Multi-list Food Recommender Systems for Healthier Choices

Edis Asotic

Department of Information Science and Media Studies

University of Bergen

Supervisor: Prof. Dr Christoph Trattner

Co-Supervisor: Dr Alain D. Starke

Delivery Date: June 1, 2021

Abstract

Recipe websites are a popular destination for home cooks to discover new recipes and find what to cook. However, the most popular way of recommending recipes to users is trough similarity and popularity-based recommendations, which previous research has shown tend to be unhealthy. Building upon knowledge on how diverse sets of options increases satisfaction, this thesis investigates whether a multi-list recommender interface can support healthier food choices compared to traditional single-list interfaces, as well as increase choice satisfaction. As diverse set of options may introduce choice overload to users, explanations were investigated in terms of how they affect user evaluation with regards to choice difficulty, perceived diversity and understandability. A developed recommender system was used in a online study ($N = 366$), where users could select recipes from recommendations, as well as answering short questionnaires regarding their choices. The analysis showed that a multi-list recommender system was not able to support healthier food choices. However, users who interacted with the multi-list interface found it more satisfactory compared to single-list users. No significant evidence was found that explanations could mitigate choice difficulty. This thesis provides novel work on the utilization of multi-list recommender systems with explanations in the food recommender domain, which can further be expanded with considering other factors such including personalized recommendations in the multi-list interface.

Acknowledgments

I would first like to thank my supervisor Prof. Dr Christoph Trattner, for supervising and guiding this thesis, as well as setting up the research design. I would also like to thank my co-supervisor Dr. Alain Starke for all his help, especially regarding statistics and continuous feedback. My thanks also goes to family and friends, who have provided tremendous support during the work on this thesis.

Finally, i would like to thank the Faculty of Social Sciences at the University of Bergen, for their monetary support, as well as DARS Research Group for an additional grant which allowed for additional participants in the study.

Bergen, Norway, June 2021

Edis Asotic

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Recipe websites are a popular destination for home-cooks to find new recipes to cook, and as the COVID-19 pandemic that hit globally in 2020 imposed several restrictions, restaurants across the world were forced to close. Studies have reported an increase in home-cooking and shopping of ingredients for such activities [35, 8], both due to closed restaurants as well as people picking up a new hobby. This makes recipe websites even more actual to visit for a lot of people. One example of a popular recipe website site among the large quantity that exists is allrecipes.com, the world's largest food-focused social network [3]. On their page users can find a large amount of recipes generated by users, staff and brands. With a large quantity of options for their users, allrecipes.com uses recommender systems to recommend recipes to their users. Food recommender systems presents recipes to users based on both shared properties of recipes and based on what he or she liked in the past [54]. For example, users who indicate to favor a certain type of recipes, will be presented more recipes with similar ingredients in future sessions. Better user experience trough personalization can result in higher user retention for allrecipes.com. Recommender systems similar to this are used in a lot of domains to recommend items to users, most common applications are in movies, for example on streaming services such as Netflix, or in e-commerce such as Amazon and Ebay.

During the first year of the pandemic, researchers saw an additional increase in the intake of saturated fat [35], as well as increase in purchasing of ingredients linked to unhealthy food, such as pizza and pasta [8], making it important to focus on how to promote healthier recipes. Major diseases such as diabetes and cardiovascular diseases can be prevented trough adopting a healthier diet [9, 59], which in turn can be done trough proper home cooking [33]. However, adopting a healthy diet can be a daunting and complicated task. In an online setting, users struggle to identify healthy and unhealthy recipes due to lack of knowledge and misleading cues [16]. Current recipe websites often presents different healthiness metrics on recipes, but the informative value of these metrics largely depends on their knowledge level in regards to what constitutes a unhealthy or healthy recipe.

Food recommender systems on recipe websites can be a viable subject to research when it comes to help users change their eating habits, but these systems are currently ineffective in changing eating habits for the better, and can in some cases contribute to more unhealthy eating [16]. If a user seeks out a healthy recipe, similarity-based food recommender systems will recommend more similar, but unhealthier recipes - which is how most commercial food recommender systems behave currently. Several attempts on how food recommender systems can be used to offer healthier alternatives to users can be found in literature, such as systems that consider users health goals in recipe recommendations has been created and researched [18, 1, 16]. Said systems allow for users to input their health goals or dietary needs, in order to recommend recipes that take the goals and needs into account. Although several attempts have been made in regards to creating recommender systems which promotes healthy recipes, little research been conducted on how the presentation of these recommendations can affect the decision process when users need to chose what to eat [54]. Recommender system research is often concerned with algorithmic accuracy, while other factors that can influence user experience such as diversification in recommendations are often overlooked in research [28, 32].

Humans are subject to several decision-making biases during the process of making decisions. Choice overload is one example, which describes how too many options to choose from can make it more difficult to make a decision, and can also create unsatisfactory results after the choice has been made [25]. Even though recommender systems are proposed as a solution to mitigate such biases [2] they still exists in recommender systems [7], and remain relatively unexplored - especially in the domain of food recommender systems. Lastly, humans are subject to encounter the issue that our preferences are not always fixed in every decision making scenario. Research suggest that preferences are often constructive, meaning that we construct preferences "on the go", depending on the context. This is due to several factors, such as biases which makes it difficult to create preferences for many different situations, or that humans often has multiple goals in a decision making situation [6].

With regards to eating and food choices, humans can often have different goals at different times. Sometimes we may seek to eat what we want without special consideration, other times we focus on different goals such as weight loss, reduce risk of diseases or just reap the benefits of healthier eating. However, recommender systems in the food domain are often based on previous eating habits and food choices a user has made in the system. Therefore, their new goals and preferences may conflict with previous ones.

Similarity-based recommendations, which are often found on recipe websites will provide very similar recipes to the one that a user is currently looking at. Meaning that if a user is for example seeking out high caloric recipes, chances are they will be recommended more of the same recipes, with very slight variation. While this can be an effective way to provide very similar and satisfying alternatives to a user, it remains as an ineffective strategy if one seeks to change their eating habits. One solution to this problem can be to entirely exchange the set of recommendations completely with new and healthier alternatives, but this might leave the user with an unsatisfying experience with the recommender system since people have a general preference for unhealthy foods [16, 55], and doing this may leave the user with the impression of no satisfactory alternatives.

1.1 Problem

As the current state of food recommender systems often are single-list interfaces, they can only account for one factor at a time when recommending recipes - factors such as either recommending healthier alternatives of recipes, or similarity-based recommendations to cater towards current or previous preferences. This entails that recommending healthier alternatives may lead to dissatisfaction for users, as they can be perceived as not similar to their preferences. To address the issue explained, this thesis investigates a multi-list approach for recommending food recipes to users. An example of a multi-list interface is presented in figure 1.1, which is a partial screenshot taken from the streaming service HBO Nordic. Multi-list interfaces consists of providing recommendations to users in several lists at once, where each list can optimize for a specific factor or theme. Looking at other domains, multi-list interfaces have been utilized by Netflix for recommending movies, where the interface consists of up to 40 lists each with a specific theme. Themes can vary from personalized lists where movies

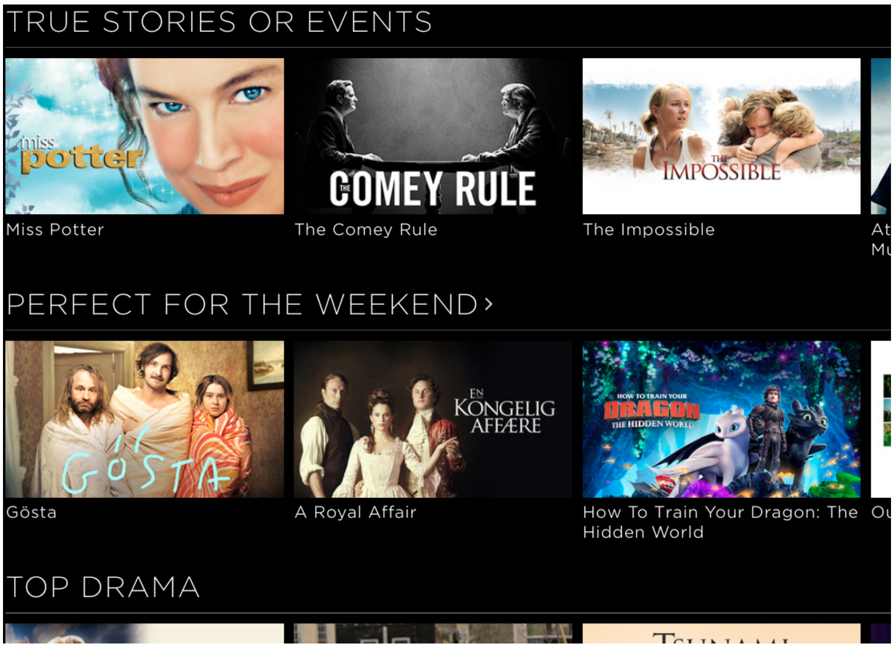


Figure 1.1: Partial screenshot of a multi-list recommender interface utilized by HBO Nordic. Depicted are three different lists, where each emphasizes a particular theme. The theme is explained trough short explanations.

are recommended based on previous likings, to showing movies in other categories [19]. Such interfaces allow for both catering to previous preferences, while also recommending slightly dissimilar alternatives that could still be appreciated by users. Furthermore, multi-list interfaces often presents explanations in conjunction with a recommended list, briefly explaining the content. Usage of explanations in multi-list recommender interfaces is thus not novel, as it has been used in other domains such as Netflix for movies and Amazon in eCommerce. However, the effects of explanations in a multi-list context has yet to be researched and tested.

