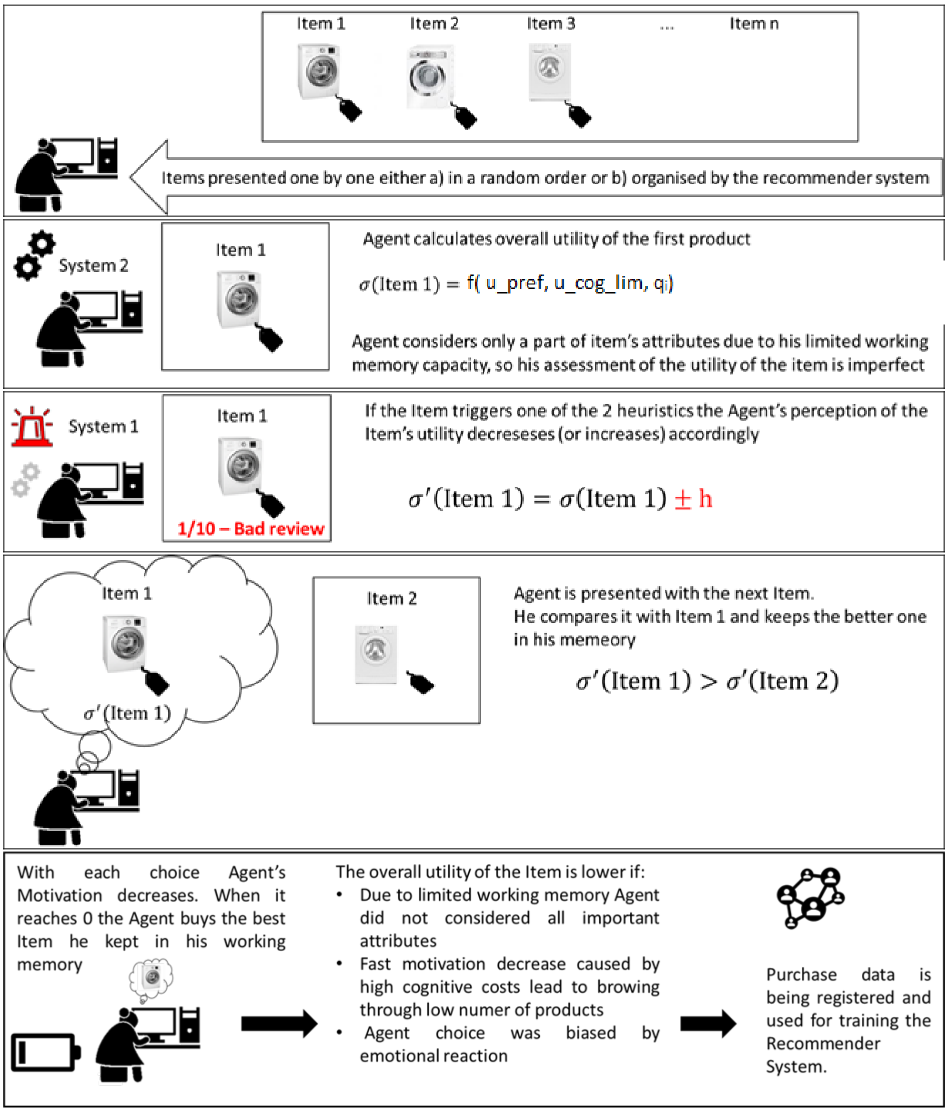

Figure 4.5: The main simulation procedure. Legend: σ (Item x) – real overall utility of the Item x for a given Agent

important for her. Knowing an agent's utility function and preferences, we can calculate both the "real" utility of the chosen product, which takes into account all product features, as well as "perceived" utility, which takes into account only selected features and is modified by agents' biases and heuristics.

Every action reduces an agent's motivation level, and when an agent's motivation level reaches 0, she chooses the best product (based on her "perceived" utility). The order in which the agent browses the products is important, since only the first n products will be considered by the agent, since the agent can only review a limited number of products. The simulation flow can be divided into two versions. In the first version, the products meeting the minimum criteria (simple product filtering) are presented to the agent in a random order. In the second version, the order in which products are presented is determined by a recommendation algorithm.

Simple product filtering

Simple product filtering used in the first version of the simulation reflects the behaviour of a user who knows that she is unable to browse through all the products available, and does not want to see the products that do not meet chosen criteria. Based on the assumption used in the simulation, all products that have a value worse than the 10th percentile value for each of the features remembered by the user as important are excluded.

4.2.2 Modeling decision support algorithms

Apart from the simple product filtering, we simulated an e-commerce platform using more complex decision making aids, namely recommender algorithms. The recommender algorithms used in the second part of the simulation determine the order in which the products are presented to the agent and are described in detail in chapter 5. **Recommendation algorithms use purchase data from the first part of the simulation as a training set**. Following the training, an algorithm proposes the product that best matches the agent's preferences as the first displayed item, the second-best as a second item, etc. Accordingly, the order in which an agent browses filtered products is changed by the recommendation algorithm.

4.2.3 Model validation

A simple simulation was designed to compare the results achieved in the simulated environment versus empirical ones in order to validate the model. The principle of the simulation was to make each agent 'buy' one from the 3 presented products **which attribute values were directly taken from Experiment 2 tasks sets** (see Section3.3.3) This was repeated for 10 subsets of 3 products representing simple and 10 subsets signifying the most onerous task.

To ensure the correct answer, the experiment's importance weights were used instead of the agents' subjective preferences. Working memory and strategy choice algorithms were kept the same as in simulation, but affective heuristics and consideration set of the AAM were not applicable. Then, the simulation was run treating each set of 3 attributes as a full set of available products. If the agent's final purchase at the end of each short 3-item simulation was the item that was the correct answer in the task, the choice was marked as 'correct', otherwise 'incorrect'.

The results are presented on 4.6. Simulated agents achieved average scores that were similar to those of real experiment participants. The simulation replicated the age effect, with older agents achieving lower correctness scores on average compared to younger agents. The difficulty of the task also affected performance in the simulation, with a higher proportion of correct answers for simpler tasks compared to more difficult ones. However, age-related differences in the simulation were larger than in the experiment. This could be due to simplifications in the decision-making process that did not account for factors such as motivation, learning, or individual differences. Despite this, the validation results confirm that the model adequately represents reality for research purposes. The age-related differences in the simulation are larger than the ones observed in the experiment. which may be attributed to the necessary simplification of the decision-making process, not considering other factors such as motivational aspects, learning effects, or individual differences; however, the achieved in the validation structure of results convinces us that the model's representation of reality is sufficient for the research purposes.

