Examining the Effects of Personalized Explanations in a Multi-list Food Recommender System

Lars Giske Holth

Supervisor: Prof. Dr Christoph Trattner

Co-Supervisor: Dr Alain D. Starke

Department of Information Science and Media Studies

University of Bergen

May 31, 2022

Acknowledgements

I would like to thank my supervisor Prof. Dr. Christoph Trattner, for helpful advice and feedback on the prototype and thesis. In addition, I would like to express my appreciation to my co-supervisor, Dr. Alain D. Starke, for continuously providing me with valuable feedback, scientific guidance, and expertise in statistics.

Thank you to my family and my lovely girlfriend Magdele for their support throughout the year. I would also like to thank my fellow students in room 644, Hanna, Hedda, Gina, and Nora, for their great input and support.

Lars G. Holth

Bergen, May 2022

Abstract

In the past decade, food recipe websites have become a popular approach to find a recipe. Due to the vast amount of options, food recommender systems have been developed and used to suggest appetizing recipes. However, recommending appealing meals does not necessarily imply that they are healthy. Recent studies on recommender systems have demonstrated a growing interest in altering the interface, where the usage of multi-list interfaces with explanations has been explored earlier in an unsuccessful attempt to encourage healthier food choices. Building upon other research that highlights the ability of personalized explanations to provide a better understanding of presented recommendations, this thesis explores whether a multi-list interface with personalized explanations, which takes into account user preferences, health, and nutritional aspects, can affect users' evaluation and perception of a food recommender system, as well as steer them towards healthier choices. A food recommender system was develop, with which single- and multi-lists, as well as non-personalized and personalized explanations, were compared in an online experiment (N = 163) in which participants were requested to choose recipes they liked and to answer questionnaires. The analysis revealed that personalized explanations in a multi-list interface were not able to increase choice satisfaction, choice difficulty, understanding or support healthier choices. Surprisingly, users selected healthier recipes if non-personalized rather than personalized explanations were presented alongside them. In addition, users perceived multi-lists to be more diverse and found single-list to be more satisfying.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the years, food recipe websites containing a variety of options have become an increasingly popular source of inspiration for home cooks. A search on the internet will retrieve thousands of recipe websites that offer food recipes created by either beginners or chefs [11], in just a matter of seconds. All the different cuisines, cultures and combinations of ingredients provides new ways to prepare a meal which gives an increase in options for users looking for a recipe online. However, due to the large number of options it can be challenging for the average user to find their preferred recipe, as well as to make healthy choices. For many people, food selection is driven by bad habits or poor knowledge on healthy eating which could lead to poor healthy conditions [21]. The focus on healthy eating has become a more central subject in our daily life, as lifestyle-related illnesses such as obesity and diabetes have become a growing problem [66]. World Health Organization estimated in 2016 that 39% of adults aged 18 years and older, were overweight and 13% were obese. In a world where overweight and obesity kills more people than underweight, it is important to promote healthy food options [66].

The constant growing array of recipe choices that consumers on the Internet are going through can be perceived as overwhelming. For guiding the consumers towards a recipe of their interest, a recommender system can be a helpful tool. Recommender systems consist of algorithms within a user interface that are designed to suggest relevant items to users. The relevant items could be identified based on users with similar preferences, items with similar features, or item features that are similar to the user's preferences. In recent years, recommender systems have been investigated as an effective solution to help users change their eating behavior and discover healthier recipes [60]. By suggesting a more personal and diversified food recipe selection that is healthy and may be of interest, one can possibly move them in a healthier direction. Currently, recommender systems are widely used across different online food recipe websites where the system serves its purpose to recommend new and similar recipes, but does not necessarily promote the healthiest options [60, 62]. Food recipe websites have a tendency to promote popular recipes, which in most cases are the ones that are the least healthy ones. [62, 63]. There are challenges associated with the need for information from the user in certain aspects such as nutritional needs, ratings of recipes and previous meals [60] to provide accurate and healthy recommendations. Consumers have a tendency to select recipes that satisfy their particular preferences, despite the fact that this may not correlate with the healthiness of the chosen meal and may be influenced by poor habits. There are also misconceptions out there about healthy eating which is associated with a reduced consumption of food that are usually defined as healthy [12].

Traditionally, recommender systems have been evaluated primarily on the accuracy of the recommendations. In recent years, however, interest in user-centric assessment metrics [6, 43, 49, 56] have grown where various components of the user interface, such as explanations and presentation of the recommendations [41, 48, 53, 58] have been shown to have a significant impact on the user and the quality of the recommender system. The organization of recommendations has, among other things, shown to alter the persuasiveness of a system and the satisfaction of its users [43]. When browsing through various online e-commerce websites and online streaming services, one can observe that their recommendations are presented to users in multiple lists. These lists are often optimized on different algorithms with an explanation to each list that says something about what the list contains and/or why it is relevant for the user. Having recommendations in multiple lists could possibly serve purposes such as guiding users towards new items of their interest that they were unaware of and lead to more efficient explorations among a bigger subset of items [28]. As opposed to multi-lists, single-list interfaces are also widely used on websites, especially in food recipe websites such as Allrecipes[1] (Figure 1.1), Simplyrecipes[2] and Yummly[3] to mention a few. Such lists can however just account for one factor at the time with a more limited selection. Researchers have studied users' decision-behavior across single and multi-lists when browsing movies [28], their perception of the lists and its ability to nudge users towards healthier food choices [51].

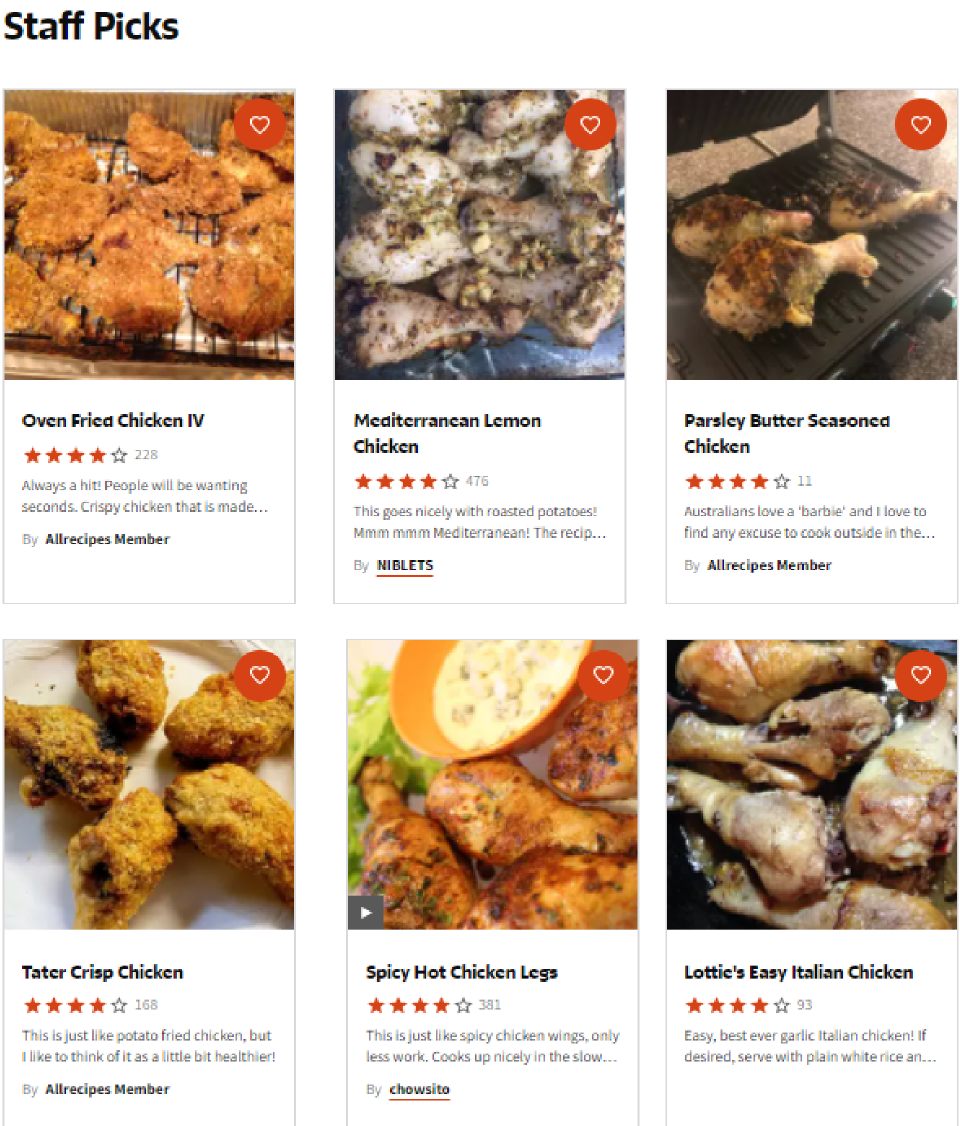

Figure 1.1: A partial screenshot of Allrecipes.com where the recipes are presented in a single-list with a single explanation.

[1] `www.allrecipes.com`
[2] `www.simplyrecipes.com`
[3] `www.yummly.com`

When consumers experience difficulty in deciding, explanations may benefit them to clarify why a recommendation was made, what qualities the recommended product(s) has, and why the recommendation is relevant to the user. Not only is a good explanation beneficial to the users, but they may also serve as an effective resource for the provider to persuade users in a certain direction, which both the receiver and the provider may use to improve their performance throughout the communication process [29]. A good explanation seeks to achieve seven possible aims among users, regardless of the domain, which in short is to inspire user trust and loyalty, boost satisfaction, make it simpler to identify what consumers want, and encourage them to test, buy or choose a recommended item [56]. However, the goals that the explanations want to achieve may vary based on the domain in which it is used [50]. A movie recommender system may emphasize helping users in making good decisions about which movie to watch, whereas an e-commerce website may prioritize persuading a user to purchase a specific product. Regarding users' demand for explanations, users want to know more about what makes a particular item worthy of being recommended [57], which has led to an emphasis on personalized explanations that combine user characteristics and item features [35, 53, 58]. In food recommender systems, where a one-size-fits-all approach is often used, recommendations aimed at personal taste and nutritional aspects can benefit consumers to follow a nutritional and healthy diet, especially when it is closely associated with personal characteristics [59]. Utilizing personal characteristics in the form of an explanation of why a recommendation was made and how it may be relevant to the user has proven to be highly persuasive [8, 24] in order to make users try or buy an item. However, little research has been conducted on how personalized explanations in the food domain, which include nutritional aspects, user preferences, and personal health, can encourage users to make healthier choices, as well as the effect of serving multiple personalized explanations at once on users' perceptions.

1.1 Problem Statement

The ever-growing selection of recipes in various food recipe websites can create challenges for users when choosing what to eat. A goal that recommender systems seek to achieve is to help users overcome these challenges by recommending similar or popular recipes. However, recommending for popularity or similarity might not always correlate with recipes being healthy [62, 63], which is becoming more and more important due to the increasing number of overweight in our society [66]. A field of research which has shown to affect users in their decision-making is to make changes to the presentation of recommendations. While a large number of food recipe websites use single-list interfaces to present recommendations (Allrecipes, Simplyrecipes and Yummly), several online streaming services (Netflix, HBO, Amazon Prime) present their recommendations in multiple lists. Until now, research has compared single- and multi-lists and examined their influence on users' decision-making in the movie domain [28], as well as the persuasiveness of multi-lists and explanations towards healthy food choices [51]. Users in the food domain were more satisfied with the choices they made while users in the movie domain found multi-list interfaces to provide a more diverse recommendations set due to being divided into several lists optimized for different algorithms. However, neither presenting food recommendations in a multi-list interface nor providing explanations were able to provide users with a higher understanding of the presented recommendations, reduce choice overload or assist them in making healthier choices [51].

Research on multi-lists within the food domain are still novel where no research have been conducted on how a personalized type of explanation could affect the users' perceptions of the multi-list interface and its recommendations, as well as its persuasive ability towards healthier food choices. This study aims to personalize the explanations in single and multi-lists by incorporating user preferences, nutritional aspects, and recipe features. As prior research has found personalized explanations to be highly persuasive [8, 24, 35] and to provide users with a better understanding of the content [58], the purpose of this thesis is to utilize personalized explanations to support users in choosing healthier recipes in multi-lists interfaces. Using personalized explanations in the food domain was done in a justification style by [41], but interesting results may be found when the personalized explanations account for multiple lists and recommendations. This thesis will also investigate how the use of personalized explanations in a multi-list interface can affect a user's choice satisfaction, choice difficulty, perceived diversity, and understanding, in addition to other recipe choice aspects. A comparison will be made between personalized and non-personalized explanations, as well as a multi-list interface and a single-list interface.

1.2 Research Questions

The goal of this thesis is to investigate the impact of personalized explanations in a multi-list food recommender system on users perception and evaluation, in addition to its ability to persuade users towards healthier recipes. In order to do so, the following research questions are raised:

- **RQ1.1:** To what extent is a users perception of a food recommender system affected by the use of personalized explanations and a multi-list interface?

- **RQ1.2:** To what extent do personalized explanations affect choice satisfaction and difficulty across multi-list and single list interfaces?

- **RQ2.1:** To what extent do personalized explanations and multi-list interfaces support healthier recipe choices?

- **RQ2.2:** Do personalized explanations and multi-list affect other recipe choice aspects, compared to non-personalized explanations and single list interfaces?

1.3 Thesis Outline

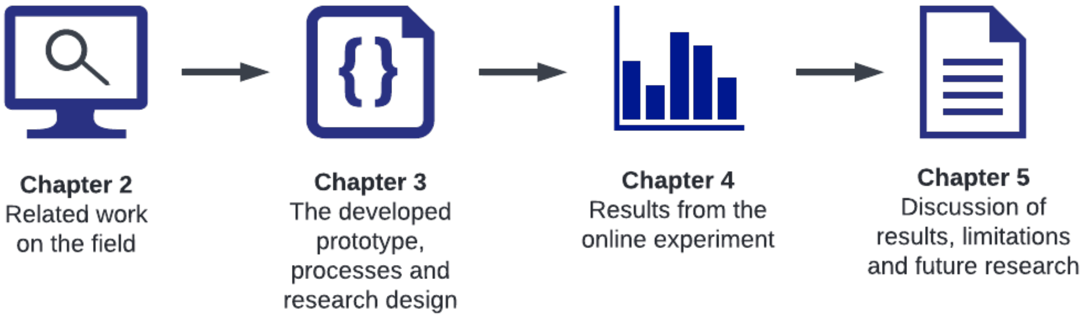

Figure 1.2: A brief overview over the remaining chapters in the thesis.

The four coming chapters are visualized in Figure 1.2 and are as follows:

- **Chapter 2** provides an overview over conducted research on the field of food recommender systems in addition to related works on explanations and multi-list interfaces.

- **Chapter 3** gives a full overview of the dataset used in the experiment, as well as technicalities and the approach used to develop the prototype. In addition, there are descriptions of the research design, experiment procedure, and participant recruitment.

- **Chapter 4** provides the results of the conducted online experiment in conjunction with the research questions.

- **Chapter 5** provides a discussion of the results, limitations and future research.