In a food recommender system context, it is expected that presenting users recipes in a multi-list interface can lead to users selecting healthier recipes, while their choice satisfaction remain positive, or even increased. Previous studies have found a relation between increased set size/diversity and choice satisfaction [25], as well as increased satisfaction when recommendations are organized [11, 38]. A multi-list recommender system allows for both increasing diversity of recommendations, as well as organization. By utilizing a multi-list recommender system, several algorithms can be used to optimize for different factors, and thus provide a more diverse recommendation set for the user. To account for users preferences, allowing for similar item retrieval can be done trough one list in the interface [46], while other lists can optimize for similar but healthier alternatives such as low fat options or alternatives with less calories. By having several algorithms that creates diversity in a multi-list interface, the hypothesis is that the chances of a user making a healthier food choice will increase, while affecting their satisfaction in a positive way. Multi-lists interfaces in recommender systems have been researched to some extent earlier, such as the work Chen & Pu [11] where they analyzes eye tracking movement of their subjects when looking at recommendations in a single list interface, vs an interface with a category structure. However this type of organizing recommendations is merely categorizing the recommendations, while the approach taken in this thesis uses different algorithms for each list.

In addition to use multi-lists, the study also investigates explanations of recommendations by accompanying each list with explanatory labels that describe the content of the recommendation set. Explanations in a recommender system can increase transparency, and expose the reasoning behind why something is recommended [51]. The aim with explanatory labels in conjunction with multi-list recommendations in a food recommender system is to increase the appeal of less similar, however healthier items. For example, the lists can be explained by using health metrics such as "recipes with less fat" or "Alternatives with fewer calories".

1.2 Research Questions

Based on the defined problem above, following research questions are raised:

- **RQ1:** To what extent can a multi-list food recommender interface support healthy food choices, compared to a single list interface?
  - **RQ1.1:** To which extent do multi-list interfaces affect how users perceive and evaluate a food recommender system compared to a single-list interface?
- **RQ2:** To what extent do recommendation explanations support healthier food choices?
  - **RQ2.1**: To what extent do explanations decrease choice difficulty and increase perception of diversity and understandability with regards to choosing a recipe from a recommender system?

1.3 Thesis Outline

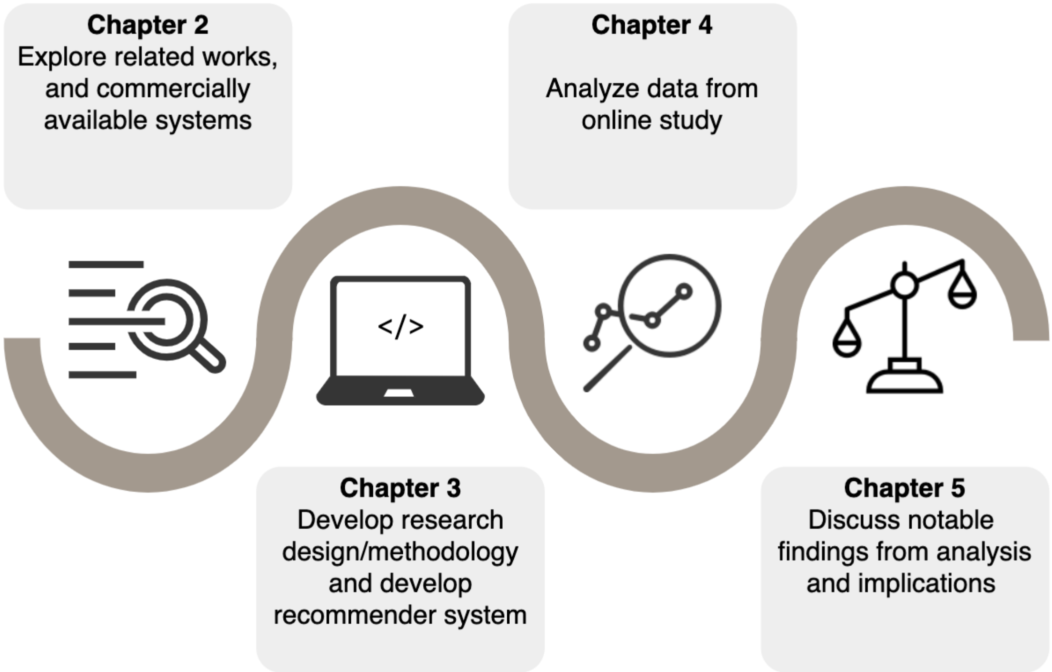


Figure 1.2: General overview of the thesis structure. Depicted are the remaining chapters in this thesis, with explanations on the work conducted for each one.

The remainder of this thesis is structure into four chapters:

- Chapter 2 Provides a thorough review of related works in food recommender systems, nudging and choice architecture, as well as explanations and choice overload.
- Chapter 3 Describes the materials for the experiment, regarding the dataset and developed prototype. Furthermore, this chapter describes the procedure and research design of the experiment, as well as measures and descriptive statistics from the experiment.
- Chapter 4 Details all results from the statistical analysis conducted in order to answer the research questions
- Chapter 5 Provides a discussion of the results, summary, limitations and future work.

Chapter 2

Background

This section presents a general introduction to recommender systems, describing different types of algorithmic approaches that are being used, as well as examples of health aware food recommender systems. Furthermore, a presentation of literature on how explanations are utilized in recommender systems is done, as several multi-list recommender interfaces utilizes explanations for explaining the theme or content of a particular list. This section also looks at how nutritional values can be used to explain healthiness of recipes, and also how they can be used as explanations in a multi-list recommender system. The concept of nudging and choice architecture is introduced, explaining how positioning and addition of items can persuade users towards different choices. Finally, a summary is presented, with an overview of how this thesis will address shortcomings in previous literature.

2.1 Recommender Systems

Recommender systems are often employed to help users find what they need, whether it is movies on Netflix, books or other products on amazon.com or a new recipe to cook for dinner. These systems normally recommends "items" of a specific type, such as a food recipe. The systems graphical interface and algorithms generates recommendations that are customized to provide suggestions for the specific type of item to the user. Various forms of recommender systems exists, where *collaborative filtering* systems are most popular. This implementation recommends items based on knowledge from the 'crowd'. Each user rates some specific items, then the system is able to recommend items to the user based on other users ratings [43]. A commonly used recommender library in recommender system literature is the MovieLens recommender system, which recommends movies to users that matches their preferences. The recommendations are based on collaborative filtering of other members movie ratings and reviews [21]. Another implementation is called *content based recommender system* and allows recommendations by using the existing information about items. This information can be genre about a movie, or nutrients in a particular food item. These kinds of recommender systems utilize information about a users preference and recommends items based on the item-specific information [41]. *Similarity-based* recommender systems are often used in the food domain, where the systems recommends items similar to a reference item, such as when a user is looking at a recipe, recipes with similar properties such as ingredients or cooking directions will be recommended. A last kind of relevant recommender system in the food recommendation domain is called *knowledge based recommender system* [41, 37]. Knowledge based recommender systems are based on explicit knowledge about the items in the recommender systems, the users preferences and also the recommender criteria. This approach is useful when users have specific requirements, for example different health goals or health issues that needs to be considered when recommending food items.

Typical usage of recommender systems on recipe websites are similarity-based, meaning that the recommended recipes share many of the properties as the reference recipe, which is evident in research on food-recommender systems such as in [56]. These properties are often ingredients, cooking directions or type of cuisine the recipe belongs to. Research of Trattner & Jannach [56] tried several similarity metrics for recipes; including titles, images, ingredients and directions, showing that recommendations based on of of these factors, or all combined leads to the highest perceived similarity for users, while title - based similarity performed particularly well.

Depicted in Figure 2.1 below is an example of a recommender system on allrecipes.com. In this particular example the reference recipe was for fried chicken, and the yielded recommendations are similar fried chicken recipes. While this system is efficient for providing very similar alternatives, it does not create a diverse set of recommendations and is therefore ineffective in changing users eating behavior.