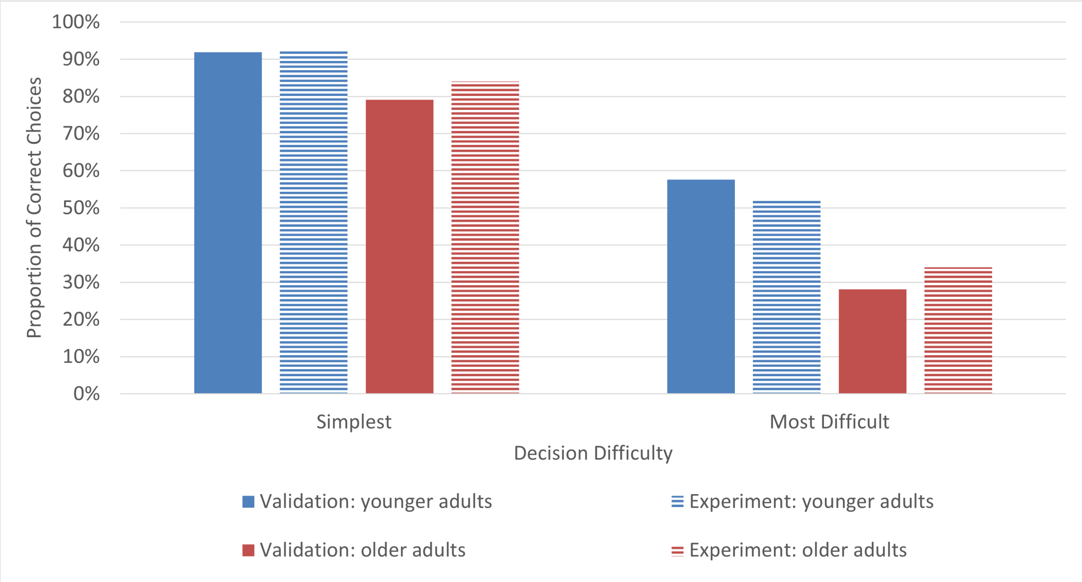

Figure 4.6: Proportion of correct choices in experimental multi-attribute decision task and validation simulation as a function of age groups and decision difficulty

4.3 Results of the simulation - baseline

The results of simulations without recommender systems can be used to describe the differences between "young" and "old" agents, and to introduce baseline results.

The aggregated results of simulations in which no recommender systems were used are shown in Figure 4.7. The x-axis shows the size of the consideration set. The y-axis shows the average utility of selected products. The dotted line shows the average utility obtained by an agent who browsed respectively 1, 2 ...n items. The solid line represents older agents. The simulation is conducted three times, firstly, on the left-skewed product set (black lines), secondly, on the right-skewed product set (gray line), and lastly on the symmetrical product set.

In Figure 4.7, the age difference is apparent: younger agents achieve on average 5 percentage points higher utility for each number of compared products, compared to the older customers, in the left-skewed product set scenario. When the product set is less optimized for customers' needs, the difference in performance is even more significant. For a symmetrical product set, the difference is an average of 6 percentage points for

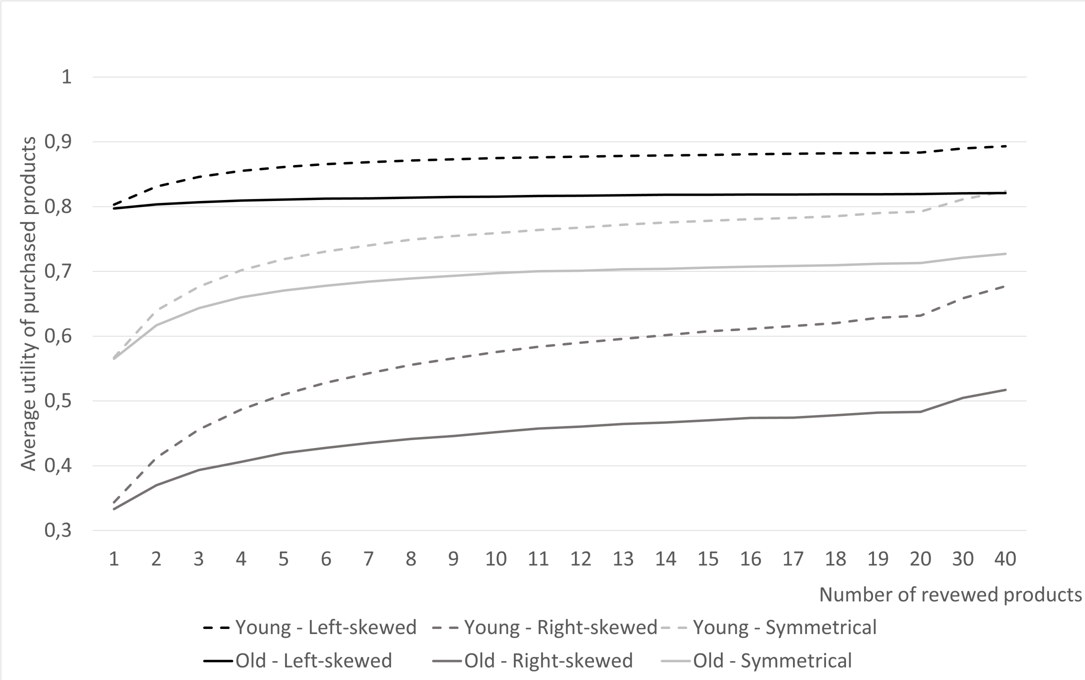

Figure 4.7: The average utility of a purchased product in a simulation without a recommender system

every consideration set size. In the case of a right-skewed product set, the difference increases to 11 p.p.

The age effect is even more pronounced when age-related differences in the size of the consideration sets (see Table 4.3) is considered. As the number of reviewed products increases, both younger and older customers experience an increase in achieved utility. However, older customers tend to stop their search after reviewing an average of 10.45 items, while younger customers consider up to 19.63 items on average. This leads to a higher utility for the final purchased product among younger customers, with a difference of 6 percentage points for left-skewed product sets, 9 percentage points for symmetrical product sets, and a significant 18 percentage points for right-skewed product sets.

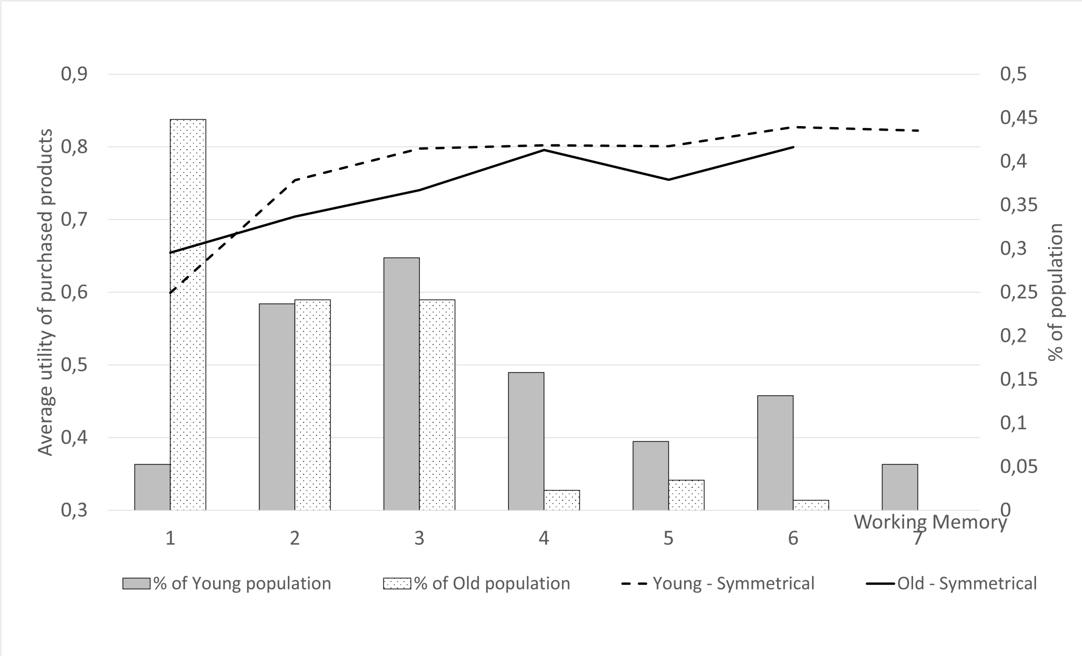

Figure 4.8: The average utility of a purchased product in a simulation without a recommender system - working memory effect

4.3.1 Working memory effect

The fact that the older agents have a lower working memory capacity impacts their ability to achieve a high utility of a purchase. Figure 4.8 shows that both younger and older users achieve higher average utility with a rise in their working memory capacity, although the marginal gain is diminishing. The fact that 47% of older users are incapable of simultaneously processing values of more than one attribute is congruent with findings described in [48]: the only strategy available for such users is effectively TTB.