Chapter 2

Background

2.1 Recommender Systems

The primary goal of a recommender system is to provide items that are useful and/or of interest to the user [49] and is commonly utilized in various e-commerce websites, streaming services and social media where it serves different purposes. Such purposes may include increasing sales in e-commerce by recommending items that will most likely suit the user, increasing diversity in movie recommendations for streaming services or increasing users satisfaction in the form of relevant and interesting recommendations in social media [49]. Most of these systems are often personalized to the users needs where implicit data such as clicks, time spent or purchases, as well as explicit data such as ratings or likes are typically considered in the recommendation process [53]. To be able to identify and predict what the users would find interesting and suitable for their needs, several different approaches could be used. Collaborative filtering is considered as the most popular and used one amongst the techniques [7]. Its main goal is to exploit gathered data about a user's past interactions and/or about opinions from existing users in the community to provide recommendations that the current user of the system would be interested in [29]. In most cases, ratings are often used as an expression of preference whereas the calculations are based on similarity in rating history between two users [53]. MovieLens is an example of a collaborative filtering movie recommender which takes the users rating and ratings from the community to predict how that user will rate future movies of interest [49]. Content-based recommendations, on the other hand, are based on item-similarity rather than user similarity. This means that users will receive recommendations on items similar to the ones that they have liked in the past. The assumption is that items with similar features will be rated similarly [7] where the calculations are made on the feature similarity between the compared items [49]. A knowledge-based recommendation would, however, recommend items based on explicit knowledge of certain features, user needs and preferences, as well as knowledge on how the features meet the users needs [53]. The greatest strength of a knowledge-based recommender compared to a collaborative-filtering and content-based recommender is that it can overcome the cold start problem (the performance of the system on new items and on new users) [49]. However, a knowledge-based recommender may suffer from knowledge acquisition bottleneck, which means that knowledge engineers who can encode knowledge from domain experts into a formal and feasible representation are required [18]. Incorporating domain knowledge into a knowledge-based recommendation process was done by Khan and Hoffmann with a system called MIKAS, a diet recommender system for changing eating habits. The purpose of the system is to recommend a menu personalized to a user's requirements by modifying it along the way. In case of a recommendation that does not meet the requirements for a healthy diet, a domain expert (dietitian) could intervene and provide additional knowledge which would then be added to the knowledge-base to ensure its adaptation [32].

2.2 Food Recommender Systems

Food recommender systems aim to recommend new recipes in various ways based on users interest, similarity to chosen recipe, taste or certain dietary goals provided by the user. In the healthy food domain, recommender systems have proven to be effective in helping users to cope with the vast amount of data related to recipes [60]. By utilizing recommender system techniques such as collaborative filtering, content-based, knowledge-based or hybrid recommender systems, one can create personalized recommendations for the users. In recent years, research contributions have investigated methods to guide users towards healthier recipes with these techniques.

Research by Mika [40] states that there are two types of food recommender systems. One type is to recommend recipes by considering user preferences (type 1) while the other type is to recommend recipes by considering nutritional needs of users (type 2). Recommending recipes by considering user preferences can be done through content-based methods to tailor recommendations to the individual tastes. Research on personalized recipe recommendation by Freyne et al. [19] used a content-based algorithm to predict the rating of a target recipe by breaking the target recipe and ingredients. The predicted rating was calculated by the users rating of the individual ingredients in the target recipe and was tested against a collaborative filtering algorithm. Results have shown that the content-based algorithm turned out to provide more accurate recommendations than the collaborative one. Another research paper [16] focused on gathering user-preferences through an interaction process that elicits users short-term and long-term recipe preferences. Combining both types of preference elicitation methods yielded positive results among users.

Research by Trang et al. [60] also identified one more recommendation type to balance user preferences and nutritional needs of users (type 3). The Health-aware recommender system by Ge et al. [21] provides recipe recommendations that takes both users preferences and health into account. The system allows users to balance both their taste and health to receive personalized recommendations. Considering nutritional information and user preferences has also been researched by Toledo et al. [59]. A pre-filtering approach was used to eliminate food that does not fit the users characteristics to further generate a meal plan with food that satisfies a users preferences and their daily nutritional requirements. Another recommender system [3] takes both user preferences and medical prescriptions to provide personalized and healthy menus through three steps. A content-based filtering approach provides the user with recommended recipes based on comparisons among features in both users profiles and recipes. Additionally, candidate menus are generated using the selected recipes and finally ranked taking the users prescriptions into account.

Musto et al. [42] created a knowledge-aware recommender system that combines knowledge about food and personal user aspects such as nutritional content, cooking experience, physical activity, BMI etc. to generate personalized recipe suggestions. Users were served three dishes from the knowledge-aware recommender (main course, second course, dessert) which was compared to dishes from a popularity-based recommender. The same authors did a Natural Language Justification study [41] which were also based on user characteristics and recipe features. From examining what drives a user towards healthier recommendations, they found that the recipes chosen based on popularity were related to taste motivations while choices in the health-aware recommendation were related to health-related reasons. Furthermore, 'because it fits my preferences' was identified as a cause for selecting healthier recipes.

An interesting comparison between the three types of recommending recipes was done by [46] where each recommendation was either optimized for user preferences, healthiness or a combination of both. The users were asked to select their preferred recipe from a list in order to determine if a healthy bias and a healthy tag could influence users decision-making. Their research showed that users are generally more inclined to select recommendations from the list optimized for user preferences than the hybrid list, while both lists were more accurate than the list optimized for healthiness. However, the users who explicitly cared about healthiness were more likely to choose recipes from the list optimized for healthiness and the hybrid list.

2.3 Explanations in Recommender Systems

Throughout the existence of food recommender systems, its main purpose has been to suggest relevant recipes to users. However, from a user's perspective, it might be difficult to determine which recipes are healthy among a large selection of recipes. In addition, most recommender approaches are popularity-based which could lead to unhealthy recommendations. While recommender systems traditionally have been evaluated on its recommendation accuracy, recent years have shown an increasing interest in user-centered aspects [58]. One factor that can affect the user is the use of explanations along with the recommendation. The goal is to give the user some sort of context as to why an item was recommended and why it could be of interest for the user. N. Tintarev and J. Masthoff [56] has identified seven possible aims of explanations in recommender systems: Transparency (1) is explaining how the system works while helping users understand how a recommendation was chosen whereas scrutability (2) allows users to provide feedback. Trust (3) is about increasing users' confidence in the system while persuasiveness (4) is about convincing users to either try or buy a product. In addition to persuading users to either try or buy an item, an explanation may also assist users to make better decisions from its effectiveness (5) and faster decisions from its efficiency (6). Finally, an explanation may improve users overall satisfaction (7) with the system. Nowadays, explanations are frequently used in popular online movie and streaming services such as Netflix (Figure 2.4), Amazon Prime Video and HBO Max (Figure 2.1) in their recommendations to the users.

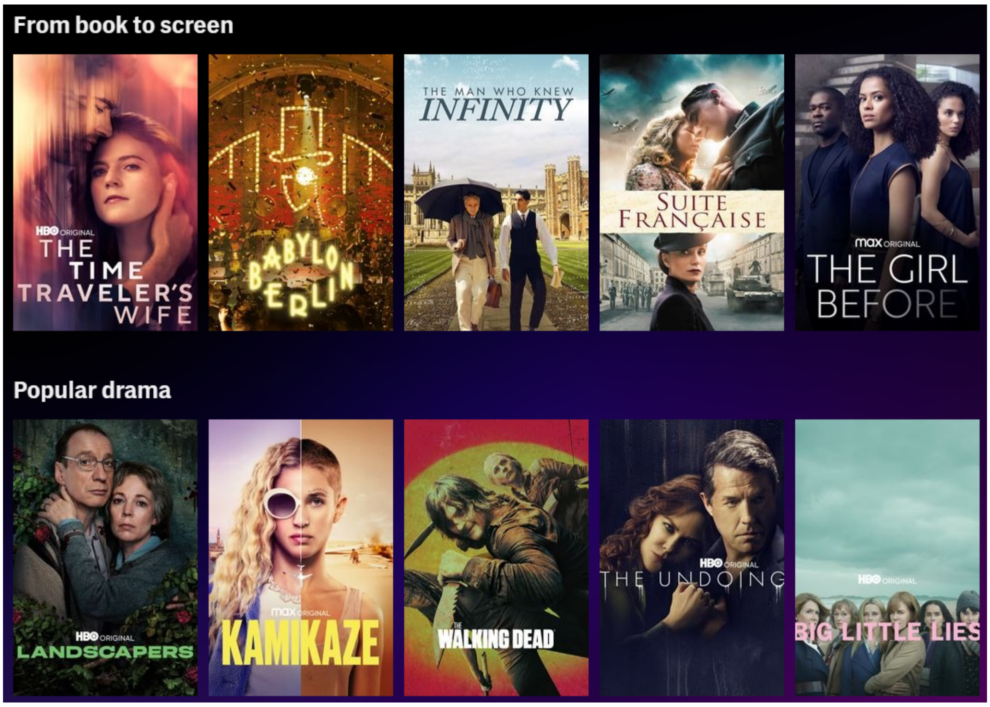

Figure 2.1: A screenshot of the multi-list interface in HBO Max where explanations are utilized to provide users with an understanding of the content.

2.3.1 Explanation Types and Interfaces

The explanation may follow the style of the underlying algorithm even though the recommendations have been retrieved or computed differently. In the early stages of research on recommender systems, a few studies emerged on different ways of explaining recommendations. One of the first within the field, Herlocker et al. [24], addressed twenty-one different explanation interfaces for automated collaborative filtering systems. In a number of experiments in a web-based movie recommender system called MovieLens, they found out explaining a collaborative-filtering recommendation in a histogram were the most compelling among the users. In addition, did explanations in general yield positive feedback amongst users. In comparison to Herlocker et al. [24], Biligic and Mooney [4] did a study where they also included content-based and user-history-based explanations and investigated the impact on users' satisfaction. Three different explanation types were evaluated by the users; neighbor style, keywords style and influence style. Neighbor style explanations which explains collaborative filtering based recommendations turned out to create overestimation of the quality of an item which could lead to mistrust from users. However, keywords style explanations which explain content-based recommendations and user-history-based which presents ratings previously provided by the user were found to be significantly more effective at helping users make more accurate decisions. Symeonidis et al. [54] later combined the two explanation styles in a hybrid recommendation approach called MoviExplain. They built a feature profile for each user, grouped the users into biclusters to detect matching preferences and used features in the explanation to justify the recommendations. Combining the two explanation styles supported their assumption in a user study that the combination would be more favorable among the users as it was more informative.

Vig et al. [65] explored the effects of tagsplanations, a way to explain recommendations through tags. It was evaluated based on its justification, effectiveness and mood compatibility. The study differentiates between justification and transparency in explanations, where transparency is the understanding of how the systems work and justification is the understanding of why an item was recommended. As seen in Figure 2.2, the explanations were based on two tag-components: the degree to which a tag accurately describes an item and the users preference towards a tag. In addition, four interfaces were designed and evaluated in a user experiment where they found tagsplanations to positively promote goals of justification, effectiveness and mood compatibility. Dominguez et al. [13] compared three recommendation interfaces with different types of explanations in the image domain. The results from the user experiment shows that explanations of recommendations were useful and increased user satisfaction, explainability and relevance.

Figure 2.2: The tagsplanations used in Vig et al. [65] that describe item-relevance and its relevance to a user's preference.

2.3.2 Personalized Explanations

The need for explanations tailored to users' preferences has increased in line with the complexity of recommender systems and individual differences [35]. Users might place different levels of importance on certain features due to personal preferences [58]. Serving item features in an explanation such as: "*This low calorie recipe is recommended to you as your diet goal is to lose weight*", can have a strong persuasive effect when presented with an important feature (diet goal) [24]. It has been found in the real estate domain that tailored evaluative arguments had a higher chance of persuading the user to adopt a particular house compared to a non-tailored evaluative argument [8]. For instance, the argument: the house has a good location was persuasive to a user that cared about the location.

Since most of the research that had been done was on the persuasive power of personalization, Tintarev and Masthoff [58] decided to investigate the impact of a personalized feature-based explanations ability to help users make good decisions (effectiveness) and its ease of use/enjoyment (satisfaction). More specifically, they investigated what makes an explanation effective, how the effectiveness could be best evaluated and if user satisfaction could have an impact on potential trade-offs. In order to represent the most important features in the explanation, the selection of features were personalized as users may differ in what they find important and their individual taste. The experiments were done in the movie and camera domain with three different types of explanations: A baseline condition where the explanation is not personalized nor described any item features. A non-personalized feature-based explanation that describes item features, but the features were not tailored to the user. A personalized feature-based explanation that both describes item features and tailored them to the users interests. Results from the experiment showed that explanations in general can help users in decision-making. In two domains across three experiments, results showed that personalized explanations hindered effectiveness, but increased satisfaction. However, satisfaction is only increased if it results in information that is meaningful to a user. A similar study on personalized explanations by Kouki et al. [35] found that personalized content-based explanations were found to be more persuasive than non-personalized popularity-based explanations.

Chang et al. [9] developed personalized natural language explanations which were inspired by how people explain word-of-mouth recommendations through a crowd-sourcing and computational process. They compared their personalized natural language based explanations to tag-based explanations [65] and examined the impact, finding it to positively affect trust and satisfaction among users. Gedikli et al. [22] measured ten different explanation types (personalized and non-personalized) by its efficiency, effectiveness, persuasiveness, perceived transparency and satisfaction of users in the movie domain. The experiment found non-personalized content-based tag cloud explanations to be most effective and accepted by the users in addition to increasing perceived transparency and satisfaction.

A study by Svcrek et al. [53] examined the combined use of basic explanation styles in order to provide an appropriate type of personalized explanation to each user in the news domain. They developed an interface called ExplORe (Figure 2.3) and made the personalized explanations in either a content-based or collaborative-filtering way, where content-based provided an explanation containing feature-similarity in the article (..because you like the topic basketball) whereas collaborative-filtering ones included user-similarity (...because similar user Robo read it too). Their findings show that users favored content-based explanations over collaborative filtering explanations, and articles with an explanation are preferred over articles without. In addition it was found that personalized explanations could increase understandability and attractiveness of recommended articles.

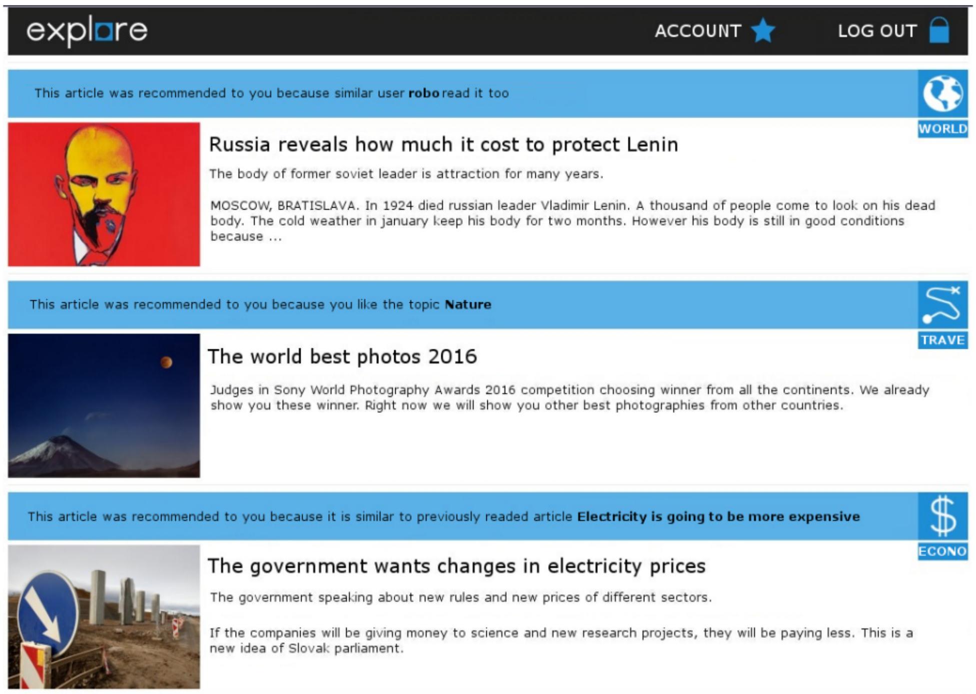

Figure 2.3: The ExplORe interface with personalized explanations developed and evaluated in the experiment by Svcrek et al. [53].