A commonly found section on recipe websites is recommending popular recipes, based on other users ratings, likes, and click history, an example of this is shown in figure 2.2 from seriouseats.com. This type of recommendation technique can be an effective way to provide

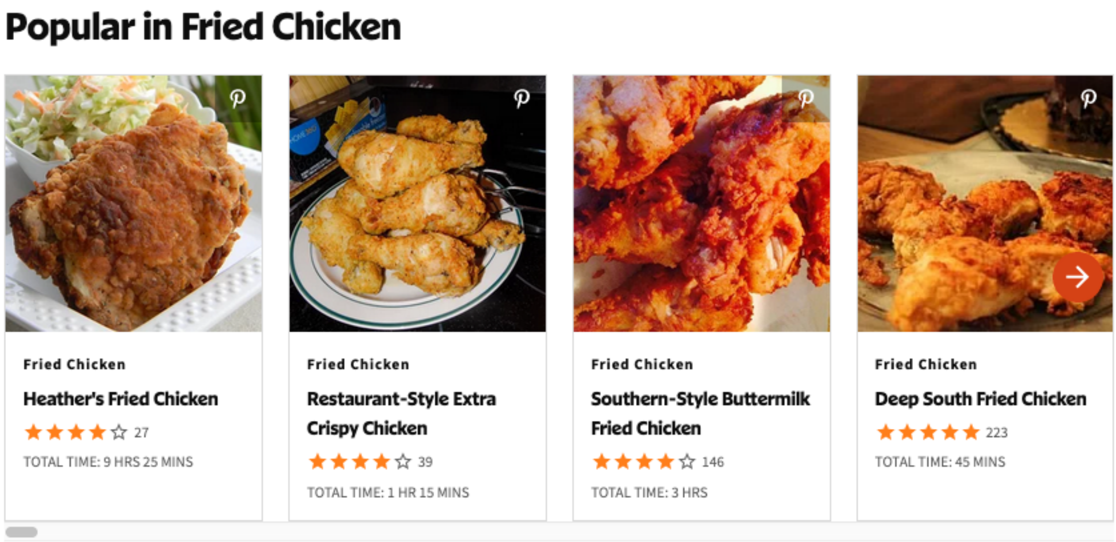


Figure 2.1: Example of similarity-based recommendations from allrecipes.com, showing several similar fried chicken recipes after a user has found a particular fried chicken recipe. The recommendation list is presented under the recipe the user currently looking at.

satisfactory recommendations to users as neighbour based recommendations tend to be popular with users [24]. This could however lead to unhealthy eating patterns as findings from research show that users often have a general preference for unhealthy foods.

2.2 Considering Health When Recommending Recipes

Research shows that recipes found online are generally unhealthy. Trattner & Elsweiler [55] investigated the health of online sourced recipes, and their findings show that only a fraction of recipes on popular sites such as allrecipes.com can be considered healthy according to standards set by WHO and Food Standards Agency (FSA). Furthermore, their investigations show that state of the art recommender systems also tend to produce unhealthy recommendations.

In literature on food recommender systems, we can find several attempts to solve the problem of unhealthy recommendations, and we can see how health can be considered and incorporated in food recommender systems. The aspect of health in food-recommender system literature has been considered mainly in two different ways; one way being generally providing healthier recipes to users that match their preferences, and the other possibility is providing attention to users health goals. Such goals can be dietary restrictions due to diseases or illnesses,

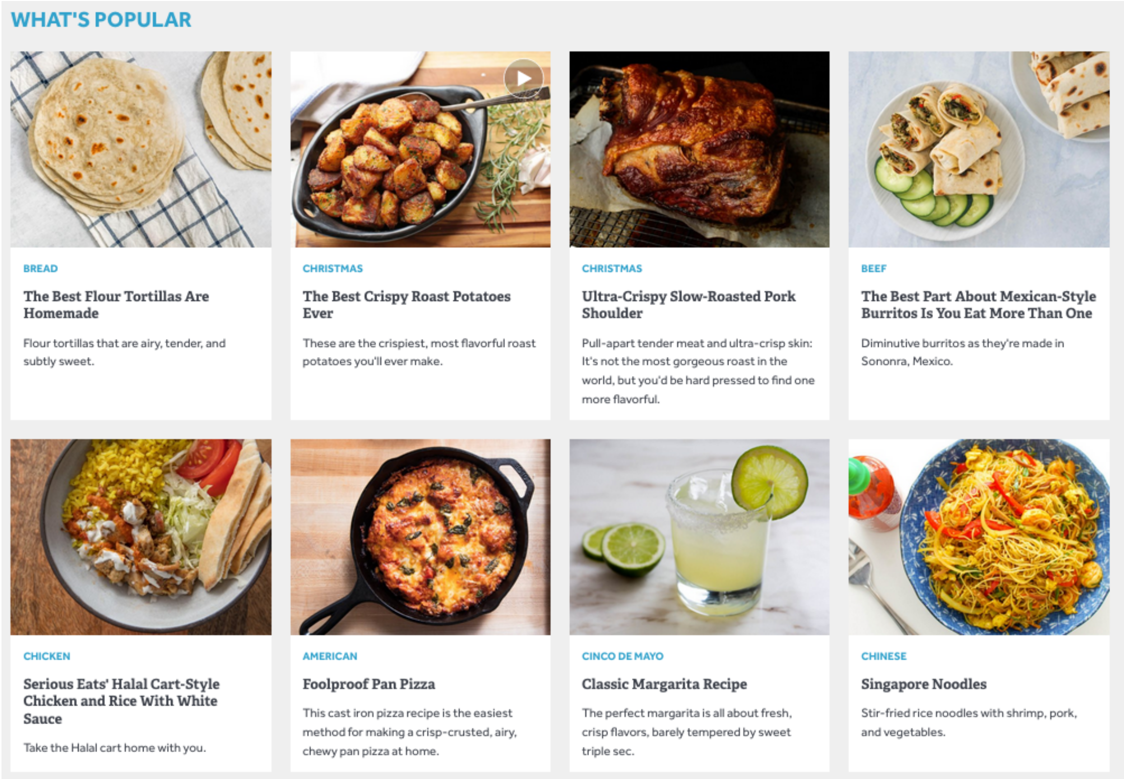


Figure 2.2: Example of popularity-based recommendations on seriouseats.com, depicted are recommendations for recipes that are popular with other visitors of the recipe-site.

or health goals in regards to loosing or gaining weight.

Agapito et al. [1] proposes a web-based system called DIETOS, which is to improve quality of life of both healthy people, but also individuals with chronical diseases. The proposed system gives personalized recommendations for food, according to the users health profile. Ge et al. [18] presents a mobile application that recommends set of recipes to users created on the basis of users preferences on ingredients and their daily calorie goals. A web system is proposed that can automatically provide entire meal plans that takes the users taste and health goals into account. By using data on age, gender and height, the system can calculate nutritional requirements for a user. Similar to DIETOS, the users rank a number of recipes in the system which builds their taste profile. In addition, the users also have the ability to directly chose some meals, and have the system recommend the rest of the meal plan. By suggesting entire meal plans the system makes it easier to eat healthy, as people often have a lack of knowledge on nutrition. Harvey & Elsweiler [22] also demonstrates how recommender systems can be used to recommend entire meal plans as opposed to only recommending recipes. Toledo et al. [60] presents a general framework for daily meal plan recommendations. The proposed framework is original in the sense that it manages user preferences and nutritional information in recommendations, which their literature review shows that most other tools do not do both.

Elsweiler et al. [16] explores the feasibility of substituting meals that would typically be recommended to users, with similar but healthier dishes. Their research shows that people have a difficult time identifying unhealthy recipes (recipes with higher fat content). This struggle is due to lack of information from the recipes, lack of previous knowledge an also misleading cues such as images. Building upon this, their research show that it is possible to replace recipes with healthier but similar, or better rated alternatives. The study shows that people have a implicit preference for fat dense foods as participants cannot tell the difference when asked, but typically select fattier recipe according to their preference. How the participants percept fat content can be influenced by the information available. Misleading cues such as images or title can bias and result in false impression of healthiness of the recipe. These biases can in return be exploited to nudge people towards choosing less fattier options of recipes.

Looking at the algorithmic approaches to healthy recipe recommendations, the work of Elsweiler et al. [16] consists of retrieving similar items to a reference recipe, based on one or more similarity factors. Upon retrieving similar items, post - filtering methods can be applied to filter out healthy recipes based on a specific metric, such as fat content. The algorithm found in [18] considers a users list of preferences and ratings on previous recommendations, while also accounting for the amount of calories in the recipes and the amount the user needs to meet the daily goal. The automated meal plan system found in [22] calculates a daily calorie requirement based on the estimated metabolic rate for the users, and then generates a set of recipes that match their taste profile, which also meets the users daily nutritional requirements. An interesting approach is proposed by Gorbonos et al. [20], Which takes user preferences as input in the form of an ingredient list, and then creates a pseduo-recipe with a healthy baseline. Furthermore, the system searches for actual recipes in a dataset that are similar to the created pseduo-recipe, thus creating a healthy-recipe first as a baseline and then find similar alternatives to that recipe.

Cataldo et al. [37] considers a knowledge-based recommender approach, where knowledge about the user as well as recipes is used to provide recipes to the user. For example, by knowing a users BMI, the system ranks recipes according to the BMI. If the BMI is over 25, the system applies a modifier which ranks low-calorie recipes higher, and if the value is under 25, then the system ranks higher-calorie recipes higher. Other factors such as mood, activity level, and health (i.e stress, sleep, depression) is also taken into account, and relevant modifiers are applied based on knowledge on the correlation between a particular user aspect and food items. Studies suggest a link between stress and the content of salt in recipes, so the system will recommend recipes to users with sodium levels for users who report high stress levels.