It is noteworthy that the distribution of working memory among younger users is quite broad as well; it is quite common for younger users to have a low working memory. As an implication, the younger users (and younger agents in our simulation) can also make sub-optimal purchasing decisions and could also benefit from specialized support.

4.3.2 Impact of affective heuristics

Table 4.4 shows The impact of affective heuristics on the average achieved utility. The utility achieved by older agents under the impact of affective heuristics is only 1 p.p. lower than the utility of the other agents, which shows that affective heuristics do not significantly harm their users and that it might be a way of saving already limited cognitive resources. The difference is more prominent for younger users, as a young agent that uses affective heuristics achieves 5 p.p. lower utility than young agents who do not use them.

Age	No heuristics	Affective heuristic	Total
Old	0.82	0.81	0.82
Young	0.91	0.86	0.88

Table 4.4: The average utility of a purchased product in a simulation without a recommender system - the impact of affective heuristics
All differences are statistically significant with p-values <0.01. The test used was the Mann-Whitney U test with sample sizes >10000.

CHAPTER 5

Effectiveness of different types of Recommender Systems

5.1 Content-Based

5.1.1 Algorithm specification

The CB (content-based) algorithm used in this research relies on a matrix of item-to-item cosine similarity, as well as the user's past purchase data. Specifically, for each item that a user has previously purchased, the algorithm generates a list of the 10 most similar items. These lists are then aggregated, and the most frequently occurring items are recommended to the user.

Algorithm 1 Content-based algorithm

function RECOMMENDATION ALGORITHM (NUMBER OF ITEMS)
    $u \leftarrow$ user
    $uList \leftarrow$ *user list*
    $iList \leftarrow$ *items list*
    $k \leftarrow numberofitems$
    $p(u) \leftarrow$ list of items purchased by user u
    $attr(n) \leftarrow$ values of attributes of item i
    for each item i in iList:
    decreasingly **sort** *iList*\j by the similarity of *attr(j)* to *attr(i)*
    $topp \leftarrow$ sum of *p(u)* for *u* in top *k* on *uList*
    recList $\leftarrow$ most frequent items in topp
    recommend items from *recList* to user n

procedure CONTENT-BASED
    call Recommendation algorithm (*number of items*)

5.2 Collaborative Filtering

5.2.1 Algorithm specification

For this simulation, we implemented a popular latent factor model: Singular Value Decomposition (SVD). The number of factors k to factor the user-item matrix is an important parameter in this approach. A higher number of factors results in higher precision while lowering the number of factors raises the model's generalization. After running the simulation with k in range (5, 10, 15, 20) it was determined that k = 15 results in the highest user's utility, which is why this parametrisation was utilised for the comparison with other recommender systems.

Algorithm 2 Collaborative-filtering algorithm

```
function RECOMMENDATION ALGORITHM (NUMBER OF FACTORS)
    k ← number of factors
    u ← user
    uList ← user list
    iList ← items list
    p(u) ← list of items purchased by user u
    attr(n) ← values of attributes of item i
    p(u) ← list of items purchased by user u
    user-item-matrix ← aggregate p(u) for all u from uList
    U, sigma, Vt ← Singular Value Decomposition (user-item-matrix, k)
    all-user-predicted-ratings ← ((U · sigma) · Vt)
    pred(u,i) ← predicted rating of item i by user u
    for each user n in uList:
        predList ← aggregate all pred(n, i)
        decreasingly sort predList by rating value
        recommend items from predList to user n

procedure COLLABORATIVE-FILTERING
    call Recommendation algorithm (number of factors)
```

5.3 Results

The effectiveness of the two most popular recommender system types is presented in Table 6.1. The table displays the results obtained solely from the optimal parameter settings for the CF and CB recommender algorithms

Table 5.1 also shows the results of a sensitivity analysis to the distribution of product utilities. As a reminder, the dataset of washing machine utilities obtained from our experimental study is heavily left-skewed. While this distribution may not be representative of all recommender system applications (for example, a user searching for a budget-friendly product), we conducted simulations with left-skewed, right-skewed, and symmetric utility distributions for products. The simulations showed that the most notable results were observed for the right-skewed and symmetric distributions.

The data shows that well-configured CF systems are capable of improving the average utility achieved by both old as well as young agents, respectively by 2 and 3 p.p. A similar improvement is not achieved in the case of the CB system: without the recommendation system, its usage in the best available configuration yields the same utility as achieved, and some configurations of the CB algorithm (not shown in the table) are actually diminishing the utility achieved by the agents.

Contrary to the results obtained in pilot versions of the simulation (see [8] the results obtained by both, old and young user groups using the traditional recommendation systems were better than the baseline. The differences in the results obtained can be attributed to variations in the simulation setup, including the design of the agents' decision-making process and the product sets used. Additionally, when presenting the final results, we only considered the best-performing versions of the recommended systems to ensure that traditional algorithms were not at a disadvantage when compared to newly proposed ones. To summarize, **the result derived in the presented version of the model is incapable of supporting the hypothesis that recommendation algorithms trained on suboptimal can exacerbate the users' decision quality**. Else, it is at least not observed for the total aggregated older population.

5.4 Quantifying Self-Induced Bias

5.4.1 Self-Induced Bias Calculation

As discussed in section 2.3.2, self-induced bias occurs when users provide a recommendation algorithm with suboptimal decision data due to their limited decision-making process, e.g. using heuristics, simpler strategies, and limited working memory. The recommendation algorithm, when trained on this input, may suggest items that are less suitable for a user's needs compared to users with similar needs who do not have the same limitations. To quantify the extent of self-induced bias, there are several key elements required:

Age, metric	Baseline	CB*	CF*
Symmetric distribution of product utility			
Old, absolute	0.70	0.73	0.77
Old, relative**	NA	0.04	0.07
Young, absolute	0.79	0.80	0.86
Young, relative**	NA	0.01	0.07
Left-skewed distribution of product utility			
Old, absolute	0.82	0.82	0.84
Old, relative**	NA	0.00	0.02
Young, absolute	0.88	0.89	0.91
Young, relative**	NA	0.00	**0.03**
Right-skewed distribution of product utility			
Old, absolute	0.45	0.54	0.60
Old, relative**	NA	0.09	0.15
Young, absolute	0.63	0.68	0.72
Young, relative**	NA	0.05	0.09

Table 5.1: The average utility of a purchased product in a simulation with a recommender system, using a left-skewed, right-skewed, and symmetric distribution of product utility
*best out of all parameters used in the simulations
**difference between the utility obtained using the recommender system and the utility without the system.
All differences are statistically significant with p-values <0.01. The test used was the Mann-Whitney U test with sample sizes >10000.

1. a group of users who share a specific characteristic impacting their decision-making ability, e.g. group of older users impacted by cognitive aging,
2. an algorithm that uses their past decisions as a training set for future recommendations,
3. a reference group as a baseline for comparison,
4. a way of measuring the quality of decisions: with and without the impact from the recommendation algorithm for both reference and impacted group.

We are able to generate all of the listed elements using AAM simulation. Algorithm 3 presents the steps for quantification of the self-induced bias in a simulated setting.