2.3.3 Explanations in Food Recommender Systems

Currently, most research on explanations is often done in domains such as products or movies. Research on the effects of explanations in the food domain have however been scarce [61] and havent flourished until recent years. A couple of research papers have used explanations in their frameworks such as a multi-modal recipe fusion framework [36] that were built to improve the performance of recipe recommendations. The artifact generated explanations derived from images or videos related to the recommended recipe. Another research paper [45] made a Food Explanation Ontology that models food and diet recommendation explanations to answer questions about food recommendations they receive from AI. In [39], Shirai et al. made a Python framework to recommend healthy food and generate explanations for each step involved in producing the recommendation. A personalized nutrition recommender system [38] included explanations for the end user to gain insights to why a recipe was recommended to them. Although explanations are included in the mentioned research papers, the effects were not investigated.

Musto et al. [41] explored natural language justifications/explanations ability to persuade users in food recommendations to make healthier food choices by emphasizing nutritional facts, health risks and/or benefits to the recommended recipes. Two recommendations were presented to the user where natural language justifications were generated and personalized based on user characteristics and the recipes features. Justification strategies that contained user features such as cooking experience or diet goals were respectively compared up against recipe features such as level of difficulty or calories. A single style justification on level of difficulty would say: “Vegetable Soup is very easy to prepare” while a comparative style would say: “Vegetable Soup is easier to prepare than Spaghetti Cacio and Pepper”. In total, eight different justification strategies were evaluated in a single or a comparative justification style. Research results have shown that the comparative justification style was effective at promoting healthiness while the justification strategy which compares each recipes features and health risks were able to cater towards a users healthy food preferences.

As found in previous research, both non-personalized and personalized explanations are commonly used in food recommender systems, but little attention has been paid to how users evaluate these explanations, as well as how they are presented with the recommendations. Positive user evaluation results have been found when providing a greater number of recommended options organized in multiple lists [10, 27, 28, 48]. By offering users with more options in the food domain, one might enable them to select from an increased number of healthier alternatives. In addition, the explanation could help users identify whats healthy and not. However, providing users with more options and explanations carries the possibility of increasing their choice overload [6].

2.4 Multi-list Recommender Systems

The extent of the benefits such as persuasion or satisfaction that users and providers achieve from a recommender system depends on the quality of the recommendations and their presentation [37, 43]. It has among other things been discovered that grouping movie genre features in a structured overview, multiple lists, has proven to be a more persuasive and satisfactory presentation method compared to a top N-items overview [43]. Another similar example of structuring recommendations in multiple lists is the organization interface designed by Pu and Chen [48]. Item recommendations are organized into different categories/multiple lists, where each category is a suggested critique/explanation and the products gathered below share specific features. After comparing the organization interface to a ranked top N-items interface in a user study, results showed that users considered the organization-based interface more efficient in making product comparisons. Also, the interface had a positive influence on users' intention to return which in fact says something about users trust in the system. The same authors have further been testing the organization interface in an eye-tracking experiment, reporting users hotspot and gaze path [10]. This time, the aim of the study was to investigate the effects of layout change where users were looking. Three different layouts were compared; a ranked top N-items layout, an organization interface with vertical layout and an organization interface with quadrant layout. From the hotspot and gaze path plot, it turns out that users in the ranked list layout tend to pay more attention to the top candidates while ignoring the remaining. Whereas in both the vertical and the quadrant organization layouts, users examined and paid attention to more recommended products. Furthermore, a higher percentage of users also successfully accomplished the task of finding and selecting a product to buy compared to the ranked list. In a study by Hu and Pu [27], the organization interface was compared to a standard list interface in order to examine perceived categorical diversity (items being of different kinds) and item-to-item diversity (items being dissimilar). Results from an in-depth user study found that organization interfaces could effectively increase users' perceived diversity of the recommendations results in the categories. This also positively influenced the perceived ease of use and usefulness of the system. Jannach et al. [28] studied the effects of multi-list interfaces on users decision-making behavior for similar-item recommendations in the movie domain. In the multi-list interface, thirty recommendations with labels attached to each list were compared to a single-list consisting of only one label and the same amount of recommendations listed below. The outcome of the experiment revealed that users favor the single-list as it allowed them to make less effort making a choice. On the other hand, multi-lists were however able to lead to more exploration and gave the user an impression of diversity and novelty in the recommendations.

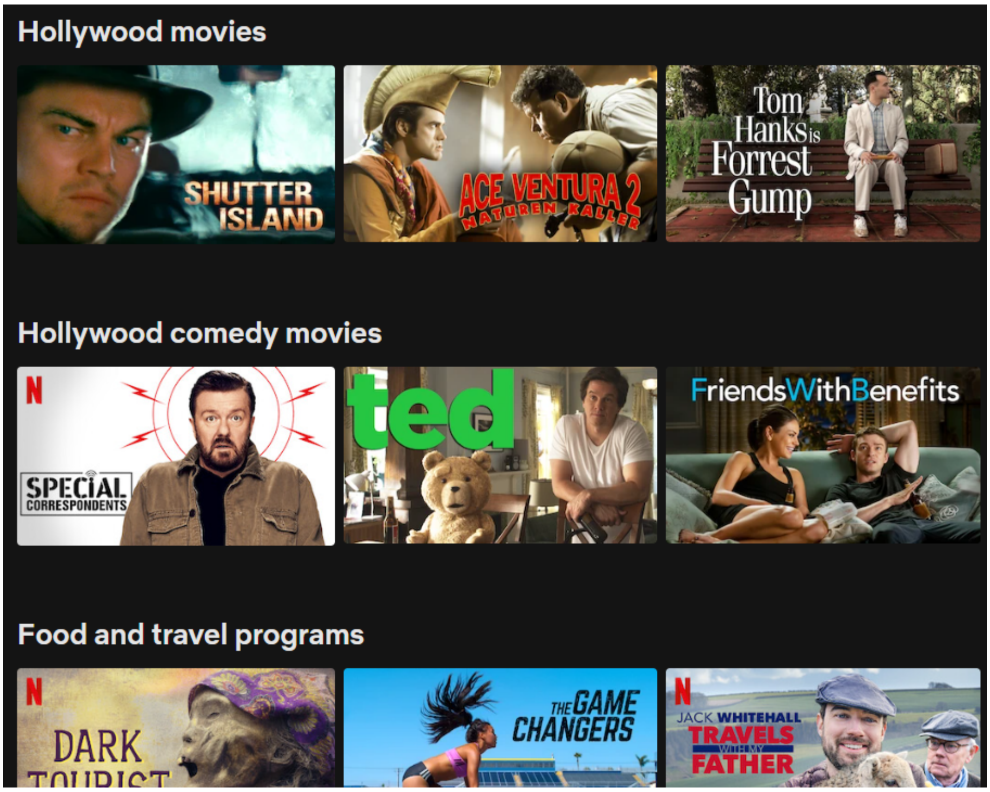

Figure 2.4: A screenshot of Netflix's multi-list interface. As depicted in the Figure, each list is accompanied with an explanation that describes its recommendations.

2.4.1 Choice Overload

An important aspect which recommender systems try to mitigate is a users experienced choice overload. The ever-increasing number of products made available to users via numerous websites may help find something of interest. However, even if the recommender system is attempting to reduce the number of options by recommending items of interest, this does not guarantee that making a decision on the next page will be any easier [23]. Studies that have been done by Bollen et al. [6] on choice overload in conjunction with recommender systems pose the question if it's necessarily better to have a large set of recommendations. They investigated the effects of presenting a large set size (20 items) compared to a smaller set size (5 items) with low versus high set quality and examined different aspects such as perceived diversity, attractiveness, choice difficulty and choice satisfaction. Their findings reveal that a large set with high quality items does not necessarily result in a higher choice satisfaction, compared to a smaller set. This is due to the fact that a larger number of high-quality recommendations is counteracted by an increased difficulty in choosing among them. A study by Beierle et al. [2] examined choice overload in related-article recommendations in digital libraries through measuring click-through rate. They found a lower click-through rate for higher numbers of recommendations and a doubled number of clicks when showing ten related articles instead of one. Their findings suggest that users quickly feel overloaded by choice.

2.4.2 Multi-list Interface for Healthier Food Choices

Latest and highly relevant work in regards to this thesis, is the research by Starke et al. [51] on multi-list interfaces. The research addressed the use of multi-list interfaces in addition to explanations in order to examine to what extent such interfaces could support users in making healthier recipe choices. The multi-list interface had a total of 25 recipe recommendations, where each list contained 5 recommendations based on multiple algorithms, such as similarity to other recipes or a recipe's fat, calorie, or carbohydrate content, with an explanation for each list. A comparison was made between the multi-list condition and a single-list condition optimized for a single algorithm containing a total of five recommendations and a single explanation. It was hypothesized that serving food recommendations in multiple lists would lead to a more diverse set of recommendations and cater towards eating goals that are not yet part of the user's profile [51]. After conducting an online user experiment, the analysis of the results showed that multi-lists were not able to support users making healthier food choices, but rather the opposite. Whereas in the single-lists with explanations, users actively chose healthier alternatives relative to what they were presented. When it came to what impact multi-lists had on users, research results showed that users are more satisfied with multi-lists compared to single-lists. Additionally, users were also more satisfied with the choices they made in the multi-lists, however, statistical evidence of perceived recommendation diversity among users was not found despite Jannach et al. [28] findings on users' impression of diversity and novelty.

The single- and multi-list interfaces were also compared with and without explanations. Analyzes of the result did not yield any significant result that explanations could persuade users to choose healthier alternatives. At the same time, no clear significance was found that explanations could increase user understandability of what was presented to them or reduce choice overload.

This thesis will, however, seek to answer the question if personalized explanations in multi-lists could yield a different result in users' perceived understandability, diversity, choice overload and in helping them make healthier choices. As explanations which includes recipe features and health risks have shown to be able to cater towards users' healthy food preferences [41], interesting results may be found when including such personalized explanations to a multi-list interface. In addition, to ensure a more accurate comparison between single- and multi-lists, this study will provide similar recommendation set sizes, 25 vs 25 instead of 5 vs 25, across both interfaces.

2.5 Summary and Differences

The cited publications have explored the use of recommender systems in the food domain, taking user preferences, health, and nutritional aspects into consideration. In addition, papers have explored several approaches of explaining recommendations (textual vs. visual) and personalizing the explanations by incorporating user preferences and item features. The mentioned research papers have also examined the impact of various recommendation presentation interfaces, as well as multi-list interfaces with explanations in the food domain. Given recent studies, there are still a number of unexplored or unaccounted-for aspects that this thesis aims to examine. Although health-aware food recommender systems include explanations for their recommendations, little research has been conducted on how users in the food domain perceive and evaluate such explanations as well as its persuasive effect in guiding users towards healthier alternatives, especially when personal and nutritional aspects are incorporated. Moreover, the presentation of food recommendations is still a little explored field where personalized explanations could have an impact on how users evaluate and perceive multi-list interfaces. In addition, personalized explanations which account for multiple recommendations at once would be a contribution in itself. The following steps will be conducted in order to address the previously described shortcomings:

- As previous research has not identified a generic explanation type in multi-lists that merely describes the content of recipes to persuade users towards healthiness or influence their evaluation of a recommender system, this study will consider user preferences, health, and nutritional recipe aspects. This is done in order to examine whether personalized explanations can provide users with a higher understanding of the presented content, greater satisfaction when choosing a recipe, assist and reduce difficulty in choosing a recipe, as well as support users in locating healthier alternatives.
- The multi-list interface will be compared to a single-list interface using a more accurate comparison of 25 versus 25 recipes as opposed to 25 versus 5 recipes, as was done in prior study. This provides a more accurate measurement of whether users perceive recommendations presented in multiple lists using different algorithms to be more diversified than recommendations presented in a single list using a single algorithm. Furthermore, both personalized and non-personalized explanations will be compared.
- An online experiment will be conducted to measure the above mentioned evaluation and perception aspects across non-personalized and personalized explanations in single- and multi-lists.

Chapter 3

Methodology

This study examined whether the use of personalized explanations in a multi-list interface could affect a user's evaluation and perception of a food recommender system, in addition to their decision to select healthier recipes. To be able to answer the posed research questions, a food recommender system was designed and organized in multiple lists with in total 25 recipes being equally distributed over 5 lists. Each list was served with a personalized explanation aimed at a user's preferences and health to explain the presented content in each list. The study compares a multi-list interface, which contains multiple algorithms, to a single-list interface, which only contains one algorithm. In addition, comparisons are made between personalized explanations and non-personalized explanations, which explain the content with a single feature. The recommender system itself used a knowledge-based approach by scoring explicit input from the users based upon their preferences and knowledge about eating healthy in order to provide accurate recommendations. This chapter goes deeper into technicalities of the prototype and the methods used.

- Section 3.1 describes the details of the dataset used in the experiment.
- Section 3.2 gives a deeper insight into the technicalities of the prototype and the prototype setup.
- Section 3.4 explains the research design method used.
- Section 3.3 describes the procedure that the user had to go through from start to finish and various variables that were measured in the study.

- Section 3.6 describes how participants were recruited to the study and further descriptive statistics on those that were recruited.

3.1 Dataset

The dataset used in the online experiment is from an Italian food community platform called GialloZafferano[1] which contains around 4671 recipes translated into english and has previously been used in studies on personalized food recommendation [41, 42]. It contains data such as the recipes name, image, difficulty, cooking time, nutrients and dietary restrictions which also were valuable in order to personalize the recommendations. Initially, the dataset had different categories such as desserts, appetizers and side dishes, but as this study mainly focuses on main courses, those recipes had to be left out. In addition, recipes that lacked useful data that were relevant to the recommendation process were also omitted. This means that the dataset eventually ended up having 1190 recipes.

3.2 Prototype

In order to conduct the experiment, two separate single- and multi-list studies with a personalized and non-personalized explanation interface as well as pre-, post-task and finishing-questionnaires were developed. A screenshot of the developed single- and multi-list interface is shown in Figure 3.1 and Figure 3.2. The front-end of the prototype was built from the ground up using web-technologies such as HTML, CSS in order to organize and present the recipes in list(s) along with a picture, description and cooking time as well as being compatible for tablet and computer users. As for the back-end of the prototype, it was built with PHP to recommend recipes and to be connected to PostgreSQL for data storage of recipes and answers. The whole infrastructure was deployed on the Heroku[2] platform, which enables running the website and data storage in the cloud. For reproducibility, the entire prototype code is available on Github[3].