2.3 Nudging and Choice Architecture

The study in this thesis is concerned about providing similar but healthier recipe alternatives without explicitly removing unhealthy alternatives trough multi-list interfaces. This is a concept of both nudging and choice architecture. A common definition of a nudge is any aspect of the choice architecture that alters peoples behavior in a predictable way but without forbidding any options or significantly changing their economic incentives [49]. The concept of choice architecture is concerned with how the organization of choices in a context can affect a persons decision. A typical example of how choice architecture is present in our daily lives is the placement of products in supermarket checkout, which caters to humans impulsiveness when shopping. Promoting specific products in this area can be highly viable, as the checkout register is one spot in a supermarket that every customer has to visit, and are held captive until it is their turn [5].

The persons who are in charge of how alternatives are presented in a choice scenario (i.e what items to place in a supermarket checkout) are often called 'choice architects', which can alter choice architectures to create nudges. Since products found at the checkout in a supermarket are often unhealthy items such as chocolate bars, or other sweets [14], researchers have conducted studies on how choice architectures can be altered to make consumers make healthier purchases at checkout, or in other words *nudge* them to do so. Kroose et al. [30] show that by adding healthier alternatives to the checkout area, and still keeping the unhealthy products the consumers were successfully nudged into making healthier food purchases at checkout. By keeping both options at the same place, the customers are not taken away any options. Similar conclusions were found in a longer study by Van Gestel et al. [57] where they tried re-positioning healthy items to the checkout area, and providing unhealthier items other places in the store. Even though the items are removed from the checkout area, the customers are still free to seek them out in other parts of the store, thus this is also not breaking the nudge definition. Similarly to real life situations, placement of items in recommender systems is also important. The way that the recommendations are presented to the user affects how the user makes decisions or choices. Studies show that users pay more attention to first few items in a list, than items lower down [7]. Similar observations are made in grid layouts, where users pay more attention to top left-items.

Johnson et al. [27] provides an overview of several tools or techniques that can be used by choice architects. These tools are divided in two categories; tools used for structuring the choice task, and tools used for describing the choice options. Some of the techniques that can be used is for example reducing the amount of alternatives to reduce the choice overload, or make limited time windows, often employed in for example gift card settings. Another tool that is described is the use of attributions for describing or explaining different choice options. People often make choices by weighing pros and cons of attributes on different alternatives, or choices are often made by looking at attributes and predicting how satisfied we will be in the future with a particular alternative. Recommender systems often utilize explanations of recommendations to describe what has been recommended to users, and these explanations can be formed based on the attributes of items in the recommendation set.

2.4 Explanations in Recommender Systems

Explanations in recommender systems can yield several advantages. Tintarev & Masthoff [50] show seven advantages of providing explanations in a recommender system; by providing explanations, the system can be transparent to its users (1), allowing scrutability (2) and also increase trust (3). Further they can be effective to help users make a decision (4), and increase the efficiency (5) thus helping users make decisions faster. This can help increasing the overall satisfaction (6) while using a recommender system. Furthermore, a benefit for the system is that explanations can also be persuasive (7), and convince them to try or buy something. Explanations persuading users to make a choice can be considered as a nudge, by framing a specific item in a particular way. I.e in a food recommender setting where the systems goal is to make users eat healthy, persuasive explanations can be beneficial for the user, by highlighting particular abilities about a recipe, making it more attractive to the user.

A lot of examples on explanations in recommender systems exist in various domains, both in research and commercial settings. Herlocker et al. [24] conducted one of the first major works on evaluating explanations in recommender interfaces. Studies on twenty-one implementations of explanation interfaces, by using the MovieLens system. The study concludes that best performing explanations are based on the ratings of neighbours, and histogram presentation performed better than a table presentation. Since the movie domain is the most popular application of recommender systems, it is natural that a lot of research literature on explanations is concentrated around this domain. The work of Symendiondis et al. [48] from 2009 show an implementation of a movie recommender system called "MoviExplain" which provides explanations for why a particular movie was recommended. The explanations are formed based on similar movie features between previous rated movies and the recommended movie. With regards to multi-list recommender systems, explanations are utilized by netflix in their multi-list interface, which labels each list presented to the user, according to the theme of the list [19], thus explaining the content as well as providing separation between the lists.

In food recommender systems, research on explanations is however scarce, according to the state-of-the-art summary by Trattner & Elsweiler [54]. Some implementations of explanations in food recommender systems is the work of Elahi et al. [15] which presents a prototype of interactive food recommender systems for groups in planning their meals. The users can add tags on what kind of ingredients or type of food they do or do not want. The system then recommends a recipe, and provides a explanation that informs the user *why* the particular recipe was recommended. The work of Harvey & Elsweiler [22], mentioned in earlier sections also includes explanations in their meal plan recommender system, where each recommended meal plan is accompanied with nutritional values, and how far off these are from the users ideal values. Similar to this, Leipold et al. [31] created an mobile application for logging daily nutrients, which also recommends new recipes based on previously consumed nutrients. Each recommendation includes nutritional values, and explains how much the recipe fills the daily macro nutrient goals of the user.

Cataldo et al. [36] explored natural language explanations/justifications on food recommendations. Based on both user characteristics, recipe features and domain knowledge, the recommender system produced natural language explanations which both informed users about healthier choices as well as being personalized. Some examples of user characteristics were BMI, cooking skill level, or health goals, while recipe features were nutritional values, recipe difficulty or ingredients. A combination of these features could produce either single style justifications such as “Vegetable soup has 462 calories, please consider it since your goal is to lose weight”, or comparative style which compared two different recipes; “Vegetable Soup is easier to prepare than spaghetti cacio and Pepper. They could be more adequate to your cooking skills, which are low”. Their findings show that the comparative style of justifications were more effective with regards to healthier food choices.

2.5 Designing explanations

When utilizing explanations in recommender systems, it is important to think trough what types of explanations to include, and what can be considered a good explanation. Tintarev & Masthoff [51] provide guidelines on how to design a "good" explanation, and how to evaluate this. In short, these guidelines tells us that we should consider wanted benefits from using explanations (1), be aware that the evaluation of the explanations are related and confounded with the functioning of the recommendation engine (2), We should think about how the presentation of recommendations and interaction with them affect each other and the explanations (3), and at last we should consider the relationship between the algorithm and the type of explanations we chose to generate (4).

In a lot of countries, food items we buy and eat every day are required to disclose the ingredients in the product, and also nutritional content, such as amount of calories, fat, carbohydrates, sugar, fiber etc. [17]. The ingredient list and the nutritional content on products provides explanations on how healthy or unhealthy a particular food item is, for example a calorie dense or high caloric product is most often classified as an unhealthy food item. The utilization of nutritional information in recommender systems is a possible way to disclose the healthiness of a particular recipe, in a similar way. Examples of this is shown in Harvey & Elsweiler [22], where they disclose the nutritional values such as calories, carbohydrates and sugar of a recommended meal plan, and how far off this is from ideal values that the particular user should consume. The popular recipe site allrecipes.com also provides nutritional values on their recipes, as seen in figure 2.3, which in turn can help identify if a recipe suits their goals or needs, and thus assist them in the decision process of whether to cook a recipe or not.

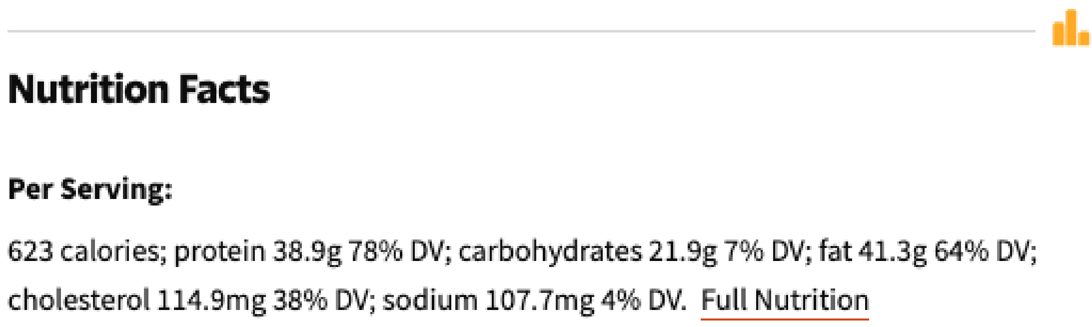


Figure 2.3: Screenshot of nutrition facts displayed at the bottom of a recipe on allrecipes.com, informing the user about nutritional content of the particular recipe.

From literature in other fields, we see several examples of studies that show how people use information on nutritional content found on product packaging, to make an informed choice about their food purchases. Studies done by Kretuer & Brennan [29] show that participants that have healthier diets report that they read the amount of nutritional content on products more often than participants with unhealthier diets. Also shown in their studies is that participants who has diet related health problems use these values to identify which foods they can or cannot consume. Similar findings where also found in a study conducted by Nayga et al. [39] in 1998, where results show that participants on a special diet were more likely to read the disclosed values on products while shopping.