Algorithm 3 Quantifying self-induced bias caused by a recommender system

```
function U(USERSGROUP)
    u(usersGroup, x) ← average utility achieved by users in usersGroup under x condition
function SELF-INDUCED BIAS (USER-GROUP-CRITERIA, RECOMMENDATION ALGORITHM)
    allList ← all users
    impList ← users meeting user-group-criteria
    refList ← users not meeting user-group-criteria
    x-b ← baseline scenario
    x-rec ← scenario where users decisions are made under recommender system impact
    Self-induced-bias = [u(refList, x-b) - u(impList, x-b)] - [u(refList, x-rec) - u(impList, x-rec)]
```

5.4.2 Self-Induced Bias of CB and CF algorithms

The self-induced bias of the examined recommender systems, resulting from age-related changes in cognitive functions, can be measured by computing the disparity in absolute utilities attained by younger and older agents, as demonstrated in Table 5.2. The initial column displays the contrast in utilities between younger and older agents when basic filtering is employed without any recommendation system. This value serves as a benchmark for self-induced bias. The following columns demonstrate the self-induced bias when various recommendation systems are employed. In the instance of

Product preference distribution, metric	Baseline	CB*	CF*
Symmetric distribution of product utility			
Absolute*	0.09	**0.07**	0.09
Rec-sys**	NA	**-0.02**	0.00
Left-skewed distribution of product utility			
Absolute*	**0.06**	0.07	0.07
Rec-sys**	NA	+0.01	+0.01
Right-skewed distribution of product utility			
Absolute*	0.18	0.14	**0.12**
Rec-sys**	NA	-0.04	**-0.06**

Table 5.2: *Self-Induced Bias: Difference between the absolute performance of young and old agents
**Self-Induced Bias Caused by the Recommender system: difference vs baseline

the left-skewed distribution of utility, Content-Based, Collaborative Filtering algorithms, and the newly proposed Decision-competency based algorithm perform worse than the baseline. In contrast, the CB and CF algorithms slightly enhance the baseline for the right-skewed and symmetric datasets.

Chapter 6

Proposal of New Recommendation System

6.1 Preferences Twin

The Preferences Twin algorithm is the first of two algorithms proposed in this thesis that aims to explicitly combat self-induced bias. This algorithm uses explicit user preferences and chooses a younger user with similar preferences to the older one.

To identify agents with similar preferences, the algorithm utilizes a standard k-NN algorithm with a k value of 1. The algorithm purposely selects only "young" agents from among these similar users to avoid suggesting items chosen through a more biased and less efficient decision-making process by "older" agents. The algorithm then suggests items that were purchased by the most similar "young" agents to the agent in the simulation.

The Preferences twin algorithm relies on a method for evaluating user preferences. In our model, we have adopted an approach where preferences are represented as weights assigned to the most significant product attributes. However, there are various decision support research methods available for expressing user preferences. A systematic review of them [117] highlights 2 methods traditionally used for the purpose- Value Function Elicitation and Analytic Hierarchy Process and provides examples of using these methods in the e-commerce space. Systems that utilize a trusted network often incorporate similarity in opinions among users to enhance the quality of recommendations [118].

Algorithm 4 Proposed new algorithms

```
function RECOMMENDATION ALGORITHM (SELECTION CRITERIA)
    u ← user
    uList ← user list
    p(u) ← list of items purchased by user u
    pref(u) ← list of preferences of user u
    age(u) ← age of user u
    wm(u) ← working memory capacity of user u
    for each user n in uList:
        decreasingly sort uList\u by the similarity of pref(u) to pref(n)
        remove users from uList based on (selection criteria)
        topp ← sum of p(u) for u in top k on  uList
        recList ← most frequent items in topp
        recommend items from recList to user n

procedure PREFERENCES TWIN
    selectionCriteria ← age(u) = "old"
    call Recommendation algorithm (selectionCriteria)
procedure DECISION-COMPETENCY BASED
    selectionCriteria ← wm(u) < 5
    call Recommendation algorithm (selectionCriteria)
```

6.2 Decision competency-based

The Decision-competency based algorithm is the second algorithm proposed in this thesis. This algorithm utilizes explicit information regarding a user's decision-making abilities. Due to the extensive input data required for our cognitive ability model, we have chosen to concentrate on one of the most critical model parameters, namely working memory size. The Decision-competency based algorithm selects a user with similar preferences, but a working memory of at least 5.

To implement the Decision-competency based algorithm, information about a user's working memory size is required in addition to their preferences. Fortunately, there are various effective and well-researched tests available for measuring working memory, such as the visual pattern span task outlined in Section 4.1.2 [109].

6.3 Collaborative Filtering With Working Memory Threshold

Finally, I would like to introduce Collaborative Filtering With a Working Memory Threshold algorithm (CFD-WMT). The core of this algorithm is the collaborative filtering (CF) algorithm, which was implemented in Section 5.2.1. Upon the implemented algorithm I imposed a new restriction, the same as for the new algorithm described in Section 6.2; the user-item matrix was newly generated for each user taking into account only decisions made by the users with working memory of at least 5.

Algorithm 5 CFD-WMT algorithm

function RECOMMENDATION ALGORITHM (NUMBER OF FACTORS, SELECTION-CRITERIA)
    $k \leftarrow$ *number of factors*
    $u \leftarrow$ user
    uList $\leftarrow$ *user list*
    iList $\leftarrow$ *items list*
    $p(u) \leftarrow$ list of items purchased by user u
    $attr(n) \leftarrow$ values of attributes of item i
    $p(u) \leftarrow$ list of items purchased by user u
    user-item-matrix $\leftarrow$ **aggregate** p(u) for all u from uList
    for each user n in uList:
        remove users from *user-item-matrix* based on (selection criteria)
        U, sigma, Vt $\leftarrow$ Singular Value Decomposition (user-item-matrix, k)
        all-user-predicted-ratings $\leftarrow$ ((U $\cdot$ sigma) $\cdot$ Vt)
        $pred(u,i) \leftarrow$ predicted rating of item i by user u
        predList $\leftarrow$ **aggregate** all pred(n, i)
        decreasingly **sort** predList by rating value
        recommend items from *predList* to user n

procedure CFD-WMT
    $selectionCriteria \leftarrow$ *wm(u)* < 5
    call Recommendation algorithm (*number of factors*)

6.4 Results

Impact on achieved utility

The effectiveness of the three proposed recommender algorithms is presented in Table 6.1. **All comparisons of average values in the table are statistically significant with p-values<0.01** (we used the Mann-Whitney U-test with sample sizes>10000).

The decision-competency-based algorithm clearly outperforms other algorithms for the right-skewed and symmetric product utility distributions. When compared to the baseline approach (product filtering without recommendations), the use of this recommendation algorithm results in a difference of up to 24 p.p. (for the right-skewed product utility distribution) or 15 p.p. (for the symmetric distribution) in the utility of older agents. Notably, the decision-competency-based algorithm also improves results achieved by younger adults, although by a smaller margin (15 p.p. and 9 p.p for the right-skewed and symmetric product utilities, respectively).

The CFD-WMT algorithm provided the best overall results for the left-skewed product utility distribution. It increased the utility obtained by older users by 6 p.p. and by younger users by 3 p.p. The Preference Twin algorithm had a marginal impact on the utility achieved by younger agents but improved the average utility obtained by older agents by 3 p.p. The second proposed algorithm, which considers not only preferences and age but also decision-making competency, increased the utility achieved by younger users by 3 p.p., which is comparable to the best collaborative filtering algorithm. For older users, the new algorithm was more effective than the CF algorithm and increased the utility obtained by older users by 4 p.p.

It should be noted that the proposed algorithms also led to improved results for younger users when compared to the baseline scenario. All three algorithms, namely Preferences Twin, Decision-competency based, and CFD-WMT, outperformed the benchmarks for younger agents. This is primarily due to the cognitive ability diversity, such as working memory, that exists among the younger population. Additionally, a comparison to traditional recommender systems can be found in Table 6.2 reveals, that all three proposed algorithms outperform the CB and CF algorithms. Both young and older users derive benefits from the new design of recommender systems that considers the differences in users' cognitive abilities.