[1] www.giallozafferano.it
[2] www.heroku.com
[3] www.github.com/larsholth97/personalized-exp-multilist

3.2.1 Technical Overview

In the beginning of each study, the users had to go through a pre-questionnaire in order to obtain information about their preferences and health. The information given was used in a knowledge-based approach to score recipes based on general knowledge about healthy eating and how well it fit the user's profile. This scoring system was inspired by the work of Cataldo et al. [41] where a knowledge-aware modifier was used to score a recipe based on a general understanding of food choices encoded as rules. When a rule is satisfied, the recipes score will be updated where the recipes with the highest score will be the most relevant for the user. All the applied health-rules in the prototype were derived from governmental health sites such as NHS (United Kingdom), USDA (United States) and FHI (Norway). The rules include, among other things, recommended nutritional guidelines for people who want to lose weight, diabetics and underweight. The same way of scoring the recipes were also applied for non-nutrition-oriented rules such as cooking experience or preferred cooking time. Once the score was calculated, the recipes were shuffled and further split to guarantee that both personalized and non-personalized explanation interfaces had an unique set of recipes. Each unique set was re-ranked based on its score where top 350 recipes were included further in the process.

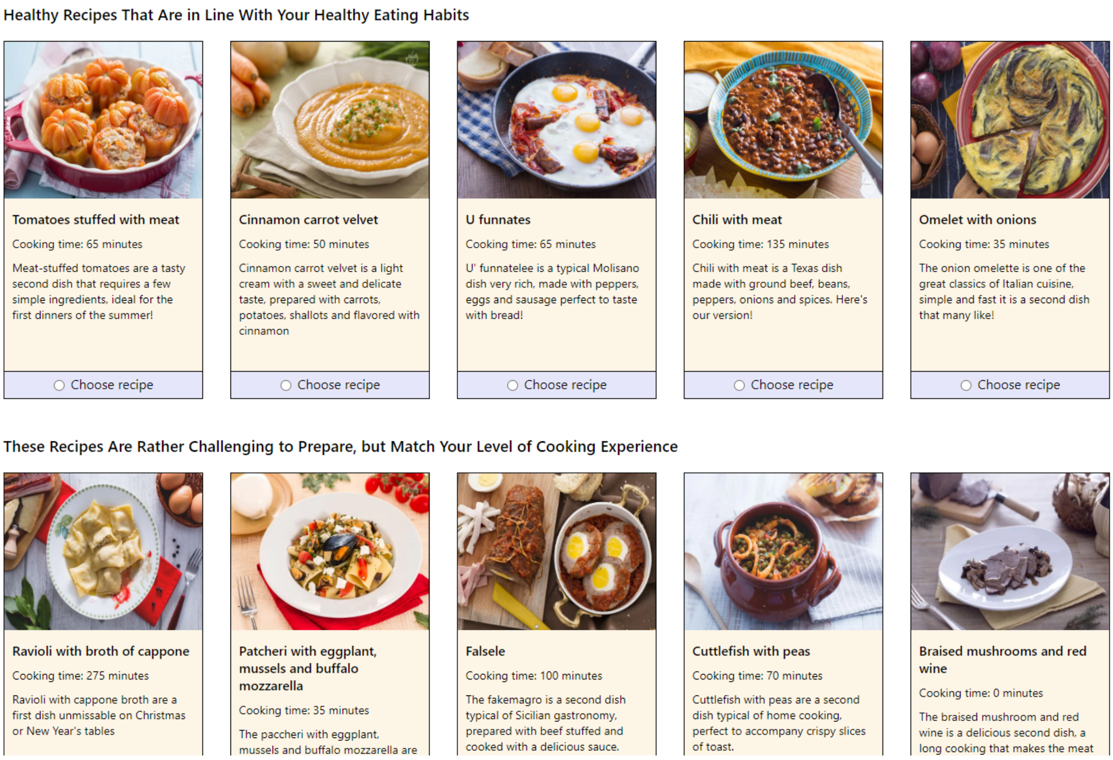

Figure 3.1: Depicted is a partial screenshot from the developed prototype of a multi-list interface with personalized explanations. Each recipe is presented with a picture, title, cooking time and a short description in 5 lists.

3.2.2 Explanations

The explanations in the prototype are related to the information provided in the pre-questionnaire to ensure that they are of relevance to the user. However, the actual structure of the explanations differs from non-personalized to personalized explanations. While a non-personalized explanation describes the content with a single feature such as Low-calorie recipes, a personalized explanation takes the feature and links it explicitly to the user's preferences or health. For instance, if a person enters their diabetes dietary restriction, the personalized interface could provide an explanation saying: "Low-calorie recipes that also fit your diabetes dietary restriction". It is important to point out that the features that are linked are in line with knowledge obtained from the governmental health sites. In total, there were 29 personalized and 13 non-personalized pre-made explanations to choose from, where all the explanations are shown in Table 3.1. Depending on how much information the user chooses to provide, the number of personalized and non-personalized explanations generated in each session varied between 5-8 across all conditions. For the purpose of exposing users to different explanations, all the generated personalized and non-personalized explanations were shuffled. Users in the multi-list study were shown five explanations in each condition whereas single-list users were only shown one in each condition.

Table 3.1: An overview of all the explanations used in the prototype. There were 13 non-personalized and 29 personalized explanations.

Non-Personalized Explanations	Personalized Explanations
Low-Calorie Recipes	Low-Calorie Recipes That Also Fit Your Diabetes Dietary Restriction Low-Calorie Recipes That Match Your Weight-Loss Goal Low-Calorie Recipes That Match Your Body Mass Index and High Level of Physical Activity Low-Calorie Recipes That Match Your Body Mass Index and Low Level of Physical Activity
High-Calorie Recipes	High-Calorie Recipes That Match Your Weight-Gain Goal
Low-Fat Recipes	Low-Fat Recipes That Match Your Body Mass Index and High Level of Physical Activity Low-Fat Recipes That Match Your Body Mass Index and Low Level of Physical Activity Low-Fat Recipes That Match Your Weight-Loss Goal
High-Fat Recipes	High-Fat Recipes That Match Your Weight-Gain Goal High-Fat Recipes That Match Your Body Mass Index and High Level of Physical Activity High-Fat Recipes That Match Your Body Mass Index and Low Level of Physical Activity
Challenging Recipes to Try	These Recipes Are Rather Challenging to Prepare, but Match Your Level of Cooking Experience
Recipes That Are Easy to Cook	These Recipes Are Very Easy to Prepare, Which Matches Your Low Level of Cooking Experience These Recipes are Easy to Prepare, Which Matches Your Low Level of Cooking Experience
Healthy Recipes That Meet Dietary Intake Guidelines	Healthy Recipes That Are in Line With Your Healthy Eating Habits Healthy Recipes That Could Improve Your Unhealthy Eating Habits Healthy Recipes That Fit Your Lactose-Free Dietary Restriction Healthy Recipes That Fit Your Gluten-Free Dietary Restriction Healthy Recipes That Fit Your Vegetarian Dietary Preferences
Recipes With a Long Cooking Time	Recipes With a Long Cooking Time, but No Longer Than Your Preferred Cooking Time Recipes With a Long Cooking Time That Match Your Preferred Cooking Time and Your-Diabetes Dietary Restriction
Recipes With a Short Cooking Time	Recipes With a Rather Short Cooking Time, Which Matches Your Preferences
Low-Sugar Recipes	Low-Sugar Recipes That Match Your Body Mass Index and Low Level of Physical Activity Low-Sugar Recipes That Match Your Body Mass Index and High Level of Physical Activity
High-Protein Recipes	High-Protein Recipes That Match Your Weight-Gain Goal High-Protein Recipes That Match Your Body Mass Index and High Level of Physical Activity
High-Fiber Recipes	High-Fiber Recipes That Match Your Body Mass Index and High Level of Physical Activity High-Fiber Recipes That Match Your Body Mass Index and Low Level of Physical Activity
Recipes Low in Saturated Fat	Recipes Low in Saturated Fat That Match Your Weight-Loss Goal

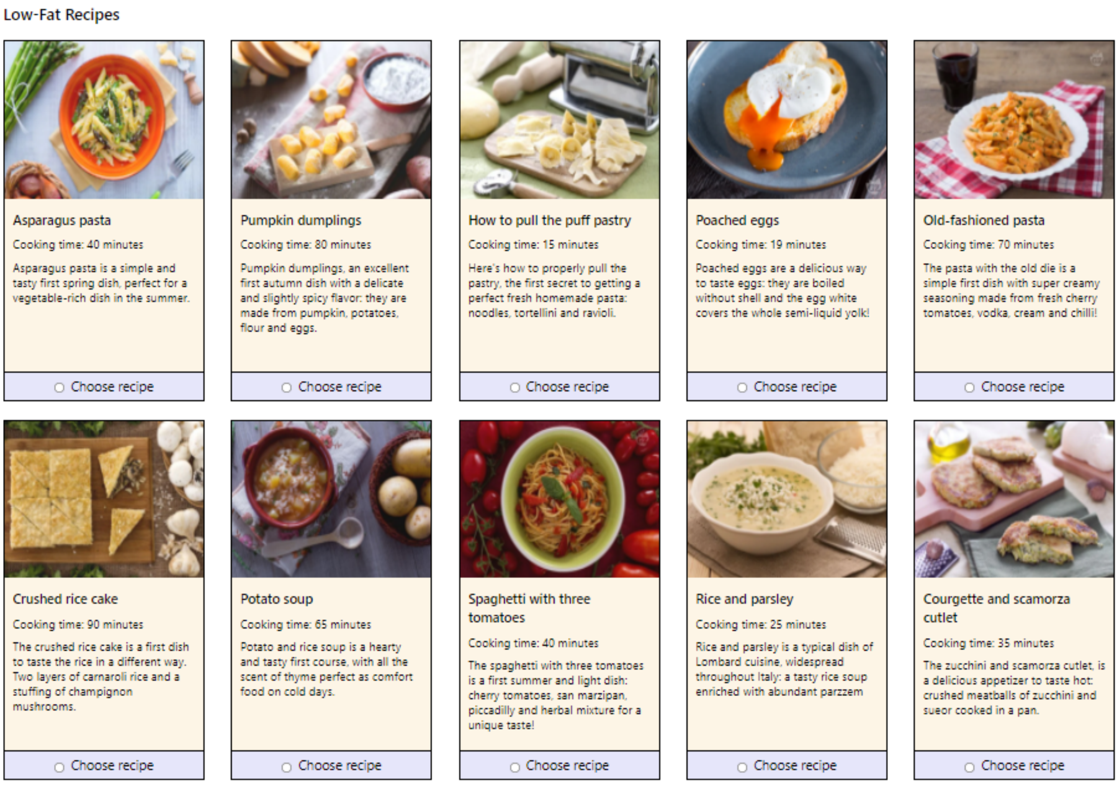

Figure 3.2: A partial screenshot from the prototype of a single-list interface with non-personalized explanations.

3.2.3 Post-filtering Approach

Following the creation of the explanations, a post-filtering method was applied in order to fill the lists according to the given feature the explanation contains. By using a function to sort each recipe, the lists could either be sorted and filled from low to high or from high to low. This means that a user who wants to lose weight could receive a list with the personalized explanation: "Low-fat recipes that match your weight-loss goal" and the non-personalized explanation: "Low-fat recipes". The lists will thus be filled with recipes retrieved from the scored set and sorted by fat content from low to high. A single-list would then eventually end up with 25 sorted recipes under a single explanation, while a multi-list interface would have 25 recipes distributed under 5 lists and sorted according to each explanation.

3.3 Research Design

To be able to compare and evaluate the effect of the interfaces, a 2x2 mixed research design, seen in Figure 3.3, was used in the user experiment. This is a type of experimental design that makes it possible to understand the effect of manipulated independent variables between and within a group. In this experiment, the users were distributed approximately equally between the two studies: one group was shown a single-list interface while the other group was shown a multi-list interface, which was the manipulation between the groups. In addition, a manipulation was performed within each group where users were presented with personalized and non-personalized explanations along with the list(s). As displaying recipe recommendations in a single-list with a non-personalized explanation is the most prevalent among various food recipe websites such as GialloZafferano.it and Allrecipes.com, this was then used as a baseline for this study. Furthermore, non-personalized explanations in conjunction with single- and multi-lists have already been researched while personalized explanations are yet to be explored.

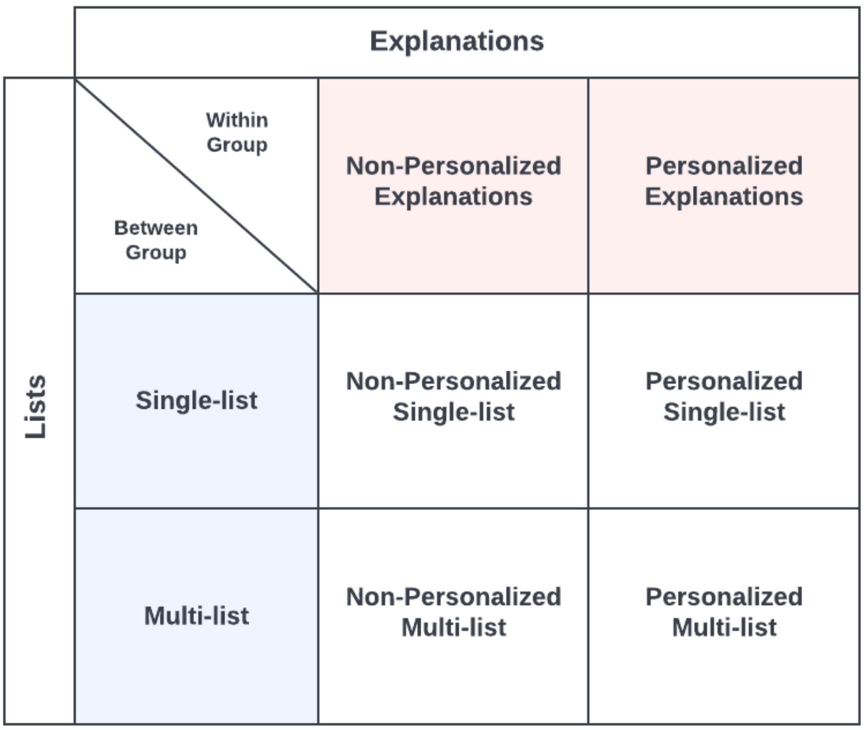

Figure 3.3: The figure above shows the 2x2 mixed research design used for the user experiment. As depicted, the manipulation between the groups were the types of lists, whereas manipulations within the group was the type of explanation presented.

3.4 Procedure

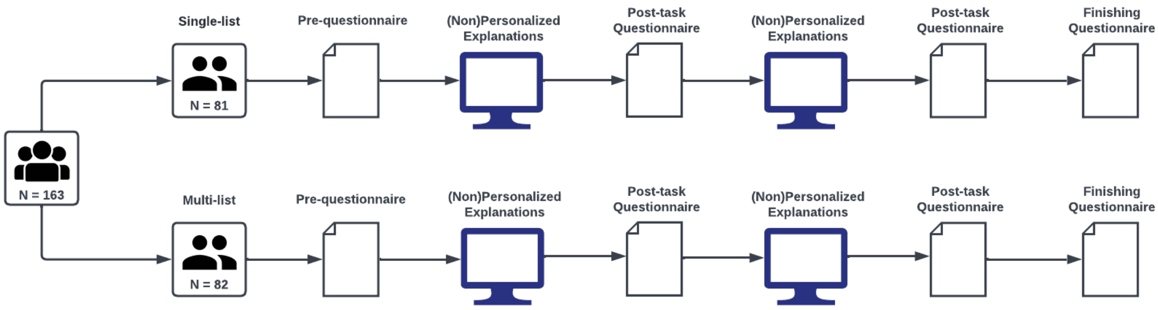

Figure 3.4: An overview of the procedure with four questionnaires and two recipe choice tasks.