It is important to acknowledge that nutrition label reading can be difficult as well. Shine et al. [45] identifies in their research that some respondents have difficulties understanding the labels, and some also pointed out lack of time as a reason to not read nutrition labels on products, which may indicate that people find it time consuming. Rothman et al. [42] also identifies that people in general tend to struggle with nutritional labels on products, especially with calculating serving sizes and nutrients in a particular serving. For example, respondents struggled with calculating how much carbohydrates a half bagel consisted. Strong correlations were found between the struggle of label reading and low numeracy and literacy. The authors also identifies that those who had less struggles had high income and high education backgrounds. Based on this, consideration has to be made in regards to how to incorporate such information in for example recommender systems, as people may have different expertise and experience with it.

In addition to explanations, an important aspect Johnson et al. [27] addresses is the amount of choices or options to present. A choice architect needs to balance between two criteria. Presenting more options can be beneficial for the user as it increases their chances for a preference match, but it also increases the cognitive load since the user needs to evaluate more options. The answer for what the right balance is dependant on personal characteristics of the decision maker, meaning that it may not exist one-fits-all solution. For example, older adults prefers less options than younger adults, as they have less processing capacity. A series of studies by Schwartz et al. [44] show that different types of people need different set sizes. Maximizers who seek to maximize their outcome prefer smaller set sizes as these users search for the best option, and therefore larger set size equals more time spent on searching. Satisficers however do not always consider all options, and stops after finding a "good enough" solution. In other words, considering the set size of recommendations is important, especially in avoiding choice overload.

2.6 Choice Overload

Choice overload is a term that explains the problem where too many choices can give humans difficulties in deciding. Iyengar & Lepper [25] provided one of the first works that examines the possibility of the choice overload hypothesis by conducting three studies in both field and laboratory settings. Their work demonstrate a contradiction to the notion of 'having more choices is better', as their studies described above show that having more choices might appear desirable at first sight, but it can have consequences with regards to choice difficulty, and leave users with less satisfaction with their choice. Diehl and Poynor [13] also show how larger choice sets affect satisfaction in a decision scenario. When a person is presented with a large assortment, their expectations also rise in regards to finding better items. Higher expectations can then lead to a stronger disappointment if the large assortment does not provide a suitable match for their expectations.

Some examples of directly addressing choice overload can be found in recommender systems literature. Bollen et al. [7] has conducted experiments on this matter to see if recommender systems are prone to choice overload. The authors investigate the effect of different set sizes (5 vs 20 items) and also between low and high quality on perceived variety, attractiveness of the recommendation set, choice difficulty and satisfaction with chosen item. Their findings are that larger sets with only good movies does not equal higher choice satisfaction in comparison to smaller set sizes. The increased attractiveness of the set gets counteracted by the increase of choice difficulty. In a study from 2012, Cremonesi et al. [12] show how personalized recommendation can lead to information overload since it stimulates the user to seek more alternatives before deciding on their final choice.

2.7 Multi-list Interfaces in Recommender systems

As an effort to reduce choice overload in recommender systems, several researchers have studied organization based or 'multi-list' interfaces. Pu and Chen [40] compared a single list interface of recommendations with simple explanations to a multi list interface, where each list had explanations on what items the particular list contained. The interfaces were evaluated by prompting users to select a product from a recommendation set which they would purchase if given the opportunity. Furthermore users opinions were captured on a 6-item questionnaire, inquiring on overall opinions with the interface. Three constructs were made based on the questionnaire, where the authors looked at *percieved competence, cognitive effort and intention to return*. Their results show that a multi-list interface was perceived more helpful, as it allowed for easier comparison between items and induced less cognitive strain, even if time spent on making a decision was equal between the two interfaces. The study also shows that users had a higher intention to return to the multi-list interface, and on average built more trust to this presentation form.

As mentioned in the introduction, eye tracking studies conducted by the same authors [11] investigated attractiveness of organized vs non-organized interfaces, as well as user satisfaction in a choice context where the main task was to find a product the participant would purchase if given the opportunity. Saving the particular item meant that the user found it satisfactory, as opposed to quitting the experiment if no satisfactory items were found. Their results show that interfaces organized in lists by category were more attractive than single list interfaces. Multi-list interfaces are often accompanied with short explanations on what each list contain, or the difference between them. For example in this study each list was accompanied with titles describing the attributes of the products in each list.

Netflix is an example of a commercial service with a recommender system that implements a multi-list interface. Gomez-Uribe & Hunt [19] describes the layout of the homepage on Netflix and underlying algorithms, which consists of approximately 40 different lists, and up to 75 videos per list. Each list has recommendations that are generated by one single algorithm. Each row is also labeled with explanations on what type of movies each list contains to make the recommendations more transparent and intuitive. A screenshot of the multi-list interface which Netflix uses is shown in figure 2.4. However, the paper does not include any studies on the effects of the multi-list interfaces versus other types such as single-list ones, or the effects of explanations.

Jannach et al. [26] explored the effects of multi-list interfaces for similar-item recommendations, by comparing a long single-list with linebreaks to a multi-list interface separated by labels. The results from their study showed that single-list interfaces allowed for less effort when making a choice, compared to a multi-list interface which slowed down the decision process. However, multi-list interfaces allowed for more exploration, and left the users with an impression of more diverse and novel recommendations.

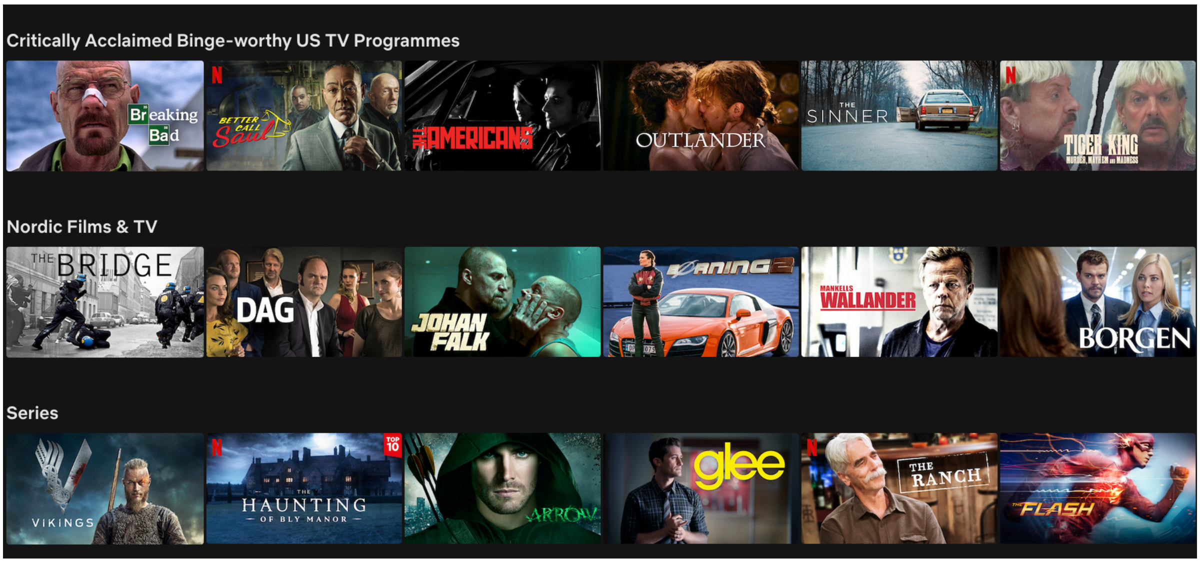


Figure 2.4: Screenshot of Netflix landing page, utilizing explanations in a multi-list interface which explains the theme or content of each presented list. Depicted are three lists of recommendations, which each optimize for different factors, such as Nordic films & TV.

Similar interfaces are found in e-commerce sites such as amazon, such as demonstrated in figure 2. Amazon provides several lists, with explanations formed on the basis of item similarity on items you have previously bought, to neighbour similarity showing which items you have bought that customers often buy again. As shown in Herlocker et al. [24], neighbor similarity explanations are proven to be effective.