Age, metric	Baseline	Preference twins	Decision competency-based	CFD-WMT
Symmetric distribution of product utility				
Old, absolute	0.70	0.80	**0.85**	0.83
Old, relative*	NA	0.10	**0.15**	0.13
Young, absolute	0.79	0.87	**0.89**	0.86
Young, relative*	NA	0.08	**0.09**	0.06
Left-skewed distribution of product utility				
Old, absolute	0.82	0.85	0.86	**0.88**
Old, relative*	NA	0.03	0.04	**0.06**
Young, absolute	0.88	0.89	**0.93**	**0.93**
Young, relative*	NA	0.00	**0.03**	**0.03**
Right-skewed distribution of product utility				
Old, absolute	0.45	0.67	**0.69**	0.63
Old, relative*	NA	0.22	**0.24**	0.18
Young, absolute	0.63	0.77	**0.78**	0.72
Young, relative*	NA	0.14	**0.15**	0.10

Table 6.1: The average utility of a purchased product in a simulation with a recommender system, using a left-skewed, right-skewed and symmetric distribution of product utility *difference between the utility obtained using the recommender system and the utility without the system.
All differences are statistically significant with p-values <0.01. The Mann-Whitney U test was used with sample sizes >10000.

Reduction of Self-Induced Bias

With respect to the impact of the proposed algorithms on the self-induced bias, Table 6.3 shows that all three proposed algorithms were effective in lowering the difference in utility achieved by older and younger users for symmetric and right-skewed distribution of product utility. The most effective algorithm was CFD-WMT for symmetric distribution, reducing the difference by 6 p.p. For right-skewed distribution, the results obtained by Decision competency-based and by CFD-WMT were both at the

Age, metric	CB	CF	Preference twins	Decision competency-based	CFD-WMT
Symmetric distribution of product utility					
Old	0.73	0.77	0.80	**0.85**	0.83
Young	0.80	0.86	0.87	**0.89**	0.86
Left-skewed distribution of product utility					
Old	0.82	0.84	0.85	0.86	**0.88**
Young	0.89	0.91	0.89	**0.93**	**0.93**
Right-skewed distribution of product utility					
Old	0.54	0.60	0.67	**0.69**	0.63
Young	0.68	0.72	0.77	**0.78**	0.72

Table 6.2: The average absolute utility of a purchased product in a simulation with a recommender system, using a left-skewed, right-skewed, and symmetric distribution of product utility, comparison of all recommender systems.

same level of 9 p.p. On the other hand, Preference Twins was slightly less effective with an 8 p.p. improvement over the benchmark.

Product preference distribution, metric	Baseline	CB	CF	Preference twins	Decision competency-based	CFD-WMT
Symmetric distribution of product utility						
Absolute*	0.09	0.07	0.09	0.07	0.04	**0.03**
Rec-sys**	NA	-0.02	0.00	-0.02	-0.05	**-0.06**
Left-skewed distribution of product utility						
Absolute*	0.06	0.07	0.07	**0.04**	0.07	0.05
Rec-sys**	NA	0.01	0.01	**-0.02**	0.01	-0.01
Right-skewed distribution of product utility						
Absolute*	0.18	0.14	0.12	0.1	**0.09**	**0.09**
Rec-sys**	NA	-0.04	-0.06	-0.08	**-0.09**	**-0.09**

Table 6.3: *Self-Induced Bias: Difference between the absolute performance of young and old agents

**Self-Induced Bias Caused by the Recommender system: difference vs baseline

CHAPTER 7

Conclusions

7.1 Summary of Contributions

This thesis addresses the issue of adapting e-commerce recommender systems to accommodate cognitive aging. The primary contribution of this study is the verification of Hypothesis 1, which suggests that due to the cognitive constraints of older customers, their product choices in e-commerce are less optimal than those of younger customers. This hypothesis was validated through both an experiment involving actual participants and other research methods, described in Section3.3 as well as by computer simulations, as described in Section 4.3.

Another significant contribution of this thesis is the identification of the issue of self-induced bias in recommender systems. This unique form of data bias arises when users make suboptimal choices, leading to the teaching of recommendation algorithms using data consisting of suboptimal decisions. To address this problem, I proposed a simulation-based approach for measuring the self-induced bias of a recommendation algorithm. This measurement approach entails a simulation model imbibing the utility of a simulated agent. The self-induced bias of a recommendation algorithm refers to the difference in average utility between two groups of agents, specifically the biased class (older consumers) and the normal class (younger consumers) when the recommendation algorithm is employed by all agents in the simulation.

The quest to quantify and explore self-induced bias has resulted in the creation and advancement of a practical simulator that models consumer decision-making in e-commerce systems. This simulator incorporates pertinent insights from psychology and marketing research and is made publicly available under an open-source license, representing another significant contribution of this thesis.

The second hypothesis considered in this thesis was that recommendation algorithms trained on suboptimal can further worsen the users' decision quality. The results of the verification of this hypothesis are inconclusive. In the first round of the simulated experiments, the recommendation algorithms were demonstrated to worsen some users' decision quality causing a vicious loop of wrong decisions. The experiments presented in the thesis do not demonstrate such phenomena for the aggregated older population, although it was present in individual cases.

This thesis also examined the hypothesis that traditional recommender systems are susceptible to self-induced bias among older users. Our simulation study confirmed this hypothesis, showing that Content-Based and Collaborative Filtering algorithms, even in their optimal settings, exhibit higher levels of self-induced bias compared to the newly proposed algorithm

Finally, this thesis proposes new algorithms aimed at decreasing self-induced bias. The algorithms are designed based on the hypothesis that decision quality can be improved if the system considers the cognitive limitations of some users. The hypothesis is supported by evidence that recommender systems, which remove lower-quality decisions from the training set and consider users' cognitive limitations, provide better recommendations for both younger and older users. Additionally, these new algorithms exhibit less self-induced bias among older users compared to algorithms that don't account for cognitive limitations.

7.2 Limitations

7.2.1 Aging Agent Model Limitations

The results described in this thesis are based on our simulation model. This model denotes an improved version of the model described in [8]. Unlike the model's previous version, we have removed several unrealistic assumptions and limitations. These include the assumption that agents optimize an objective function - this has been replaced by the use of realistic decision strategies studied in empirical psychological research. All model parameters are derived either from our own experiments or from the most relevant psychological or marketing research. However, no matter how realistic the model is, a computer simulation would not be able to fully replicate real human behavior.

The population of agents and items was sourced from real-life experiment participants and e-commerce platforms. However, these datasets may not be able to reflect other possible markets (with different products or a different customer population). We have attempted to mitigate this limitation by performing a sensitivity analysis of our results to the shape of the product utility distribution.

7.2.2 Recommender System Implementation Limitations

We have tested multiple representative versions (calibrations) of the mainstream recommender systems. Nevertheless, our tested versions of the Content-Based or Collaborative Filtering algorithms still may not represent the vast universe of all ways in which the CF and CB systems can be designed. Therefore, the results obtained from our versions may not be universal for all such recommendation systems.

Moreover, instead of real-world datasets as input for training the recommender systems, I was able to avoid the typical data sparsity issues experienced in reality due to using the results of simulated traffic on an e-commerce platform. In practise, the implementation of recommender systems must address the issues of data sparsity, cold-start, scalability, as well as other challenges.

7.2.3 Barriers in implementing the proposed algorithms in commercial setting

The algorithms proposed in this thesis are predicated on the underlying knowledge about user preferences or user cognitive abilities. In practice, this information may be difficult to obtain. As such, the algorithms proposed in this thesis may be viewed as concepts for future recommendation system development. Realistic algorithms could obtain user preferences using preference elicitation methods [117], and could estimate user cognitive abilities using psychological tests, such as working memory tests [109].