The whole experiment was split into a single- and multi-list study where procedures in both interfaces were similar, consisting of four screens with accompanying tasks. The whole procedure are depicted in Figure 3.4. In the first task, the users were asked to fill out a pre-questionnaire which included questions about demographic information such as age and gender in addition to questions regarding height and weight in order to calculate their BMI, which was taken into account when recommending recipes. Furthermore, the users had to fill out their importance of eating healthy, their self-reported eating habits, cooking experience and health consciousness on a 5-point scale. Other questions regarding food preferences and health-related goals such as their preferred cooking time, physical activity, diet goals and diet restrictions were also given. Diet goals included losing weight and gaining weight while diet restrictions included diabetes, gluten-free, lactose-free and vegetarian diets. Most of the questions in the pre-questionnaire were either obtained or inspired by the work of Cataldo et al. [42]. Upon completion of the pre-questionnaire, it was random whether a personalized or non-personalized explanation interface was shown first across single- and multi-list conditions. Both explanation interfaces presented 25 recipes and asked the users to choose only one recipe that they liked the most and would prepare for tomorrow night, or in the near future. Following the recipes, they were given a post-task questionnaire with 13 statements to rate on a likert scale of 1-7, based on their level of agreement. The statements regarded their understandability, perceived helpfulness, choice difficulty and choice satisfaction to the interface and their choice of recipe. After selecting two recipes and rating all the statements, a finishing-questionnaire with four statements was given to address the users perceived diversity, which was also to be assessed from on a likert scale from 1-7. The statements used in the post-task and finishing questionnaire were partly based on previous studies, with understandability derived from [1, 33], perceived helpfulness from [26, 52], choice difficulty from [1, 30] choice satisfaction from [1, 33] and perceived diversity from [1, 30]. All the questions used in this experiment can be seen in Table 3.2.

Table 3.2: An overview over all the questions used in the experiment divided into the different questionnaires.

Phase	Questions	Input
Pre Questionnaire	To which gender do you most identify?	Scale (Male, Female, Other/Prefer not to say)
	What is your age?	Scale
	How important is eating healthy to you?	Range 1-5 (Very Unimportant to Very important)
	What is your weight? (in kilos)	Numeric Scale
	What is your height? (in centimeters)	Numeric Scale
	How would you rate your eating habits?	Range 1-5 (Very unhealthy to Very healthy)
	To what extent do you agree with the follow statement: "My health depends on the foods I consume."	Range 1-5 (Completely disagree to Completely agree)
	How do you rate your level of cooking experience?	Range 1-5 (Very low to Very high)
	What is your maximum preferred cooking time (in minutes)?	Numeric Scale
	How often do you engage in physical exercise per week?	Range 1-3 (Little to Alot)
	Do you have any weight-related goals?	Options (Lose weight, Gain weight, No goal)
	Do you have any dietary restrictions?	Options (Diabetes, Vegetarian, Lactose, Gluten-free)
Post-Task Questionnaire	I like the recipe Ive chosen.	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	I think I will prepare the recipe Ive chosen.	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	I know many recipes that I like more than the one I have chosen	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	I would recommend the chosen recipe to others.	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	The task of choosing a recipe was overwhelming.	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	I changed my mind several times before choosing a recipe.	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	Comparing the recommended recipes was easy.	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	I could easily find recipes on this page.	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	This page helped to discover new recipes.	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	A page like this helps me make better recipe choices	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	I understood why the recipes were recommended to me.	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	I could understand how the recipes were based on my preferences.	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	The recommendation process was NOT clear to me.	Likert Scale 1-7 (Completely Disagree to Completely Agree)
Finishing Questionnaire	Several recipes in each list of recommended recipes differed strongly from eachother	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	The recommendation lists included recipes from many different categories	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	Both interfaces contained recipes that were similar to each other	Likert Scale 1-7 (Completely Disagree to Completely Agree)
	No two recipes seemed alike	Likert Scale 1-7 (Completely Disagree to Completely Agree)

3.5 Measures

To be able to understand the impact personalized explanations and multi-list interfaces have on users, in addition to which aspects could affect a users choice and evaluation, the following aspects were measured: healthiness of the chosen recipe, position of the chosen recipe in the interface, recipe characteristics, user characteristics and user evaluation aspects.

3.5.1 WHO Score

A way to measure the overall healthiness of a recipe is to calculate the WHO score. The score ranges from 0-7 where a score of 0 is seen as very unhealthy while 7 is very healthy. The calculation itself was done according to the approach from Howard et. al. [25] where the 7 most important nutrients (i.e. proteins, carbohydrates, sugars, sodium, fats, saturated fats and fibers) were included in addition to the scale from 0-7. The WHO health score has also been used in other food recommender system studies [44, 62]. In this experiment, the WHO score was used in one of the lists showing recipes with a score from high to low. The score was used in the experiment to see how healthy the recipes chosen by the users were and analyzed in order to address RQ2.1. Analyses were also conducted to examine its impact on the user evaluation aspects choice satisfaction and difficulty.

3.5.2 Choice Metrics and List Layout

To be able to keep track of the order in which recipe choice the user makes in the list interfaces, data about its position were collected in the database. The position variables were denoted as pos-y and pos-x for the multi-list interface, where pos-y referred to which list/row the recipe was selected from, with values ranging from 0-4 and pos-x denoted the position of the recipe in the list, which ranged from 1-5. For the single-list interface, a pos variable denoted the position of where the recipe was chosen from in a range from 0-24. In addition, the variable pos-x was used to examine more closely where in the specific list the recipe was chosen from. For each recipe choice the user makes, data about which explanation-type the recipe was chosen from and all the other explanations presented were also gathered.

3.5.3 Personal Characteristics

In the pre-questionnaire, users were asked several personal questions which were used in order to score recipes based on their preferences and health. The questions addressed their gender, age, their importance of eating healthy, weight and height. Moreover, users were asked to rate their eating habits (very unhealthy to very healthy), health consciousness (completely disagree to completely agree), and cooking experience (very low to very high) on a scale from 1 to 5. In addition, maximum preferred cooking time, their engagement in physical exercise, weight goals and any dietary restrictions. The questions regarding eating habits, health consciousness and cooking experience were further included in the analysis to determine its impact on user evaluation aspects such as choice satisfaction and difficulty in addition to its impact on the healthiness of the chosen recipe (WHO Score).

3.5.4 Recipe Characteristics

Beside measuring the overall healthiness of the chosen recipe, it was also addressed if any other recipe characteristics were affected by the use of personalized explanations and multi-list interfaces. More specifically, nutrient density of fat, saturated fat, salt, protein, sugar, carbohydrates and fiber were examined in the chosen recipe to address RQ2.2.

3.5.5 User Evaluation Aspects

To address RQ1.1 and RQ1.2, users were asked about their evaluation and perception of the interface following each task of selecting a recipe (post-task questionnaire) in terms of how satisfied they are with the choice they made, difficulty in making a choice, perceived helpfulness and their perceived understanding. Furthermore, after completing both tasks (finishing-questionnaire), users were questioned regarding their perception of the recommendations' diversity. All the questions were assessed on a likert scale from 1-7 (completely disagree to completely agree) and further included in a factor analysis in order to better understand the data and identify any underlying dimensions. Eventually, Cronbachs alpha were applied on the retrieved factors to determine how closely the items are related as a group. The results from the factor analysis can be seen in Table 3.3.

Table 3.3: The results from the factor analysis on the post-task questionnaire items and finishing questionnaire items.

Factor	Item	Loading
Choice Satisfaction	I like the recipe I've chosen.	0.8192
$\alpha = 0.83$	I think I will prepare the recipe I've chosen.	0.8712
	I know many recipes that I like more than the one I have chosen.	
	I would recommend the chosen recipe to others.	0.9083
Choice Difficulty	The task of choosing a recipe was overwhelming.	0.8465
$\alpha = 0.66$	I changed my mind several times before choosing a recipe.	0.8217
	Comparing the recommended recipes was easy.	-0.5899
Perceived Helpfulness	I could easily find recipes on this page.	
	This page helped to discover new recipes.	
	A page like this helps me make better recipe choices	
Understandability	I understood why the recipes were recommended to me.	0.8947
$\alpha = 0.85$	I could understand how the recipes were based on my preferences.	0.8986
	The recommendation process was NOT clear to me.	-0.8399
Percevied Diversity	Several recipes in each list of recommended recipes differed strongly from eachother	0.8768
$\alpha = 0.69$	The recommendation lists included recipes from many different categories	0.8768
	Both interfaces contained recipes that were similar to each other	
	No two recipes seemed alike	

Factor Analysis. The factor analysis was conducted on the post-task questionnaire and finishing questionnaire items separately due to the post-task questionnaire being answered multiple times by a single user. The initial analysis on post-task questionnaire items identified three factors: choice satisfaction, difficulty and understandability. However, one factor which regarded perceived helpfulness and a questionnaire item regarding choice satisfaction were left out due to factor loading less than 0.5. Further, Cronbach's alpha tests were performed to measure the internal consistency of each factor. The test showed an alpha score of $\alpha = 0.85$ for items regarding understandability and a score of $\alpha = 0.83$ for choice satisfaction which could be considered as good according to [55]. For items regarding choice difficulty, an alpha score of $\alpha = 0.66$ were obtained which could be considered as acceptable.

As seen in Table 3.3, a factor analysis were also conducted on the finishing-questionnaire items which regarded users perceived diversity in the recommendations from single- and multi-list interfaces. The factor analysis was conducted on four questionnaire items, where a single factor was identified for perceived diversity. The factor loadings for two questionnaire items did not have a loading over 0.5 and were further omitted. By including the remaining questionnaire items in a Cronbachs alpha test, a score of $\alpha = 0.69$ were retrieved, which is also considered acceptable.

3.6 Participants

All the participants in the study were recruited via Prolific, an online web-recruitment service that allows researchers to recruit from all over the world which yields high-quality data, according to Peer et al. [47]. The single- and multi-list studies were published separately on Prolific; however, to prevent people from taking part in both experiments, the multi-list study was conducted first. In both studies, a total of (N=163) participants were recruited, with (N=81) in the single-list condition and (N=82) in the multi-list condition. From the total number of participants, 54,6% identified as female and 44,7% as male. The mean age was 30,1 with a standard deviation of 10,8.

Prolific allows participants to be pre-screened before the survey is published. Because the survey was conducted in English, a filter based on English fluency was used to guarantee that all participants understood the task. To eliminate any individuals with bad intentions, a 98 percent approval rate was used. In addition, users were also filtered out on whether they have any dietary restrictions. This filter was applied since omitting a large number of recipes (desserts, appetizers, and side dishes) could result in relatively few recommendations when users select more than two or three dietary restrictions in the prototype. Moreover, each participant was rewarded £1.11 for their involvement in both single- and multi-list studies, which took on average 7 minutes to complete.

3.6.1 Descriptive Statistics

In the preliminary questionnaire, participants were asked to evaluate their own eating habits. On a scale ranging from very unimportant to very important, 53.3% of recruited participants (N=163) across both conditions evaluated the importance of eating healthily as important. In contrast, 49.7 percent of respondents self-assessed their own eating habits as neither healthy or unhealthy, which was also the middle option. Moreover, given the statement, "My health depends on the foods I consume", 61.3% of respondents agree on a scale ranging from entirely disagree to completely agree. In terms of physical activity and diet goals, 52.1% of respondents reported 3 hours of weekly physical activity, while 40.5% reported 6 hours, which was also considered average in the study. In addition, slightly more than half of the participants

(50.9%) had a diet goal of weight loss, whereas 36.8% had no diet goal. In terms of personal preferences related to cooking, 39.2% of respondents rated their experience as medium, while 37.4% rated it as high. The average cooking time requested by participants was 53.37 minutes. Participants could tick a box if they had any dietary limitations, although a filter for dietary restrictions were applied in Prolific, 18.4 percent of participants did have a restriction, with lactose-free diet being the most ticked option.

Chapter 4

Results

The following chapter describes the analysis methods conducted for this research and its findings. Each section is structured along the posed research questions where the following method and its results are reported below. It was examined if the usage of personalized explanations which incorporates user preferences, health and nutritional aspects in a multi-list interface could affect a users perception and evaluation of a food recommender system in addition to supporting healthier choices.

- Section 4.1 provides an overview of the results from the analysis performed in order to answer the research question if a users perception is affected by the use of personalized explanations and multi-list interfaces.
- Section 4.2 provides an overview of the results from the analysis performed in order to answer the research question regarding if personalized explanations can affect choice satisfaction and difficulty across multi- and single-list interfaces.
- Section 4.3 presents the findings of the analysis conducted to answer the question if personalized explanations and multi-list interfaces can support healthier recipe choices.
- Section 4.4 present the findings of the analysis conducted to answer the question if personalized explanations and multi-list interfaces can affect any other recipe choice aspects.
- Section 4.5 reports additional findings to identify frequent presented explanations as well as favorable explanations and recipes.

To be able to answer the research questions, a two-way ANOVA test was conducted to examine the differences in the effects of two independent variables, combined, (lists and explanations) had on a single dependent variable (understandability, WHO Score). Additionally, a two-sample t-test was performed on the lists to examine its impact on perceived diversity. Further, a regression analysis was performed on the independent variables lists, explanations and their interaction, as well as other various user aspects to examine its effect on choice satisfaction and difficulty, separately.

4.1 User's Perception (RQ1.1)

To answer RQ1.1, "*To what extent is a users perception of a food recommender system affected by the use of personalized explanations and a multi-list interface?*", a two-sample t-test on the predicted variable of perceived diversity was performed to examine how diverse users perceive the recommendations in multi-lists. In addition, a two-way ANOVA test on the predicted variable of perceived understandability was performed to investigate if personalized explanations in a multi-list interface could increase users understanding of what was presented to them.

Perceived Diversity. A two-sample t-test on perceived diversity was used as it was solely measured between the list conditions and its results are reported in Table 4.1. The results from the t-test showed higher levels of perceived diversity in the multi-list condition (M = 0.18, SD = 0.85) compared to the single-list condition (M = -0.18, SD = 1.10): $t(164)$ = -3.32, $p < 0.05$. This suggests that presenting recommendations in multiple lists where each list is optimized for different algorithms makes the users perceive the recommendations as more diverse.

Table 4.1: Results from the two-sample t-test on perceived diversity between lists.

Condition	Observations	Mean	Std. err.	Std. dev.
Single-list	162	-0.182	0.086	1.106
Multi-list	164	0.180	0.066	0.847

Perceived Understandability. In addition to measuring the users perceived diversity, their perceived understandability of the presented recommendations were also measured. Since understandability was measured across all four experimental conditions, a two-way ANOVA test was therefore conducted on the independent variables with the reported results in Table 4.2. The results from the analysis showed a marginally higher, but non-significant, increase of perceived understandability in multi-lists (M = 0.08) compared to single-lists (M = -0.08): $F(1{,}322) = 2.15$, $p = 0.143$. This marginal increase can also be seen in Figure 4.1. As for the explanations, no significant results were found $F(1{,}322) = 0.29$, $p = 0.587$ nor any interaction effect between lists and explanations $F(1{,}322) = 0.00$, $p = 0.991$.