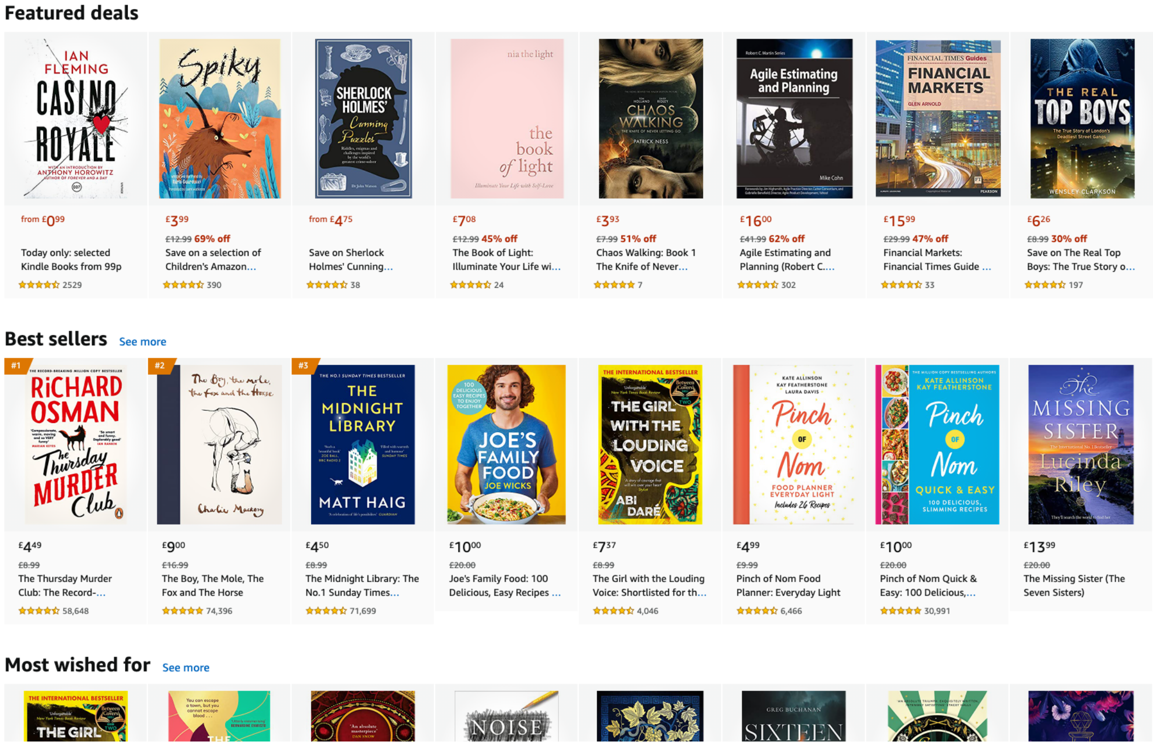


Figure 2.5: Screenshot of Amazon's interface when browsing for books. Depicted is how Amazon utilizes multi-list interface with explanations to describe the content of each recommended list.

Nanou et al. [38] studied what effects a structured recommender interface (grouped by genre) has on persuasion and satisfaction vs a unstructured representation. Their results show that the structured representation affected the systems persuasive ability positively. Users also reported more ease of use while browsing recommendations in the organization based interface. While their studies on satisfaction could not be fully concluded, it was reasonable to assume that the participants found the organization based interface to be more satisfying to use, as it was easier for 75% of the users, and the cognitive load was reduced in the decision making process. Although similar studies were conducted by Chen & Pu [11, 40], the work of Nanou et al. is done in the movie domain, where subjective opinions are more prevalent whereas Chen & Pu examined recommendations of items which can be more objectively measured and evaluated.

2.8 Summary, Differences and Contributions

In this section several works have been presented regarding food recommender systems, choice overload in recommender systems and how existing research have addressed this problem. Also included is how explanations can be utilized in recommender systems. However, several key factors can be identified which previous research has not addressed in regards to these topics. Multiple works on health-aware food recommender systems have been published, while none are concerned with the presentation of the recommendations, and how this affects the healthiness of a particular choice as well as no attention has been paid to multi-list recommender systems in the food domain. Explanations have been investigated in multiple domains, including food recommender systems. Novel work have also explored explanations and what effect they have on food choices. However, effects of explanations in multi-list interfaces have yet to be explored. This thesis will address the mentioned shortcomings by doing the following:

- A recommender system will be developed, which will utilize five different algorithms for recommending recipe alternatives. One will purely optimize for similar items, while remaining four will optimize on four health factors. The five algorithms will be used each on their on in single-list conditions, and all at once for multi-list conditions.
- Explanations for each list will be generated, which will have basis in what the particular list is containing, based on which factor it optimizes for - whether it is similar recommendations, or healthier recommendations.
- The developed recommender system will be used in a only study hosted on crowd sourcing platforms, where participants can interact with the system, and their responses will be analyzed based on healthiness of their choices as well as several different evaluation factors.

Chapter 3

Methodology

The study conducted in this thesis investigated how a multi-list interface performed in regards to nudging users towards healthier food choices compared to a single-list. In the multi-list interface, each list represented one of five implemented algorithms, where four of them optimized for different health factors, and the last provided similar recipes based on computed similarity. The health factors the algorithm optimizes for are; amount of fat, fiber, carbohydrates and calories. These four factors among others are used by Food Standards Agency (FSA) in their 'traffic light' food labeling system, which is in place to more easily distinguish healthy foods vs. unhealthy foods. The different amounts of these nutrients in food determines the healthiness of a recipe or particular food item [10].

Also tested in the study was how explanations supported the participants in making healthier food choices, and what type of explanations were most effective. The explanations served as aids to help the user understand what makes the recommendations healthier and help them make a healthy choice.

The approach for the study consisted of conducting a online user experiment on the survey platform Prolific, as well as the crowdsourcing platform Amazon Mechanical Turk. To conduct the experiment, a recommender system was developed, which served participants recommendations of food recipes while using several algorithms that optimize for different health factors of the recipes.

3.1 Dataset

For recommending recipes the developed prototype uses recipes from an Allrecipes.com dataset, which is composed of over 58,000 recipes and used in several other studies in the food recommender system domain [54, 55, 56, 53]. A smaller subset of 935 recipes was sampled, trough randomized queries, across five different categories: Casseroles, Roasts, Salads, Pasta and Chicken dishes. The same sampling process was applied for the reference recipes, while also excluding the recipes that are used in the recommendation set. In total the recommender system prototype used a pool of 28 randomly selected recipes from the same five categories listed above.

3.2 Prototype of Recommender System

For conducting the experiment, a prototype of a recommender system was developed consisting of several parts and technologies. The majority of the front-end interface was developed from the ground up, using standard web-development technologies: HTML, JavaScript, CSS with Bootstrap and PHP. Some aspects regarding data saving was incorporated from Mouselab WEB[1], The participant interacted with a front-end interface which allowed them to browse recommendations of recipes, read short descriptions, view images and finally select the recipe that appealed to them most.

Further interaction was trough answering short questionnaires after each selection task, and also answering longer questionnaires in the introduction, as well as at the end. In total, the the application presented users with a consent form, introduction survey, five tasks and a finishing survey that addressed overall satisfaction and choice difficulty. For each task the prototype presented the user with a reference recipe consisting of an image, title and cooking directions. The reference recipe was randomly selected from a pool of 28 recipes. Below the reference recipe the system provided the participant with either 5 or 25 recommendations, depending on which condition the participant was assigned. A partial screenshot is provided below in figure 3.1, of the recommender interface. In this figure the structure with reference recipe, recommendations and questionnaire is presented.

[1] A tool for creating experiments that monitor decision makers [34].