7.2.4 Barriers in measuring self-induced bias in commercial setting

Quantification of the self-induced bias proposed in the thesis requires measurement of utility achieved by the users. While it is straightforward in simulated models, where the preferences of the user are known, as well as to what degree users' needs are fulfilled with the chosen items, in real life the objective evaluation of the value provided to the customer is a complex issue requiring taking into account a variety of factors. Some techniques for measuring recommender system users' satisfaction were presented in [119], [120], and [121].

7.3 Future Work

In the thesis, a potential path for future work based on the research elucidated is to address the limitations regarding the implementation of the proposed algorithms in a commercial setting. More incisive research concerning extracting user preferences and assessing user decision-making competency would make the proposed algorithms more useful in a business environment.

Sharing the implementation of Aging Agent Model in an easy-to-use and interactive environment is another potentially beneficial initiative. One of the ways of achieving this is by integrating them with one of the existing platforms mentioned in Section 2.1.5.

References

[1] M. Steverson, World health organisation, ageing and health.
URL https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

[2] Pew-Research, Internet/Broadband Fact Sheet, Pew Research Center, 2021.
URL https://www.pewresearch.org/internet/fact-sheet/internet-broadband/

[3] S. Fedushko, T. Ustyianovych, E-commerce customers behavior research using cohort analysis: A case study of covid-19, Journal of Open Innovation: Technology, Market, and Complexity 8 (1) (2022) 12.

[4] D. Kahneman, Thinking, fast and slow, Macmillan, 2011.

[5] J. Pawlowska, K. Rydzewska, A. Wierzbicki, Using cognitive models to understand and counteract the effect of self-induced bias on recommendation algorithms, Journal of Artificial Intelligence and Soft Computing Research 13 (2) (2023).

[6] K. Rydzewska, J. Pawłowska, R. Nielek, A. Wierzbicki, G. Sedek, Cognitive limitations of older e-commerce customers in product comparison tasks, in: IFIP Conference on Human-Computer Interaction, Springer, 2021, pp. 646–656.

[7] R. Nielek, J. Pawlowska, K. Rydzewska, A. Wierzbicki, Adapting algorithms on the web to deal with cognitive aging, Multiple Pathways of Cognitive Aging: Motivational and Contextual Influences (2021) 368.

[8] J. Pawlowska, R. Nielek, A. Wierzbicki, Lost in online stores? agent-based modeling of cognitive limitations of elderly online consumers, in: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, Springer, 2019, pp. 204–213.

[9] P. Lops, M. De Gemmis, G. Semeraro, Content-based recommender systems: State of the art and trends, Recommender systems handbook (2011) 73–105.

[10] C. Desrosiers, G. Karypis, A comprehensive survey of neighborhood-based recommendation methods, Recommender systems handbook (2011) 107–144.

[11] G. Adomavicius, A. 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) (2005) 734–749.

[12] G. Adomavicius, R. Sankaranarayanan, S. Sen, A. Tuzhilin, Incorporating contextual information in recommender systems using a multidimensional approach, ACM Transactions on Information systems (TOIS) 23 (1) (2005) 103–145.

[13] R. Burke, Hybrid recommender systems: Survey and experiments, User modeling and user-adapted interaction 12 (4) (2002) 331–370.

[14] L. Sharma, A. Gera, A survey of recommendation system: Research challenges, International Journal of Engineering Trends and Technology (IJETT) 4 (5) (2013) 1989–1992.

[15] S. Sali, Movie rating prediction using singular value decomposition, Machine Learning Project Report by University of California, Santa Cruz (2008).

[16] T. Hofmann, Probabilistic latent semantic indexing, in: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, 1999, pp. 50–57.

[17] A. Mnih, R. R. Salakhutdinov, Probabilistic matrix factorization, Advances in neural information processing systems 20 (2007).

[18] M. Kunaver, T. Požrl, Diversity in recommender systems–a survey, Knowledge-based systems 123 (2017) 154–162.

[19] E. Lex, D. Kowald, P. Seitlinger, T. N. T. Tran, A. Felfernig, M. Schedl, et al., Psychology-informed recommender systems, Foundations and Trends® in Information Retrieval 15 (2) (2021) 134–242.

[20] E. Rich, User modeling via stereotypes, Cognitive science 3 (4) (1979) 329–354.

[21] N. A. ALRossais, D. Kudenko, Evaluating stereotype and non-stereotype recommender systems., in: KaRS@ RecSys, 2018, pp. 23–28.

[22] M. F. Rutledge-Taylor, A. Vellino, R. L. West, A holographic associative memory recommender system, in: 2008 Third International Conference on Digital Information Management, IEEE, 2008, pp. 87–92.

[23] D. Bollen, M. Graus, M. C. Willemsen, Remembering the stars? effect of time on preference retrieval from memory, in: Proceedings of the sixth ACM conference on Recommender systems, 2012, pp. 217–220.

[24] H. Ebbinghaus, Memory: A contribution to experimental psychology, Annals of neurosciences 20 (4) (2013) 155.

[25] H. Yu, Z. Li, A collaborative filtering method based on the forgetting curve, in: 2010 International conference on web information systems and mining, Vol. 1, IEEE, 2010, pp. 183–187.

[26] L. Ren, A time-enhanced collaborative filtering approach, in: 2015 4th International Conference on Next Generation Computer and Information Technology (NGCIT), IEEE, 2015, pp. 7–10.

[27] A. Chmiel, E. Schubert, Using psychological principles of memory storage and preference to improve music recommendation systems, Leonardo Music Journal 28 (2018) 77–81.

[28] Z. Yang, J. He, S. He, A collaborative filtering method based on forgetting theory and neural item embedding, in: 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), IEEE, 2019, pp. 1606–1610.

[29] J. R. Anderson, M. Matessa, C. Lebiere, Act-r: A theory of higher level cognition and its relation to visual attention, Human–Computer Interaction 12 (4) (1997) 439–462.

[30] L. Van Maanen, J. N. Marewski, Recommender systems for literature selection: A competition between decision making and memory models, in: Proceedings of the 31st Annual Conference of the Cognitive Science Society, Austin, TX: Cognitive Science Society, 2009, pp. 2914–2919.

[31] D. Kowald, P. Seitlinger, C. Trattner, T. Ley, Long time no see: The probability of reusing tags as a function of frequency and recency, in: Proceedings of the 23rd International Conference on World Wide Web, 2014, pp. 463–468.

[32] C. Trattner, D. Kowald, P. Seitlinger, S. Kopeinik, T. Ley, Modeling activation processes in human memory to predict the reuse of tags, The Journal of Web Science 2 (2016).

[33] D. Kowald, P. Seitlinger, S. Kopeinik, T. Ley, C. Trattner, Forgetting the words but remembering the meaning: Modeling forgetting in a verbal and semantic

tag recommender, in: Mining, Modeling, and Recommending'Things' in Social Media, Springer, 2013, pp. 75–95.

[34] D. Kowald, E. Lex, The influence of frequency, recency and semantic context on the reuse of tags in social tagging systems, in: Proceedings of the 27th ACM Conference on Hypertext and Social Media, 2016, pp. 237–242.

[35] C. Stanley, M. D. Byrne, Comparing vector-based and bayesian memory models using large-scale datasets: User-generated hashtag and tag prediction on twitter and stack overflow., Psychological Methods 21 (4) (2016) 542.

[36] M. C. Mozer, R. V. Lindsey, Predicting and improving memory retention: Psychological theory matters in the big data era, in: Big data in cognitive science, Psychology Press, 2016, pp. 43–73.

[37] L. Li, W. Chu, J. Langford, R. E. Schapire, A contextual-bandit approach to personalized news article recommendation, in: Proceedings of the 19th international conference on World wide web, 2010, pp. 661–670.