Table 4.2: Results from the two-way ANOVA on users perceived understandability between lists and explanations.

Condition	**df**	**SS**	**MS**	*F*	*p-value*
Explanation	1	0.294	0.294	0.29	0.587
List	1	2.153	2.153	2.15	0.143
Explanation:List	1	0.0001	0.0001	0.00	0.991
Residual	322	322.55	1.0017		

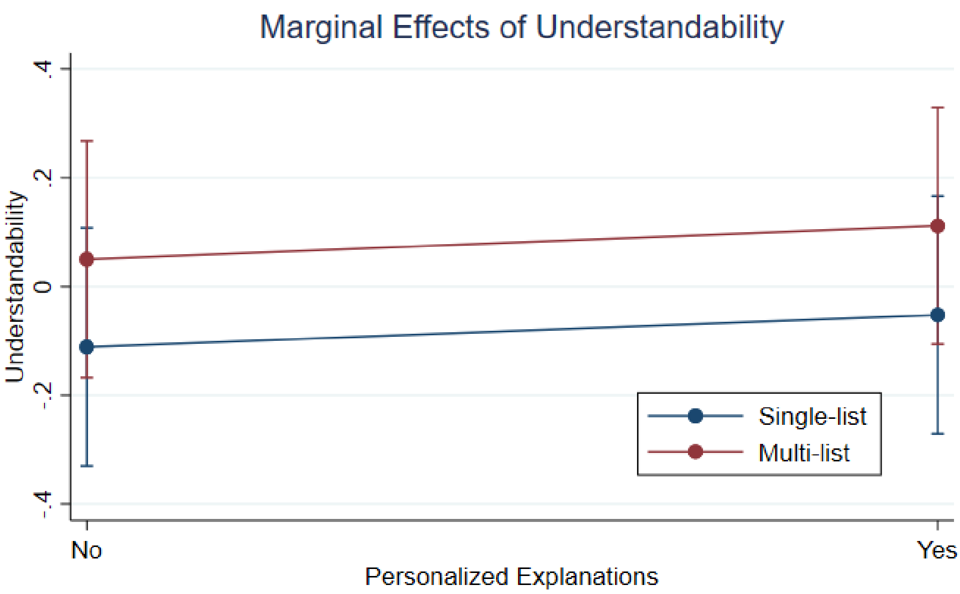

Figure 4.1: Marginal effects of understandability across lists and explanations.

4.2 Choice Satisfaction and Choice Difficulty (RQ1.2)

In order to answer RQ1.2, "*To what extent do personalized explanations affect choice satisfaction and difficulty across multi-list and single-list interfaces?*", a multiple regression analysis was conducted separately on the post-task questionnaire items regarding choice satisfaction and choice difficulty based on the experimental conditions (lists and explanations), in addition to user perceptions, personal characteristics and interaction data. The results on all the variables are shown in Table 4.3.

Choice Satisfaction. A multiple regression analysis was conducted on choice satisfaction to examine if multi-lists and personalized explanations could affect users satisfaction in making a choice. As seen in Table 4.3, the results from the analysis found no statistical significance on the explanations ($\beta = 0.009$, $p = 0.893$), which indicates that personalized explanations could not affect users choice satisfaction across single- and multi-lists. The list conditions, on the other hand, differed unexpectedly, as multi-lists led to a lower level of choice satisfaction compared to a single-list interface ($\beta = -0.172$, $p = 0.027$), which also can be seen in Figure 4.2. However, no interaction effect was found between the explanations and lists ($\beta = 0.063$, $p = 0.440$).

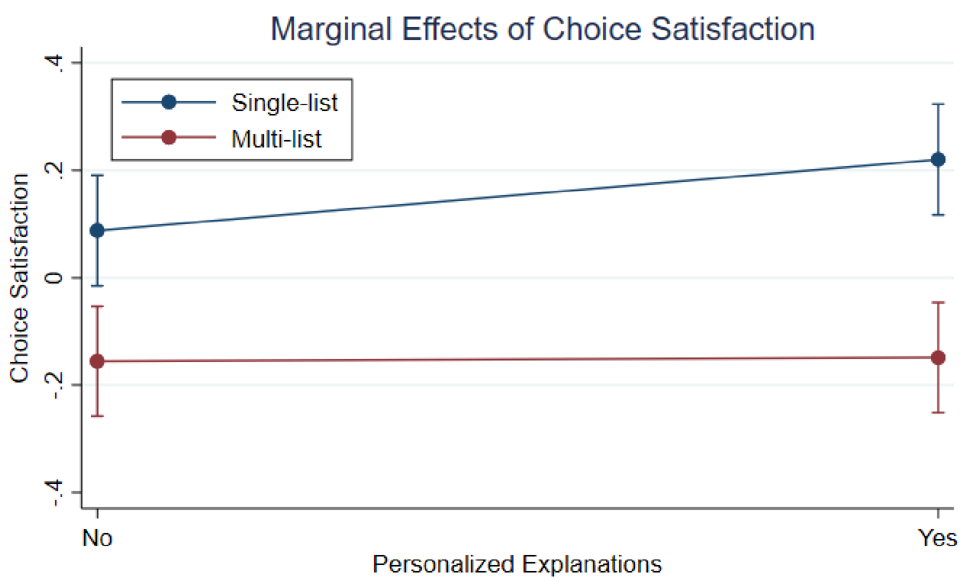

Figure 4.2: Marginal effects of choice satisfaction between lists and explanations. Lower levels of choice satisfaction were found in multi-lists compared to single-list.

Choice Difficulty. Regarding if personalized explanations could affect the choice difficulty among users across multi- and single-list interfaces, a multiple regression analysis was performed on the predictors, lists and explanations, and are listed in Table 4.3. The results from the analysis did not yield any statistical significance between the explanations ($\beta = -0.008$, $p = 0.918$), nor any statistical significance between the list conditions ($\beta = 0.012$, $p = 0.891$). Additionally, there was not found any interaction effect between the lists and explanations. The following results indicate that personalized explanations could not affect choice difficulty in either multi- or single-list interfaces in any significant way.

Other Aspects

Several other aspects such as users perceptions, personal characteristics and interaction data was also included in the multiple regression analysis to examine whether it could affect a users choice satisfaction and difficulty. The variable on users' perceptions includes perceived diversity and perceived understandability, whereas personal characteristics consists of eating habits, cooking experience and health consciousness. In addition, interaction data consists of the healthiness of the chosen recipe (WHO score) and the position of the chosen recipe. All the variables used in the analysis and its results are shown in Table 4.3.

Table 4.3: The results from the separately conducted regression analysis on the dependent variables choice satisfaction, choice difficulty and several independent variables.

	Choice Satisfaction			Choice Difficulty		
Choice Aspects	Beta	Std. Err.	p-value	Beta	Std. Err.	p-value
Multi-list	-0.172	0.155	0.027	0.012	0.177	0.891
Personalized	0.009	0.134	0.893	-0.008	0.152	0.918
Multi-list & Personalized	0.063	0.188	0.440	0.039	0.214	0.674
Cooking Experience	0.120	0.059	0.019	0.193	0.067	0.001
Health Consciousness	0.131	0.072	0.009	-0.096	0.082	0.091
Eating Habits	0.030	0.061	0.548	-0.076	0.069	0.180
List Position	0.025	0.043	0.729	-0.151	0.048	0.035
Who Score	0.027	0.029	0.591	0.030	0.033	0.585
Perceived Diversity	0.200	0.050	0.000	0.067	0.057	0.239
Understandability	0.341	0.049	0.000	-0.135	0.060	0.026
Choice Difficulty	-0.103	0.381	0.037			

User Perceptions. A multiple regression analysis was used to examine the impact of choice evaluation aspects from RQ1.1, perceived understandability, and perceived diversity on choice satisfaction and difficulty. It was found that users who indicated a higher understandability were more satisfied with their choices ($\beta = 0.341$, $p < 0.001$) and reported less difficulty in making them ($\beta = -0.135$, $p = 0.026$). Furthermore, users who perceived the presented items as more diverse reported higher levels of choice satisfaction ($\beta = 0.200$, $p = < 0.001$), however, it was not found to affect choice difficulty significantly ($\beta = 0.067$, $p = 0.239$).

Personal Characteristics. The following personal characteristics: cooking experience, health consciousness and eating habits were included in a multiple regression analysis to examine its impact on choice satisfaction and choice difficulty. As seen in Table 4.3, users with higher levels of cooking experience indicated higher levels of choice satisfaction ($\beta = 0.120$, $p = 0.019$) as well as higher levels of choice difficulty ($\beta = 0.193$, $p = 0.001$). In addition, higher levels of health consciousness was found to positively affect choice satisfaction ($\beta = 0.131$, $p = 0.009$), however, no statistically significance was found for its impact on choice difficulty ($\beta = -0.096$, $p = 0.091$). Moreover, no statistical significance was found on eating habits impact on choice difficulty ($\beta = -0.076$, $p = 0.180$) and choice satisfaction ($\beta = 0.030$, $p = 0.548$).

Interaction Data. Regarding interaction data, the following variables were included in a multiple regression analysis to examine their impact on choice satisfaction and choice difficulty: list position, which defines the vertical position of the chosen recipe, and WHO score, which determines the healthiness of the chosen recipe. For the list position seen in Table 4.3, it was found that users who selected their recipe higher up in the interface, experienced higher levels of choice difficulty ($\beta - 0.151$, $p = 0.035$), but no effect on choice satisfaction ($\beta = 0.025$, $p = 0.729$). The WHO score could not positively affect choice difficulty ($\beta = 0.030$, $p = 0.585$) and choice satisfaction ($\beta = 0.027$, $p = 0.591$).

4.3 Personalized Explanations & Multi-lists for Healthier Choices (RQ2.1)

In order to answer RQ2.1, "*To what extent do personalized explanations and multi-list interfaces support healthier recipe choices?*", the WHO score, which determines the overall healthiness of a chosen recipe, was included in a two-way ANOVA test between list and explanation conditions. As reported in Table 4.4, the findings from the analysis indicated that the WHO score of the chosen recipe was lower in the personalized explanation interface (M = 2.21) compared to a non-personalized explanation interface (M = 2.78): $F(1, 322) = 10.16$, p = 0.0016. This suggests that providing explanations that take nutritional aspects and users health into account unexpectedly encouraged users to make less healthy choices. Further, there was not found any statistical significance between the list conditions $F(1{,}322) = 0.04$, $p = 0.837$, nor any interaction effect between explanations and lists $F(1{,}322) = 0.65$, $p = 0.420$.

Table 4.4: The results of the two-way ANOVA on the WHO Score between lists and explanations. Findings show a significant result between non-personalized and personalized explanations.

Condition	**df**	**SS**	**MS**	*F*	*p-value*
Explanation	1	26.61	26.61	10.16	0.0016
List	1	0.11	0.11	0.04	0.837
Explanation:List	1	1.70	1.70	0.65	0.420
Residual	322	843.15	2.61		

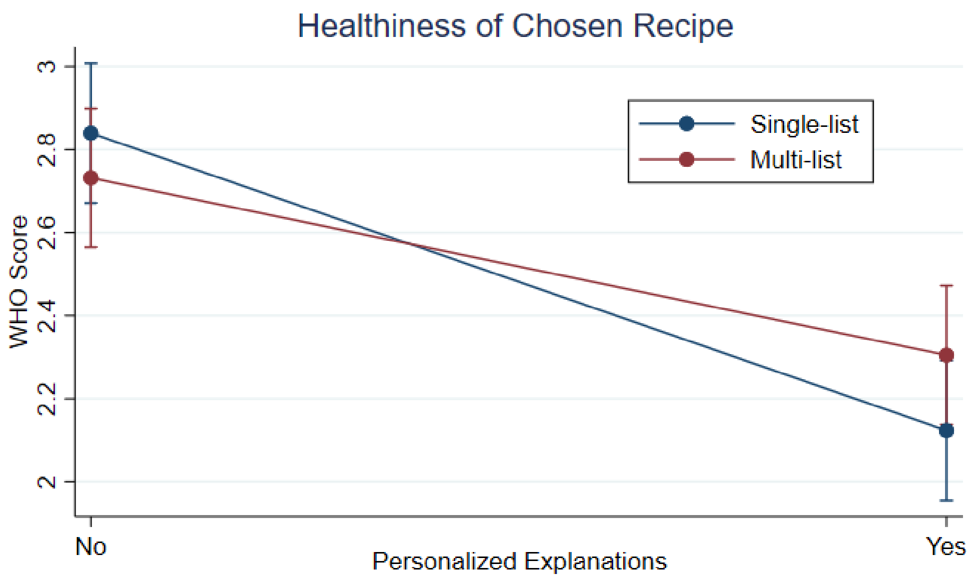

Figure 4.3: Marginal effects of the WHO Score between list and explanation conditions.

Personal Characteristics. As there was found statistical significance between the explanation conditions, it was also examined if any of the personal characteristics from the pre-questionnaire and the explanations had an impact on the WHO score of the chosen recipe. Following characteristics such as cooking experience, eating habits and health consciousness were computed as continuous variables and examined in multiple two-way ANOVA tests between the explanations on the WHO score. Results from the analysis, reported in Table 4.5, found an interaction effect between eating habits and explanations $F(1, 322) = 6.01$, $p = 0.015$. By taking a closer look at the WHO score in the different user answers across the explanation conditions, one can observe that users who self-reported 'very unhealthy' eating habits chose healthier recipes (M = 3.77) compared to users with 'very healthy' eating habits in the non-personalized explanation condition (M = 1.93): $F(1, 322) = 6.01$, $p = 0.015$. This suggests that users with rather unhealthy eating habits could make healthier choices with non-personalized explanations. For the remaining personal characteristics, there was no statistically significant interaction effect between explanations and cooking experience $F(1,322)$ = 3.62, p = 0.058, nor any statistical significant interaction effect between explanations and health consciousness $F(1,322) = 0.22$, $p = 0.641$.

Table 4.5: Results from the two-way ANOVA test on the WHO Score between explanations and eating habits.

Condition	df	SS	MS	*F*	*p-value*
Explanation	1	25.85	25.85	10.13	0.001
Eating Habits	1	8.13	8.13	3.19	0.075
Explanation:Eating Habits	1	15.34	15.34	6.01	0.015
Residual	322	871.5	2.68		

4.4 Other Recipe Choice Aspects (RQ2.2)

To answer RQ2.2, "*Do personalized explanations and multi-list affect other recipe choice aspects, compared to non-personalized explanations and single list interfaces?*", an analysis was performed on the density of the following nutrients in the chosen recipe: fat, saturated fat, sugar, carbohydrates, proteins, fiber and sodium. As some of the lists and explanations not only addressed the healthiness of the presented recipes, but also more specific nutrients, interesting findings were made when examining the impact of the lists and explanations on various nutrients in the chosen recipe. A two-way ANOVA analysis was conducted on each of the mentioned nutrients between explanations and lists. Significant results were identified for the nutrients fat, saturated fat, salt, and protein, but not for sugar, carbohydrates, or fiber.

Fat. Reported in Table 4.6, the initial two-way ANOVA test on fat density of the chosen recipe found statistical significance in the main effect of explanations, where chosen recipes in the personalized explanation interface had a higher fat density ($M = 48.25$) compared to the non-personalized explanation interface ($M = 38.79$): $F(1, 322) = 3.93$, $p = 0.048$. No interaction effect was found between lists and explanations $F(1,322) = 0.55$, $p = 0.460$.