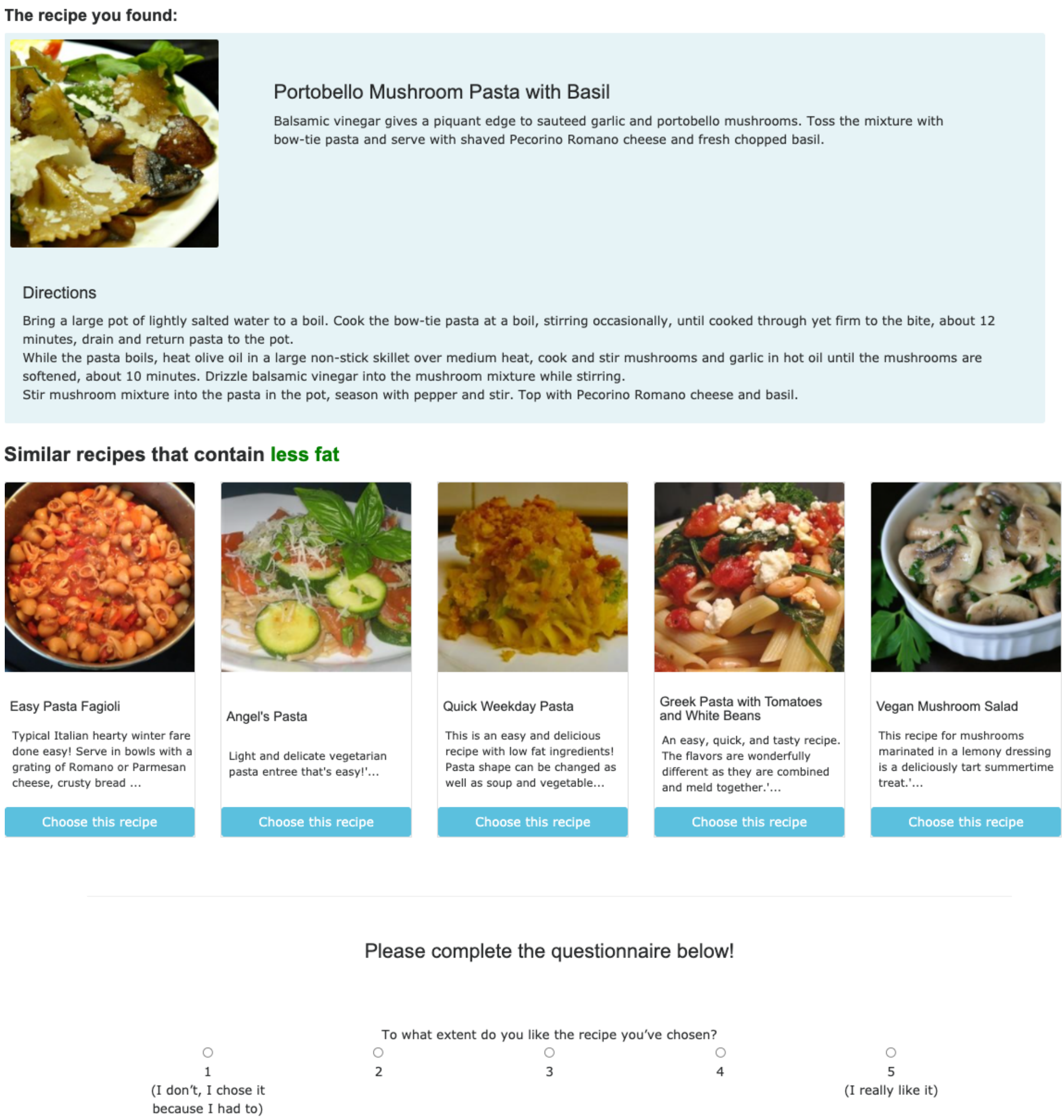


Figure 3.1: Partial screenshot of the recommender system. Depicted here is the single-list condition, with explanations, where five recipes are presented which are similar to the reference recipe, but contain less fat.

The back-end retrieved recipes from the compiled dataset, which were indexed by the the PHP framework *Zend Lucene*, a framework that has previously been used for querying similar recipes [47]. The framework allowed for indexation of all the recipes in the recommendation set, which was in turn utilized for retrieving similar recipes. The approach of retrieving similar recipes is formally represented below:

$$rec_k(r_i) = \underset{r_j \in R \setminus r_i}{\overset{k}{\operatorname{argmax}}} \{sim(r_i, r_j)\}, \tag{3.1}$$

Where $R \setminus r_i$ represents all recipes in the recommendation set where the reference recipe r_i is excluded, and $sim(r_i, r_j)$ is a similarity function. As shown in the work of Trattner & Jannach [56], similarity functions based on recipe titles performs particularly well in terms of how users perceive similarity between recipes. Based on the reference recipe title that was retrieved from the front end, Zend Lucene gathers similar items from the recommendation set by calculating the Term Frequency-Inverse Document Frequency (TD-ITDF) score. Upon retrieving a list of similar recipes based on the calculated scores, the back-end applied post-filtering, which sorted the recipes according to which list the recipes was requested for, and the top-5 recipes were selected.

In example, if the retrieval was for the "similar alternatives with less calories" list, the back-end first retrieved all similar recipes trough the approach explained above, and then sorted the list of similar recipes based on the amount of calories the recipe contained. For the "Similar recipes" list, no post-filtering was conducted as *Zend Lucene* returned a sorted list based on similarity scores. Multi-list interfaces made five requests simultaneously for a reference recipe, as the interface consisted of all lists each time.

3.3 Procedure

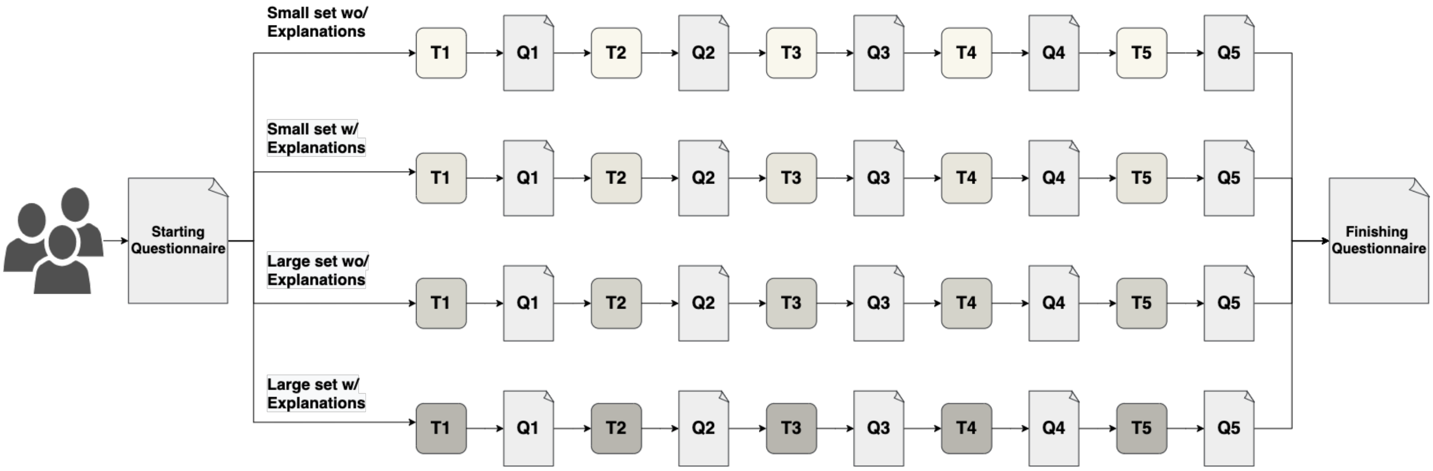


Figure 3.2: Overview of the procedure in the online experiment. Each participant is assigned one of four conditions after the initial questionnaire. The participants completes a sequence of five tasks and five short questionnaires before answering a final questionnaire on completion.

The overall procedure of the online experiment is depicted in figure 3.2. Each participant started off with answering a initial questionnaire that was presented for them and included questions about basic demographic data such as age, gender, nationality and education. Following these questions, the participants were further asked to input their dietary preferences, such as if they were vegan, vegetarians, carnivore, or pescatarian. All questions can be seen in Table 3.1. As seen in figure 3.2, upon completion of the questionnaire, each participant got assigned one random condition out of four; multi-list with explanations, multi-list without explanations, single-list with explanations and single-list without explanations. In every condition the participants had to complete a sequence of five tasks, which is depicted in figure 3.2 as T1-T5. After completion of each task there was a short questionnaire afterwards which is depicted as Q1-Q5.

Every task presented the user with a interface that had a single reference recipe and a recommendation set for other recipes. The size of the recommendation set that the participant was presented with depended on which condition the user was assigned. For example if the user was assigned the condition "large set, no explanations", he or she was presented with a reference recipe and a recommendation set that contained 25 recipes, but no specific explanation. In order to answer the research questions, in each task the participant was prompted to choose *one* specific recipe from the recommendation set which the participant was most likely to prepare at home. After making a decision, the participant was then sent to the next page and presented with a short set of questions to assess satisfaction with the chosen recipe and also the recommendation set in general. Each question was captured on a 5 - point Likert scale. As seen in figure 3.2, after completing the whole sequence of tasks (T1-T5), the participants were presented with a longer finishing questionnaire with questions regarding choice difficulty, percieved diversity and understandability.

Table 3.1: All questionnaire items asked in throughout the entire user study, categorized by which of the three surveys the questions were part of.

Survey	**Aspect**	**Survey Question**
Intro	Demographics	Age, Nationaility, Gender
	Eating behavior	Do you have any dietary restrictions?
		Which of the following goals are the most important when searching for a recipe online?
Post-Task	Satisfaction	To what extent do you like the recipe you've chosen?
		How likely are you to actually prepare the recipe you've chosen?
		How much do you like the list of recommended similar recipes?
Finishing	Choice Difficulty	I changed my mind several times before choosing a recipe.
		I think I selected the most attractive recipe from each list.
		I was in doubt between multiple recipes.
		The task of choosing a recipe was overwhelming.
	Diversity	The lists of recommended recipes were varied.
		The recommendation lists included recipes from many different categories.
		Several recipes in each list differed strongly from each other.
		Most recipes were of the same type.
	Understandability	I understood why recipes were recommended to me.
		The explanations of recipes, such as 'similar recipes', were clear to me.
		I did not understand the presented explanations.

3.3.1 Research Design

The online user experiment was subject to a 2x2 between-subjects design as the different conditions used in the study on one hand varied the recommendation set size, and on the other hand varied on the usage of explanatory labeling on the presented recommendations. In total, four conditions were investigated. The condition *Small recommendation set, without explanations* can be described as the baseline condition since minimal changes are made to the recommender system and interface. As seen in the previous chapter, in figure 2.1 and 2.2, this is a common way of presenting recommendations in commercial food recommender systems, such as allrecipes.com, and seriouseats.com. In this condition the effectiveness of a single-list interface without explanations was investigated by presenting the participant with one reference recipe, and a small recommendation set of five recipes. No explanations are accompanying the recommendations other than 'Similar recipes'.

To contrast this, the condition *Small recommendation set, with explanations* was used. Participants who were assigned this condition got presented with the same set size but also an explanatory label that describes which healthiness factor from the FSA score the the recommendation set optimized for. I.e "recipes that are similar, but low fat/carbohydrates/calories or high in protein". A partial screenshot of this condition was provided earlier, in figure 3.1.