[38] J.-C. Shi, Y. Yu, Q. Da, S.-Y. Chen, A.-X. Zeng, Virtual-taobao: Virtualizing real-world online retail environment for reinforcement learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 4902–4909.

[39] D. Rohde, S. Bonner, T. Dunlop, F. Vasile, A. Karatzoglou, Recogym: A reinforcement learning environment for the problem of product recommendation in online advertising, arXiv preprint arXiv:1808.00720 (2018).

[40] E. Ie, C.-w. Hsu, M. Mladenov, V. Jain, S. Narvekar, J. Wang, R. Wu, C. Boutilier, Recsim: A configurable simulation platform for recommender systems, arXiv preprint arXiv:1909.04847 (2019).

[41] M. R. Santana, L. C. Melo, F. H. Camargo, B. Brandão, A. Soares, R. M. Oliveira, S. Caetano, Mars-gym: A gym framework to model, train, and evaluate recommender systems for marketplaces, in: 2020 International Conference on Data Mining Workshops (ICDMW), IEEE, 2020, pp. 189–197.

[42] B. Shi, M. G. Ozsoy, N. Hurley, B. Smyth, E. Z. Tragos, J. Geraci, A. Lawlor, Pyrecgym: A reinforcement learning gym for recommender systems, in: Proceedings of the 13th ACM Conference on Recommender Systems, 2019, pp. 491–495.

[43] L. Bernardi, S. Batra, C. A. Bruscantini, Simulations in recommender systems: An industry perspective, arXiv preprint arXiv:2109.06723 (2021).

[44] J. Huang, H. Oosterhuis, M. De Rijke, H. Van Hoof, Keeping dataset biases out of the simulation: A debiased simulator for reinforcement learning based recommender systems, in: Fourteenth ACM conference on recommender systems, 2020, pp. 190–199.

[45] R. Cabeza, Hemispheric asymmetry reduction in older adults: the harold model., Psychology and aging 17 (1) (2002) 85.

[46] D. Kahneman, Attention and effort, Vol. 1063, Citeseer, 1973.

[47] J. Cerella, Age-related decline in extrafoveal letter perception, Journal of Gerontology 40 (6) (1985) 727–736.

[48] T. A. Salthouse, R. L. Babcock, Decomposing adult age differences in working memory., Developmental psychology 27 (5) (1991) 763.

[49] E. L. Glisky, Changes in cognitive function in human aging, Brain aging (2007) 3–20.

[50] P. A. Reuter-Lorenz, C.-Y. C. Sylvester, The cognitive neuroscience of working memory and aging. (2005).

[51] W. R. Miller, M. E. Seligman, Depression and learned helplessness in man., Journal of abnormal psychology 84 (3) (1975) 228.

[52] E. C. Nisbet, O. Kamenchuk, Russian news media, digital media, informational learned helplessness, and belief in covid-19 misinformation, International Journal of Public Opinion Research 33 (3) (2021) 571–590.

[53] G. Sedek, D. N. McIntosh, Intellectual helplessness, in: Personal control in action, Springer, 1998, pp. 419–443.

[54] G. Sedek, M. Kofta, T. Tyszka, Effects of uncontrollability on subsequent decision making: Testing the cognitive exhaustion hypothesis., Journal of Personality and Social Psychology 65 (6) (1993) 1270.

[55] U. Von Hecker, G. Sedek, Uncontrollability, depression, and the construction of mental models., Journal of Personality and Social Psychology 77 (4) (1999) 833.

[56] W. Bruine de Bruin, A. M. Parker, B. Fischhoff, Decision-making competence: More than intelligence?, Curr. Directions in Psych. Science 29 (2) (2020) 186–192.

[57] T. M. Hess, Subjective perceptions of cognitive costs: Determinants and impact on motivation and engagement. (2022).

[58] T. M. Hess, Selective engagement of cognitive resources: Motivational influences on older adults' cognitive functioning, Persp. Psychol. Science 9 (4) (2014) 388–407.

[59] J. M. Morris, User interface design for older adults, Interacting with Computers 6 (4) (1994) 373–393.

[60] R. W. Morrell, Human factors guidelines for web site usability: Health-oriented web sites for older adults, in: Older adults, health information, and the World Wide Web, Psychology Press, 2001, pp. 99–118.

[61] F. Nunes, M. Kerwin, P. A. Silva, Design recommendations for tv user interfaces for older adults: findings from the ecaalyx project, in: Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility, 2012, pp. 41–48.

[62] S. Liu, S. Joines, Developing a framework of guiding interface design for older adults, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 56, SAGE Publications Sage CA: Los Angeles, CA, 2012, pp. 1967–1971.

[63] B. B. Neves, F. Vetere, Ageing and digital technology: designing and evaluating emerging technologies for older adults, Springer, 2019.

[64] L. Nurgalieva, J. J. J. Laconich, M. Baez, F. Casati, M. Marchese, A systematic literature review of research-derived touchscreen design guidelines for older adults, IEEE Access 7 (2019) 22035–22058.

[65] B. Knowles, V. L. Hanson, Y. Rogers, A. M. Piper, J. Waycott, N. Davies, Hci and aging: Beyond accessibility, in: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1–8.

[66] A. Wierzbicki, Trust and fairness in open, distributed systems, Vol. 298, Springer, 2010.

[67] P. Gajane, M. Pechenizkiy, On formalizing fairness in prediction with machine learning, arXiv preprint arXiv:1710.03184 (2017).

[68] M. J. Kusner, J. Loftus, C. Russell, R. Silva, Counterfactual fairness, Advances in neural information processing systems 30 (2017).

[69] N. Grgic-Hlaca, M. B. Zafar, K. P. Gummadi, A. Weller, The case for process fairness in learning: Feature selection for fair decision making, in: NIPS symposium on machine learning and the law, Vol. 1, Barcelona, Spain, 2016, p. 2.

[70] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, R. Zemel, Fairness through awareness, in: Proceedings of the 3rd innovations in theoretical computer science conference, 2012, pp. 214–226.

[71] M. B. Zafar, I. Valera, M. G. Rogriguez, K. P. Gummadi, Fairness constraints: Mechanisms for fair classification, in: Artificial intelligence and statistics, PMLR, 2017, pp. 962–970.

[72] M. Hardt, E. Price, N. Srebro, et al., Equality of opportunity in supervised learning, in 'advances in neural information processing systems' (2016).

[73] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, X. He, Bias and debias in recommender system: A survey and future directions, arXiv preprint arXiv:2010.03240 (2020).

[74] A. Olteanu, C. Castillo, F. Diaz, E. Kıcıman, Social data: Biases, methodological pitfalls, and ethical boundaries, Frontiers in Big Data 2 (2019) 13.

[75] M. D. Ekstrand, M. Tian, I. M. Azpiazu, J. D. Ekstrand, O. Anuyah, D. McNeill, M. S. Pera, All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness, in: Conference on fairness, accountability and transparency, PMLR, 2018, pp. 172–186.

[76] H. A. Simon, Models of bounded rationality: Empirically grounded economic reason, Vol. 3, MIT press, 1997.

[77] S. S. Iyengar, M. R. Lepper, When choice is demotivating: Can one desire too much of a good thing?, Journal of personality and social psychology 79 (6) (2000) 995.

[78] D. Kahneman, S. P. Slovic, P. Slovic, A. Tversky, Judgment under uncertainty: Heuristics and biases, Cambridge university press, 1982.

[79] R. H. Guttman, A. G. Moukas, P. Maes, Agent-mediated electronic commerce: A survey, The Knowledge Engineering Review 13 (2) (1998) 147–159.

[80] Y. Wan, S. Menon, A. Ramaprasad, The paradoxical nature of electronic decision aids on comparison-shopping: The experiments and analysis, Journal of theoretical and applied electronic commerce research 4 (3) (2009) 80–96.