Table 4.6: Results from the two-way ANOVA test on fat density between lists and explanations.

Condition	**df**	**SS**	**MS**	*F*	*p-value*
Explanation	1	7319.39	7319.39	3.93	0.048
List	1	2129.07	2129.07	1.14	0.285
Explanation:List	1	1018.19	1018.19	0.55	0.460
Residual	322	599873.29	1862.96		

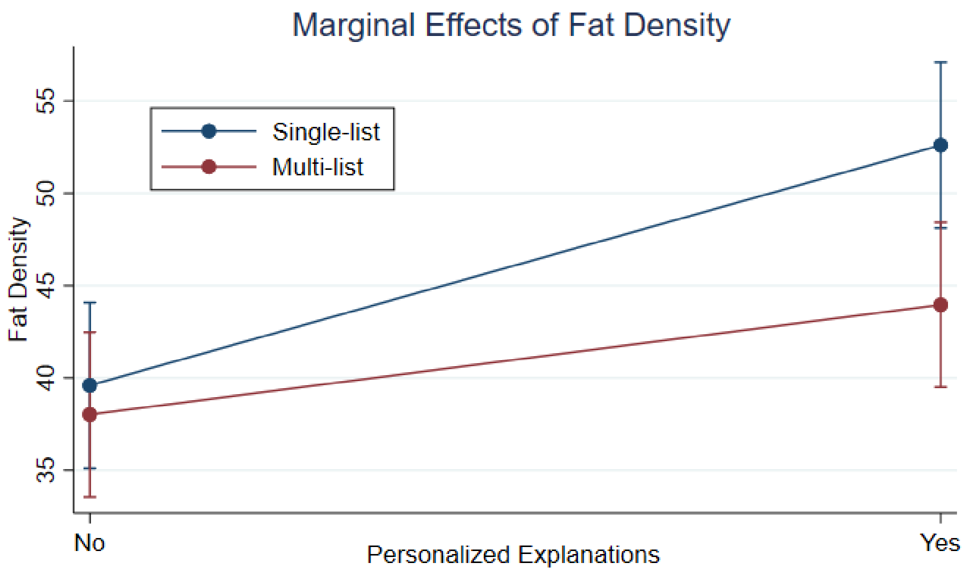

Figure 4.4: Marginal effects of fat density of the chosen recipes across list and explanation conditions. Higher levels of fat were found when served personalized explanations.

Saturated Fat. Moreover, a two-way ANOVA test was performed on saturated fat density which also found statistical significance in the main effect of explanations, showing a higher level of saturated fat in the chosen recipes with personalized explanations (M = 12.93) compared to non-personalized explanations (M = 10.21). An interaction effect was not found between lists and explanations $F(1,322) = 0.06$, $p = 0.812$.

Sodium. A two-way ANOVA test on sodium density across explanations and lists was performed, finding that users in the single-list chose recipes with higher levels of sodium (M = 649.64) compared to multi-list users (M = 495.91): $F(1,322) = 3.89$, $p = 0.05$. However, no interaction effect was found between lists and explanations $F(1,322) = 1.60$, $p = 0.207$.

Protein. A two-way ANOVA test was conducted on the density of protein in a chosen recipe across lists and explanations. Results showed that recipes chosen in the single-list condition had higher levels of protein (M = 34.79) compared to recipes chosen in the multi-lists (M = 26.25) $F(1, 322) = 4.58$, $p = 0.033$. No interaction effect was found between lists and explanations $F(1,322) = 0.00$, $p = 0.986$.

4.5 Other Findings

Besides the analyzes related to the research questions, some smaller, but interesting findings were also made in order to uncover any favorable and frequently presented explanations or recipes across the conditions. It was measured which recipe was the most popular, from which list/explanation the majority of users selected their recipe, and which explanation(s) was/were provided the most. The abovementioned measurements were counted in each of the conditions (single- and multi-list, non-personalized and personalized), where each user had to choose a recipe two times.

In the single-list condition, 'chicken in soy sauce' was chosen the most (7 times). In terms of explanations, single-list users could only choose from a single explanation, implying that the counting only applied to the most presented one. The most presented non-personalized explanations in the single-list was: 'Healthy Recipes That Meet Dietary Intake Guidelines' (14 times), which ranked the overall healthiness of the recipes (WHO Score) from high to low and 'Low-Calorie Recipes' (14 times), which ranked the recipes calorie content from low to high. As for personalized single-list, the explanation 'These Recipes are Rather Challenging to Prepare, but Match Your Level of Cooking Experience' which ranks recipes from high difficulty to low were presented the most (10 times).

In the multi-list condition, the most chosen recipe amongst users was 'Omelette (Basic recipe)' (8 times). For the non-personalized multi-list condition, when examining how many times a recipe has been chosen from a certain list in relation to how many times the list appeared to the user, one can observe that users chose a recipe from 'Recipes Low in Saturated Fat' 9 times out of 23 appearances (39.1%). However, similarly to the non-personalized single-list, most of the choices in the non-personalized multi-list condition were made in the lists/explanations: 'Low-Calorie Recipes' (12 times, 20%) and 'Healthy Recipes That Meet Dietary Intake Guidelines' (12 times, 19.7%). When it comes to personalized explanation lists with the greatest number of chosen recipes in relation to the overall amount of appearances, users chose a recipe from the list: 'Recipes With a Long Cooking Time, but No Longer Than Your Preferred Cooking Time' 9 times out of 26 appearances (34.6%). However, for the lists where the most recipes were chosen from was: 'Recipes with a rather short cooking time, Which Matches Your Preferences' (12 times, 29.2%) and 'Healthy Recipes That Are in Line With Your Healthy Eating Habits' (12 times, 21.8%).

Chapter 5

Discussion

Finding a preferred food recipe while making a healthy choice can be a challenging task due to the increasingly large number of recipes available online and the frequent presence of unhealthy foods. In addition, there is a trend among various online recipe websites, where it is common to present recommendations in a single list interface that can only account for a single algorithm and explanation. Despite the fact that various online streaming services present recommendations in multiple lists, taking into account multiple algorithms and explanations, little study has been conducted on the impact of integrating multi-list interfaces in the food domain. In addition, no research has been conducted on the effect of different types of explanations in such an interface on the user's perception and preference towards healthier choices. The goal of this thesis is to investigate the usage of personalized explanations in a multi-list food recommender system that links user preferences and health to nutritional aspects. More specifically, it will be examined how such explanations and interfaces could affect users' choice satisfaction, difficulty, perceived understandability, and diversity, in addition to supporting them in making healthier choices. Previous research have already examined structuring recommendations in multiple lists on effectiveness, intention to return and cognitive effort [48] in addition to comparing multi-lists to a top-N view on user satisfaction and persuasion [43]. Furthermore, a single-list interface have been compared to a multi-list interface and its impact on choice difficulty, choice satisfaction, perceived similarity and diversity [28]. Most recently, multi-lists with explanations have been examined in the food domain on its persuasiveness towards healthier choices, as well as its effect on users' perceived diversity, understandability, choice satisfaction and choice difficulty [51]. As much of the mentioned research on single- and multi-lists are mainly based around its interface, no research has been done on the impact of different types of explanations in such interfaces, especially not when linking users preferences and health to nutritional aspects.

Users Perception (RQ1.1). For the RQ1.1, regarding the impact of personalized explanations and multi-list interfaces on users perception, it was conducted a two-sample t-test on perceived diversity between the list conditions and a two-way ANOVA analysis on perceived understandability between the lists and explanations. For perceived diversity, the results from the analysis found that users in the multi-lists condition perceived their recommendations as more diverse compared to the single-list condition. This implies that presenting recommendations in multiple lists, where each list is optimized for different algorithms makes the user perceive recommendations as more varied. A food recommender system with a diverse set of options allows users to explore more options while also discovering potentially healthier options [14]. These results are consistent with the findings by Jannach et al. [28], where multi-lists gave the users an impression of diversity and novelty in the recommendation. Additionally, Hu and Pu [27] have also discovered in their study on organization-based interfaces that users perceive items divided in categories/lists as more diverse.

One of the aims of explanations is to provide users with transparency, which involves assisting them in understanding how the system works [56]. However, this study seeks to answer if personalized explanations and multi-list interfaces could affect users understanding in why an item was recommended. Based on the results, no significant evidence was found in this study that non-personalized and personalized explanations alone or in combination with single- and multi-lists, could help users achieve a greater understanding of the presented recommendations. Furthermore, multi-list users showed a slight, but non-significant increase in understandability. In a prior study on multi-lists and explanations by Starke et al. [51], there was also no increased understanding among users. However, Svrcek et al. [53] argued that personalized explanations led to greater understandability in the news domain, whereas Vig et al. [65] discovered that tag-explanations in the movie domain, which combine item features and user preference information, assisted users understand why a recommendation was made through a visual and textual format. When it comes to providing nutritional information through explanations, the insignificant findings of explanations in multi-lists may be specific to the food domain, where the results could be associated with lack of food literacy, which involves how a user understands nutritional information and how they are acting upon that knowledge [5]. A possible avenue of exploration could be to include a more visual explanation type as in [65] to ensure a broader understanding among users.

Choice Satisfaction and Difficulty (RQ1.2). A regression analysis was performed to analyze the impact of personalized explanations in a multi-list interface on choice satisfaction and choice difficulty. There was no significant evidence that personalized explanations alone or in conjunction with multi-lists, had a significant impact on users choice satisfaction. Previous research on personalized explanations has focused on users' satisfaction with the system, rather than their satisfaction with the choices they make, leading to Tintarev and Masthoff's [58] findings that personalized explanations were found to be satisfying among users, especially if the user found it meaningful. Dominguez et al. [13] found that explanations in general contributed to an increase in user satisfaction with the interface. The non-significant results may be due to how users perceive the quality of the explanations and recommendations, as personalized recommendations have been shown to increase the perceived quality of recommendations and further positively affect choice satisfaction [34]. Users may have considered the personalized explanations and recipes unrelated to their taste, preferences, or health, causing them not to express any satisfaction with their choices. Interestingly, statistical significant evidence was identified between single- and multi-lists, where users experienced much lower satisfaction in the multi-lists compared to the single-list interface. In comparison to several similar studies, the findings were rather surprising as Starke et al. [51] observed the exact opposite, that users were more satisfied with their choice in the multi-lists. In addition, Nanou et al. [43] found that a structured overview, similar to a multi-list interface, proved to be the most satisfying among users while Pu and Chen [48] found their organization-based interface to be highly effective in increasing users overall satisfaction. The dissimilar set sizes across single- and multi-list interfaces in this study and the research by Starke et al. [51], which is highly comparable, may account for the contradictory findings. This study had the same number of recommendations in both single- and multi-list conditions (25 vs. 25), whereas Starke et al. [51] had an unequal number of recommendations in both conditions (5 vs. 25). When comparing sets of equal size, it appears that the increase in satisfaction disappears. Another possibility is that users might have perceived a multi-list interface with such lengthy explanations to be a bit overwhelming. Strangely, this did not increase choice difficulty, as it has previously been discovered to negatively impact choice satisfaction [6].

Further analyses were done to examine if personalized explanations could have an impact on choice difficulty in single- and multi-list interfaces. One purpose that explanations seek to achieve is to assist users in recognizing useful options while discarding irrelevant ones, as well as to shorten the time it takes to perform a task [49, 56]. From the analysis, it was not found that non-personalized and personalized explanations itself or in combination with single- and multi-lists could affect choice difficulty. This is contrary to previous studies who did find a similar interface with explanations to improve users perception of the interfaces competence, resulting in users spending less time in selecting an item [48] whereas Tintarev and Masthoff [58] found that explanations in general can help users in decision-making. Even when leveling the amount of recommendations (25 vs 25), there was no significant evidence that single- or multi-lists seemed to have any impact on choice difficulty between users. Given Willemsen et al. [67] findings that diversification might be an effective way to lower choice difficulty and the fact that users in this study evaluated multi-lists as more diverse, one could have expected a decrease in choice difficulty in multi-lists.

It was also investigated which factors (user perceptions, personal characteristics and interaction data) could potentially have an impact on choice satisfaction and difficulty. The analysis revealed that users who had a better understanding of what was presented to them were more satisfied with their recipe choices and experienced less difficulty in making them. In addition, users who perceived the recommendations as more diverse did also report higher levels of choice satisfaction. These results partly reflect those of Starke et al. [51] who found users with higher understanding to perceive the recommendations as more diverse and be more satisfied with their choices. Users with more cooking experience, on the other hand, had more difficulty making a decision, which could be explained by Kamis and Davern [31] finding that users with greater degree of domain knowledge perceive recommender systems to be more difficult to use than novices. An interesting discovery reveals that users who selected their recipe higher up in the interface saw greater levels of difficulty in making their choice.

Personalized Explanations & Multi-lists for Healthier Choices (RQ2.1). To be able to answer RQ2.1, a two-way ANOVA analysis was performed on the WHO score between explanations and lists. The results from the analysis found that users chose unhealthier recipes when presented with personalized explanations compared to non-personalized ones. In other words, explanations emphasizing user's health and preferences were unable to persuade users towards healthiness; rather, users deliberately chose unhealthier recipes, which was a quite surprising result. The unexpected result could be explained by Nynke van der Laan and Orcholska [64] findings that highlighting a healthier alternative to an unhealthy choice leads to healthier purchase behavior, but that emphasizing the alternatives healthiness cancels the effect out. On the other hand, previous research on personalized explanations suggests that serving explanations with features of importance to the user could have a strong persuasive effect [24] as users place different levels of importance on certain features [58]. It was therefore expected that providing personalized explanations that ties nutritional factors to users health would boost the likelihood of users finding an explanation/list of importance and further persuade them towards healthier alternatives. Another potential factor that could have had an impact between the conditions, which also could be considered a limitation, is the unequal amount of explanations across non-personalized and personalized conditions. This resulted in some explanations being presented more frequently than others. As observed in Section 4.5, the list optimized for recipes with the highest WHO Score occurred the most (12 times) in the non-personalized condition across single- and multi-lists, which in fact increases the likelihood of users choosing a recipe from a list with high WHO Score in the non-personalized condition.

In terms of the user interface, it was not found that multi-lists alone or in combination with explanations, could support healthier food choices which is contrary to Starke et al. [51] findings, that multi-list users chose unhealthier recipes than single-list users. A possible explanation may again point to the differences in set sizes (5 vs 25) between single- and multi-lists in [51], which could affect the availability of healthy versus unhealthy recipes. Once the set sizes were similar, the healthiness of the chosen recipe appeared to be more balanced between the conditions.

It was also checked if any of the personal characteristics (cooking experience, eating habits, health consciousness) could have an impact on the WHO score of the chosen recipe. An interesting finding was made in the non-personalized explanation condition, where users who self-reported their eating habits as very unhealthy made healthier choices compared to users with very healthy eating habits. This suggests that users with unhealthy eating habits actually were able to choose healthy recipes without having to be assisted by personalized explanations.

Other Recipe Choice Aspects (RQ2.2). In RQ2.2, it was checked if recipe choices with personalized explanations and multi-list interfaces were affected by various nutrients (fat, saturated fat, sugar, carbohydrates, proteins, fiber and sodium). It was found that users tend to select recipes that are high in fat and saturated fat when presented with personalized explanations compared to non-personalized ones. This discovery may also be one of the determinants for the lower WHO Score findings in the personalized explanation interface (RQ2.1). Although there were explanations indicating recipes with low fat and low saturated fat, users were still led towards fatty foods. It is possible that users' decision may be due to the use of recipe images, as research indicates that users' perceptions of fat content might be influenced by the information they have access to, such as the title and images [17]. For the list interfaces, higher levels of sodium and protein were found in the single-list compared to multi-list interfaces.