[81] G. Häubl, V. Trifts, Consumer decision making in online shopping environments: The effects of interactive decision aids, Marketing science 19 (1) (2000) 4–21.

[82] Y. Park, U. Gretzel, Influence of consumers' online decision-making style on comparison shopping proneness and perceived usefulness of comparison shopping tools (2010).

[83] A. J. Rohm, V. Swaminathan, A typology of online shoppers based on shopping motivations, Journal of business research 57 (7) (2004) 748–757.

[84] N. N. G. Kate Moran, Comparison tables for products, services, and features (2017).
URL `https://www.nngroup.com/articles/comparison-tables/`

[85] C. Institute, Ecommerce guidelines: Product comparison (2018).
URL `https://cxl.com/ecommerce-best-practices/product-comparison/`

[86] C. Michelle Lee, Comparing products: Design practices to help your users avoid fragmented comparison (2021).
URL `https://contentsquare.com/blog/comparing-products-design-practices-to-help-your-users-avoid-fragmented-comparison-7/`

[87] b. Edward Scott, 3 product comparison ux implementation details for designing user-friendly 'compare' tools (2019).
URL `https://baymard.com/blog/user-friendly-comparison-tools`

[88] S. M. Vitaly Friedman, Designing the perfect feature comparison table (2017).
URL `https://www.smashingmagazine.com/2017/08/designing-perfect-feature-comparison-table/`

[89] M. Bohanec, V. Rajkovic, Knowledge acquisition and explanation for multi-attribute decision making, in: 8th Intl Workshop on Expert Systems and their Applications, Citeseer, 1988, pp. 59–78.

[90] R. Mata, L. J. Schooler, J. Rieskamp, The aging decision maker: cognitive aging and the adaptive selection of decision strategies., Psych. and aging 22 (4) (2007) 796.

[91] R. e. a. Mata, Learning to choose: Cognitive aging and strategy selection learning in decision making., Psych. and aging 25 (2) (2010) 299.

[92] T. M. Hess, T. L. Queen, T. R. Patterson, To deliberate or not to deliberate: Interactions between age, task characteristics, and cognitive activity on decision making, Journal of Behavioral Decision Making 25 (1) (2012) 29–40.

[93] G. e. a. Chasseigne, Aging and probabilistic learning in single-and multiple-cue tasks, Experimental Aging Research 30 (1) (2004) 23–45.

[94] G. R. Samanez-Larkin, S. E. Gibbs, K. Khanna, L. Nielsen, L. L. Carstensen, B. Knutson, Anticipation of monetary gain but not loss in healthy older adults, Nature neuroscience 10 (6) (2007) 787–791.

[95] V. Pareto, Cours d'économie politique: professé à l'Université de Lausanne, Vol. 1, F. Rouge, 1896.

[96] G. Gigerenzer, D. G. Goldstein, Reasoning the fast and frugal way: models of bounded rationality., Psychological review 103 (4) (1996) 650.

[97] C.-L. Hwang, A. S. M. Masud, Multiple objective decision making—methods and applications: a state-of-the-art survey, Vol. 164, Springer Science & Business Media, 2012.

[98] J. J. Bernardo, J.-M. Blin, A programming model of consumer choice among multi-attributed brands, Journal of consumer research 4 (2) (1977) 111–118.

[99] Y. Haimes, On a bicriterion formulation of the problems of integrated system identification and system optimization, IEEE transactions on systems, man, and cybernetics 1 (3) (1971) 296–297.

[100] P. C. Fishburn, Exceptional paper—lexicographic orders, utilities and decision rules: A survey, Management science 20 (11) (1974) 1442–1471.

[101] N. Gunantara, A review of multi-objective optimization: Methods and its applications, Cogent Engineering 5 (1) (2018) 1502242.

[102] A. P. Wierzbicki, Reference point approaches, in: Multicriteria decision making, Springer, 1999, pp. 237–275.

[103] L. Thu Bui, S. Alam, Multi-Objective Optimization in Computational Intelligence: Theory and Practice: Theory and Practice, IGI global, 2008.

[104] A. Fagerlin, B. J. Zikmund-Fisher, P. A. Ubel, A. Jankovic, H. A. Derry, D. M. Smith, Measuring numeracy without a math test: development of the subjective numeracy scale, Medical Decision Making 27 (5) (2007) 672–680.

[105] I. M. Lipkus, G. Samsa, B. K. Rimer, General performance on a numeracy scale among highly educated samples, Medical decision making 21 (1) (2001) 37–44.

[106] E. Peters, D. Västfjäll, P. Slovic, C. Mertz, K. Mazzocco, S. Dickert, Numeracy and decision making, Psychological science 17 (5) (2006) 407–413.

[107] A. F. Hayes, Introduction to mediation, moderation, and conditional process analysis: A regression-based approach, Guilford publications, 2017.

[108] J. G. Bullock, D. P. Green, S. E. Ha, Yes, but what's the mechanism?(don't expect an easy answer)., Journal of personality and social psychology 98 (4) (2010) 550.

[109] R. H. Logie, E. A. Maylor, An internet study of prospective memory across adulthood., Psychology and aging 24 (3) (2009) 767.

[110] A. W. Kruglanski, D. M. Webster, Motivated closing of the mind:" seizing" and" freezing.", Psychological review 103 (2) (1996) 263.

[111] C. Dimov, P. H. Khader, J. N. Marewski, T. Pachur, How to model the neurocognitive dynamics of decision making: A methodological primer with act-r, Behavior research methods 52 (2) (2020) 857–880.

[112] B. Von Helversen, K. Abramczuk, W. Kopeć, R. Nielek, Influence of consumer reviews on online purchasing decisions in older and younger adults, Decision Support Systems 113 (2018) 1–10.

[113] R. Lambert-Pandraud, G. Laurent, E. Lapersonne, Repeat purchasing of new automobiles by older consumers: empirical evidence and interpretations, Journal of Marketing 69 (2) (2005) 97–113.

[114] J. R. Hauser, Consideration-set heuristics, Journal of Business Research 67 (8) (2014) 1688–1699.

[115] J. R. Hauser, O. Toubia, T. Evgeniou, R. Befurt, D. Dzyabura, Disjunctions of conjunctions, cognitive simplicity, and consideration sets, Journal of Marketing Research 47 (3) (2010) 485–496.

[116] M. Ding, J. R. Hauser, S. Dong, D. Dzyabura, Z. Yang, S. Chenting, S. P. Gaskin, Unstructured direct elicitation of decision rules, Journal of Marketing Research 48 (1) (2011) 116–127.

[117] L. Chen, P. Pu, Survey of preference elicitation methods, Tech. rep. (2004).

[118] R. B. Nozari, H. Koohi, Novel implicit-trust-network-based recommendation methodology, Expert Systems with Applications 186 (2021) 115709.

[119] P. Cremonesi, F. Garzotto, S. Negro, A. V. Papadopoulos, R. Turrin, Looking for "good" recommendations: A comparative evaluation of recommender systems, in: Human-Computer Interaction–INTERACT 2011: 13th IFIP TC 13 International Conference, Lisbon, Portugal, September 5-9, 2011, Proceedings, Part III 13, Springer, 2011, pp. 152–168.

[120] T. Silveira, M. Zhang, X. Lin, Y. Liu, S. Ma, How good your recommender system is? a survey on evaluations in recommendation, International Journal of Machine Learning and Cybernetics 10 (2019) 813–831.

[121] B. P. Knijnenburg, M. C. Willemsen, Understanding the effect of adaptive preference elicitation methods on user satisfaction of a recommender system, in: Proceedings of the third ACM conference on Recommender systems, 2009, pp. 381–384.