5.1 Limitations

This study holds a couple of limitations which could potentially have had an impact on the outcome. First, in the produced prototype, recipes were scored based on their nutritional content in grams, not their nutrient density. Commonly, the nutritional density of foods is described as the amount of selected nutrients per 100 kcal, 100g, or a serving size [15]. By evaluating the amount of nutrient content rather than its density, the scoring system becomes unequal as meal portions can vary across different recipes in the dataset. As indicated in this chapter under RQ2.1, another drawback of the prototype was the big gap in the number of pre-made explanations between non-personalized and personalized explanation conditions. This may have resulted in an increase in the occurrence of some lists, which may have affected the degree of healthiness between non-personalized and personalized explanations.

This study recruited participants via a crowdsourcing platform, which has some inherent disadvantages, such as the possibility that some users may submit incorrect information about themselves, which may affect how well they understand the tasks, or that they provide unengaging and hasty answers. What could, however, have controlled for users not rushing through is the use of attention checks during the experiment to eliminate any hurried answers. Another factor is that some users may not have been motivated, as they were not actively seeking food recipes or were generally uninterested in them. In addition, while this study recruited participants from a wide variety of countries, there may have been mismatches in cuisine for some participants, as this study was based on Italian recipes.

5.2 Future Research

Based on the findings of this study, it would be interesting to investigate further several changes and aspects to the research on personalized explanations in multi-list interfaces. A change in the prototype could take into account the mentioned shortcomings such as nutritional density, a more equal number of pre-made explanations in both conditions and further obtaining even more detailed information about the users. The information could include contextual factors such as users' current mood which have previously shown to affect food choice [20] or other factors such as sleep patterns, stress and feelings of hunger.

Reducing the number of explanations offered at the same time could be another interesting area for future research. As [35] discovered that users prefer no more than three to four explanations at once in the music domain, it would be interesting to examine whether reducing the number of personalized explanations and lists in multi-lists from five to three or four in the food domain would produce different results in terms of choice satisfaction and difficulty. By lowering the number of explanations and lists, one must determine the optimal number of recommendations for each list. This opens up the possibility of determining the most suitable amount of recommendations below each explanation and list, as recommendations should not necessarily be restricted to 25. Additionally, future research might investigate delivering varied numbers and types of explanations to different users, given that past research has shown different human types to be persuaded by varying numbers of explanations [35] and that there are variances in perceptions between beginners and experts [31]. There could also be substantial alterations to the explanations, where future research on explanations in multi-lists could focus on comparing differences in showing textual versus visual explanations or a combination. As visual explanations in previous research [24, 65] yielded positive results, it would be interesting to examine if such results are transferable to the food domain and multi-list interfaces.

Bibliography

[1] Edis Asotic. Multi-list food recommender systems for healthier choices. Master's thesis, The University of Bergen, 2021.

[2] Felix Beierle, Akiko Aizawa, and Joeran Beel. Exploring choice overload in related-article recommendations in digital libraries. *arXiv preprint arXiv:1704.00393*, 2017.

[3] Devis Bianchini, Valeria De Antonellis, Nicola De Franceschi, and Michele Melchiori. Prefer: A prescription-based food recommender system. *Computer Standards & Interfaces*, 54:64–75, 2017.

[4] Mustafa Bilgic and Raymond J Mooney. Explaining recommendations: Satisfaction vs. promotion. In *Beyond personalization workshop, IUI*, volume 5, page 153, 2005.

[5] Lauren G Block, Sonya A Grier, Terry L Childers, Brennan Davis, Jane EJ Ebert, Shiriki Kumanyika, Russell N Laczniak, Jane E Machin, Carol M Motley, Laura Peracchio, et al. From nutrients to nurturance: A conceptual introduction to food well-being. *Journal of Public Policy & Marketing*, 30(1):5–13, 2011.

[6] Dirk Bollen, Bart P Knijnenburg, Martijn C Willemsen, and Mark Graus. Understanding choice overload in recommender systems. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 63–70, 2010.

[7] Peter Brusilovski, Alfred Kobsa, and Wolfgang Nejdl. *The adaptive web: methods and strategies of web personalization*, volume 4321. Springer Science & Business Media, 2007.

[8] Giuseppe Carenini and Johanna D Moore. An empirical study of the influence of user tailoring on evaluative argument effectiveness. In *INTERNATIONAL JOINT CONFER-*

ENCE ON ARTIFICIAL INTELLIGENCE, volume 17, pages 1307–1314. LAWRENCE ERLBAUM ASSOCIATES LTD, 2001.

[9] Shuo Chang, F Maxwell Harper, and Loren Gilbert Terveen. Crowd-based personalized natural language explanations for recommendations. In *Proceedings of the 10th ACM conference on recommender systems*, pages 175–182, 2016.

[10] Li Chen and Pearl Pu. Users' eye gaze pattern in organization-based recommender interfaces. In *Proceedings of the 16th international conference on intelligent user interfaces*, pages 311–314, 2011.

[11] Sally Jo Cunningham and David Bainbridge. An analysis of cooking queries: implications for supporting leisure cooking. 2013.

[12] Maria Dickson-Spillmann and Michael Siegrist. Consumers knowledge of healthy diets and its correlation with dietary behaviour. *Journal of Human Nutrition and Dietetics*, 24(1):54–60, 2011.

[13] Vicente Dominguez, Pablo Messina, Ivania Donoso-Guzmán, and Denis Parra. The effect of explanations and algorithmic accuracy on visual recommender systems of artistic images. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pages 408–416, 2019.

[14] Larissa S Drescher, Silke Thiele, and Gert BM Mensink. A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. *The Journal of nutrition*, 137(3):647–651, 2007.

[15] Adam Drewnowski, Mathieu Maillot, and Nicole Darmon. Testing nutrient profile models in relation to energy density and energy cost. *European journal of clinical nutrition*, 63(5):674–683, 2009.

[16] Mehdi Elahi, Mouzhi Ge, Francesco Ricci, Ignacio Fernández-Tobías, Shlomo Berkovsky, and Massimo David. Interaction design in a mobile food recommender system. In *CEUR Workshop Proceedings*. CEUR-WS, 2015.

[17] David Elsweiler, Christoph Trattner, and Morgan Harvey. Exploiting food choice biases for healthier recipe recommendation. In *Proceedings of the 40th international acm sigir conference on research and development in information retrieval*, pages 575–584, 2017.

[18] Alexander Felfernig, Gerhard Friedrich, Dietmar Jannach, and Markus Zanker. *Constraint-Based Recommender Systems*, pages 161–190. Springer US, Boston, MA, 2015. doi:10.1007/978-1-4899-7637-6_5.

[19] Jill Freyne and Shlomo Berkovsky. Intelligent food planning: personalized recipe recommendation. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 321–324, 2010.

[20] Meryl P Gardner, Brian Wansink, Junyong Kim, and Se-Bum Park. Better moods for better eating?: How mood influences food choice. *Journal of Consumer Psychology*, 24(3):320–335, 2014.

[21] Mouzhi Ge, Francesco Ricci, and David Massimo. Health-aware food recommender system. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 333–334, 2015.

[22] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. How should i explain? a comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies*, 72(4):367–382, 2014.

[23] Gerald Häubl and Valerie Trifts. Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing science*, 19(1):4–21, 2000.

[24] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 241–250, 2000.

[25] Simon Howard, Jean Adams, and Martin White. Nutritional content of supermarket ready meals and recipes by television chefs in the united kingdom: cross sectional study. *Bmj*, 345, 2012.

[26] Rong Hu and Pearl Pu. A study on user perception of personality-based recommender systems. In *International conference on user modeling, adaptation, and personalization*, pages 291–302. Springer, 2010.

[27] Rong Hu and Pearl Pu. Enhancing recommendation diversity with organization interfaces. In *Proceedings of the 16th international conference on Intelligent user interfaces*, pages 347–350, 2011.

[28] Dietmar Jannach, Mathias Jesse, Michael Jugovac, and Christoph Trattner. Exploring multi-list user interfaces for similar-item recommendations. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, pages 224–228, 2021.

[29] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. *Recommender systems: an introduction*. Cambridge University Press, 2010.

[30] Yucheng Jin, Nava Tintarev, and Katrien Verbert. Effects of personal characteristics on music recommender systems with different levels of controllability. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pages 13–21, 2018.

[31] Arnold Kamis and Michael J Davern. Personalizing to product category knowledge: exploring the mediating effect of shopping tools on decision confidence. In *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, pages 10–pp. IEEE, 2004.

[32] Abdus Salam Khan and Achim Hoffmann. Building a case-based diet recommendation system without a knowledge engineer. *Artificial Intelligence in Medicine*, 27(2):155–179, 2003.

[33] Bart P Knijnenburg, Svetlin Bostandjiev, John O'Donovan, and Alfred Kobsa. Inspectability and control in social recommenders. In *Proceedings of the sixth ACM conference on Recommender systems*, pages 43–50, 2012.

[34] Bart P Knijnenburg, Martijn C Willemsen, and Stefan Hirtbach. Receiving recommendations and providing feedback: The user-experience of a recommender system. In *International Conference on Electronic Commerce and Web Technologies*, pages 207–216. Springer, 2010.

[35] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. Personalized explanations for hybrid recommender systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pages 379–390, 2019.

[36] Zhenfeng Lei, Anwar Ul Haq, Adnan Zeb, Md Suzauddola, and Defu Zhang. Is the suggested food your desired?: Multi-modal recipe recommendation with demand-based knowledge graph. *Expert Systems with Applications*, 186:115708, 2021.

[37] Juha Leino and Kari-Jouko Räihä. Case amazon: ratings and reviews as part of recommendations. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 137–140, 2007.

[38] Nadja Leipold, Mira Madenach, Hanna Schäfer, Martin Lurz, Nada Terzimehic, Georg Groh, Markus Böhm, Kurt Gedrich, and Helmut Krcmar. Nutrilize a personalized nutrition recommender system: an enable study. *HealthRecSys@ RecSys*, 2216:24–29, 2018.

[39] Deborah L McGuinness. Healthy food recommendation and explanation generation using a semantically-enabled framework.

[40] Stefanie Mika. Challenges for nutrition recommender systems. In *Proceedings of the 2nd Workshop on Context Aware Intel. Assistance, Berlin, Germany*, pages 25–33. Citeseer, 2011.

[41] Cataldo Musto, Alain D Starke, Christoph Trattner, Amon Rapp, and Giovanni Semeraro. Exploring the effects of natural language justifications in food recommender systems. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, pages 147–157, 2021.

[42] Cataldo Musto, Christoph Trattner, Alain Starke, and Giovanni Semeraro. Towards a knowledge-aware food recommender system exploiting holistic user models. In *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization*, pages 333–337, 2020.

[43] Theodora Nanou, George Lekakos, and Konstantinos Fouskas. The effects of recommendations presentation on persuasion and satisfaction in a movie recommender system. *Multimedia systems*, 16(4):219–230, 2010.

[44] Jørgen Nyborg-Christensen. Nudging healthy choices in food search through front-of-pack nutrition labels. Master's thesis, The University of Bergen, 2021.

[45] Ishita Padhiar, Oshani Seneviratne, Shruthi Chari, Dan Gruen, and Deborah L McGuinness. Semantic modeling for food recommendation explanations. In *2021 IEEE 37th International Conference on Data Engineering Workshops (ICDEW)*, pages 13–19. IEEE, 2021.

[46] Florian Pecune, Lucile Callebert, and Stacy Marsella. A recommender system for healthy and personalized recipes recommendations. In *HealthRecSys@ RecSys*, pages 15–20, 2020.

[47] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70:153–163, 2017.

[48] Pearl Pu and Li Chen. Trust building with explanation interfaces. In *Proceedings of the 11th international conference on Intelligent user interfaces*, pages 93–100, 2006.

[49] Francesco Ricci, Lior Rokach, and Bracha Shapira. Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer, 2011.

[50] Frode Sørmo, Jörg Cassens, and Agnar Aamodt. Explanation in case-based reasoning–perspectives and goals. *Artificial Intelligence Review*, 24(2):109–143, 2005.

[51] Alain Starke, Edis Asotic, and Christoph Trattner. serving each user: Supporting different eating goals through a multi-list recommender interface. In *Fifteenth ACM Conference on Recommender Systems*, pages 124–132, 2021.

[52] Alain Starke, Martijn Willemsen, and Chris Snijders. Effective user interface designs to increase energy-efficient behavior in a rasch-based energy recommender system. In *Proceedings of the eleventh ACM conference on recommender systems*, pages 65–73, 2017.

[53] Martin Svrcek, Michal Kompan, and Maria Bielikova. Towards understandable personalized recommendations: Hybrid explanations. *Computer Science and Information Systems*, 16(1):179–203, 2019.

[54] Panagiotis Symeonidis, Alexandros Nanopoulos, and Yannis Manolopoulos. Moviexplain: a recommender system with explanations. In *Proceedings of the third ACM conference on Recommender systems*, pages 317–320, 2009.

[55] Mohsen Tavakol and Reg Dennick. Making sense of cronbach's alpha. *International journal of medical education*, 2:53, 2011.

[56] Nava Tintarev and Judith Masthoff. A survey of explanations in recommender systems. In *2007 IEEE 23rd international conference on data engineering workshop*, pages 801–810. IEEE, 2007.

[57] Nava Tintarev and Judith Masthoff. The effectiveness of personalized movie explanations: An experiment using commercial meta-data. In *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 204–213. Springer, 2008.

[58] Nava Tintarev and Judith Masthoff. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction*, 22(4):399–439, 2012.

[59] Raciel Yera Toledo, Ahmad A Alzahrani, and Luis Martinez. A food recommender system considering nutritional information and user preferences. *IEEE Access*, 7:96695–96711, 2019.

[60] Thi Ngoc Trang Tran, Müslüm Atas, Alexander Felfernig, and Martin Stettinger. An overview of recommender systems in the healthy food domain. *Journal of Intelligent Information Systems*, 50(3):501–526, 2018.

[61] Christoph Trattner and David Elsweiler. Food recommender systems: important contributions, challenges and future research directions. *arXiv preprint arXiv:1711.02760*, 2017.

[62] Christoph Trattner and David Elsweiler. Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proceedings of the 26th international conference on world wide web*, pages 489–498, 2017.

[63] Christoph Trattner, Dominik Moesslang, and David Elsweiler. On the predictability of the popularity of online recipes. *EPJ Data Science*, 7(1):1–39, 2018.

[64] L Nynke van der Laan and Oliwia Orcholska. Effects of digital just-in-time nudges on healthy food choice–a field experiment. *Food Quality and Preference*, page 104535, 2022.

[65] Jesse Vig, Shilad Sen, and John Riedl. Tagsplanations: explaining recommendations using tags. In *Proceedings of the 14th international conference on Intelligent user interfaces*, pages 47–56, 2009.

[66] WHO. Obesity and overweight. 2020. URL: `https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight`.

[67] Martijn C Willemsen, Bart P Knijnenburg, Mark P Graus, LC Velter-Bremmers, and Kai Fu. Using latent features diversification to reduce choice difficulty in recommendation lists. *RecSys*, 11(2011):14–20, 2011.