Participants who were assigned the condition *Large recommendation set with no explanations* were presented with a reference item, and recommendation sets containing 25 recipes that varies in similarity between the recommendation sets. In this condition the recommendations were visually presented in a single large list, however the recommendations were computed by five different algorithms which makes it multi-list, even if they are not visually separated. This condition is depicted below, in figure 3.3.

In order to investigate how explanations affected user choices in a multi-list interface, the condition *Large recommendation set with explanations* was used. In this condition the interface presented participants with one reference item and with a total recommendation set of 25 recipes each, with varied similarity computed by the five algorithms. In this interface the recommendations were presented in five visually separated lists containing five recipes each. Each list had explanatory labels attached which explained the content of a particular list. These explanations were formulated based on which factor the list optimized for. The interface in this condition is depicted in 3.4. Lists that contained recipes with a low fat content had an explanation such as "recipes that are similar but contains less fat". Similarly lists with recipes that had low caloric content were presented with the explanation "Recipes with low calories", and at last the computed similarity list had the explanation 'Similar recipes'.

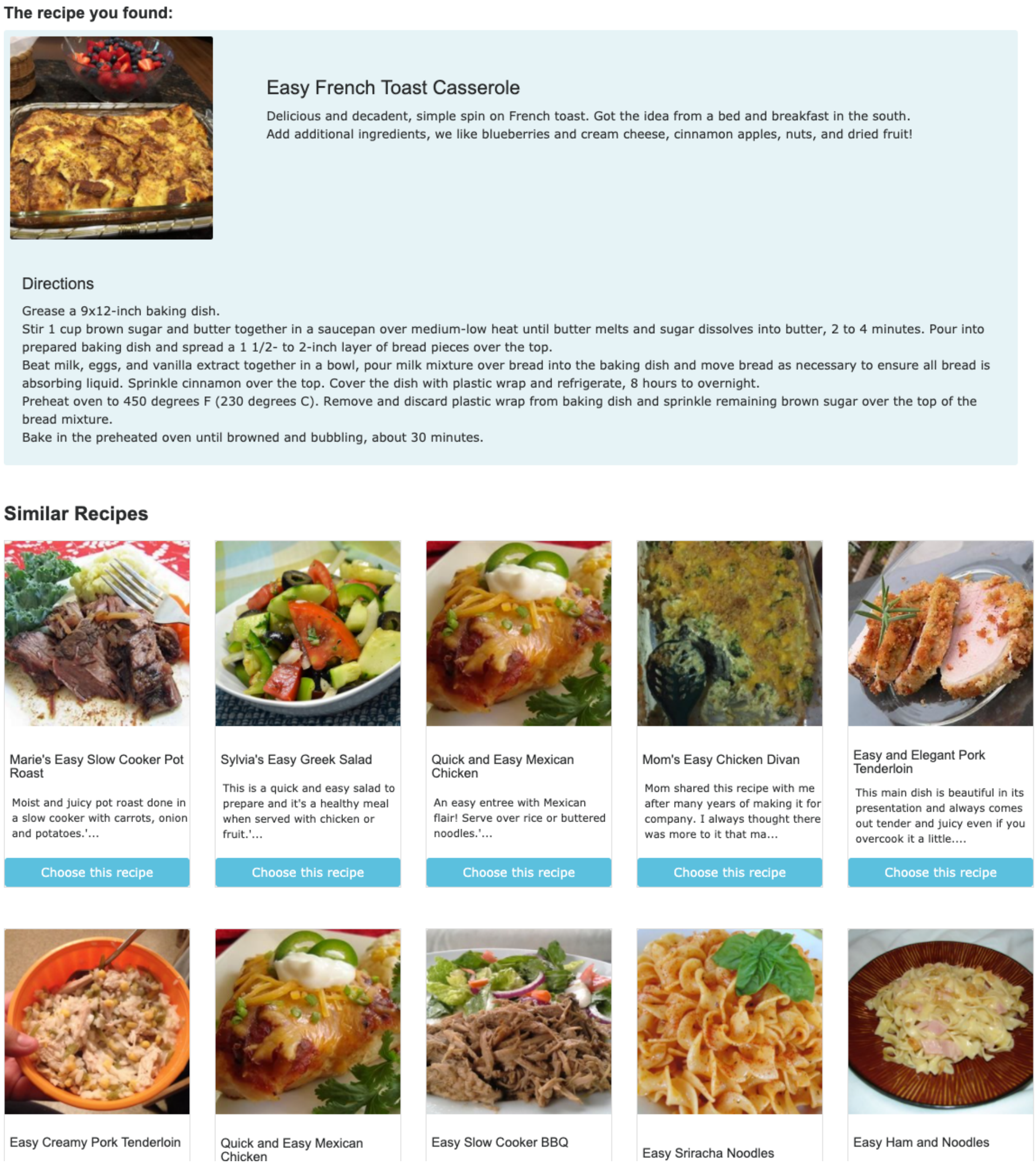


Figure 3.3: Partial screenshot of the prototype. Depicted is the multi-list condition without explanations, where participants were presented with a reference recipe and recommendations in five different lists, which were not labeled individually.

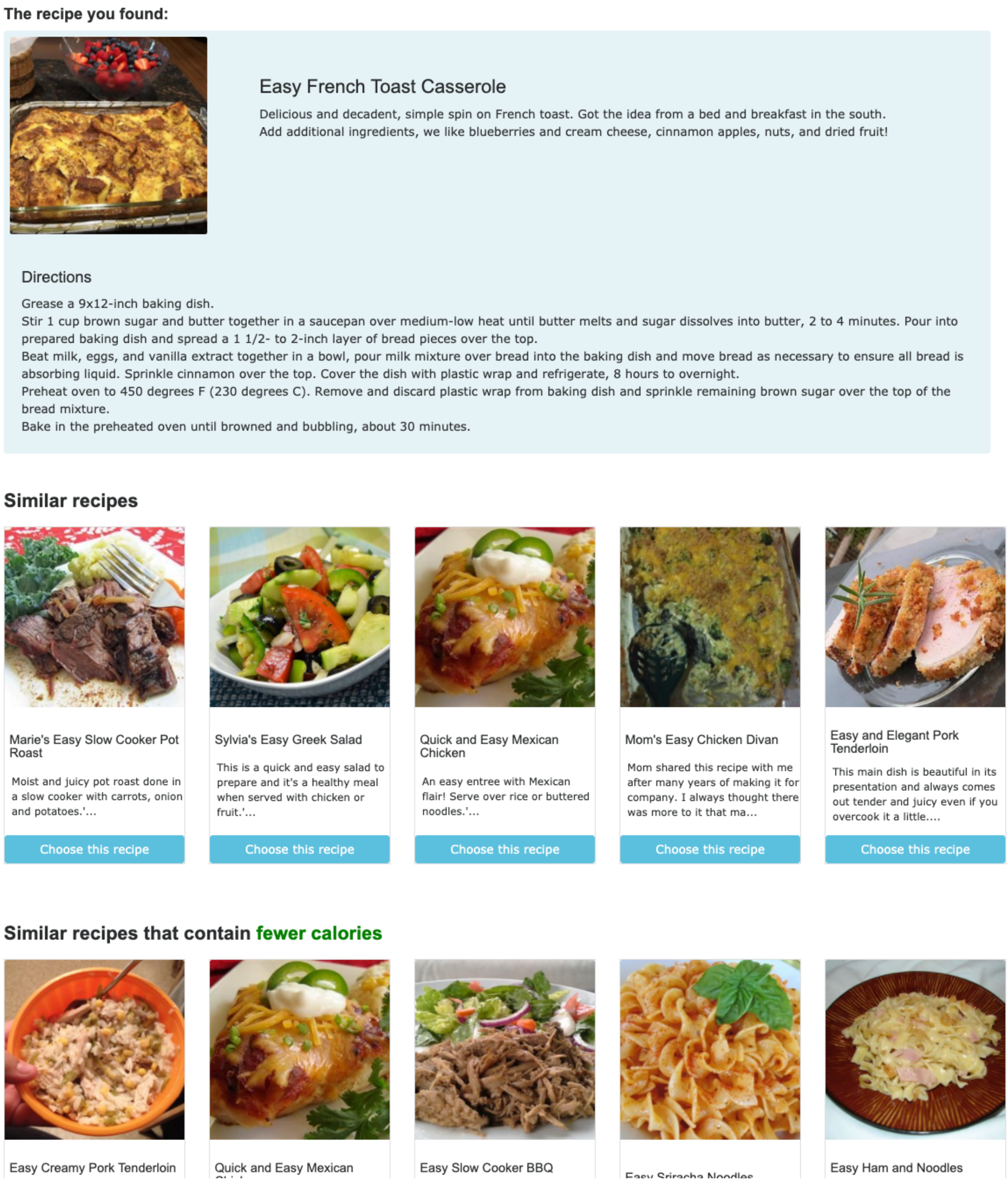


Figure 3.4: Partial screenshot of prototype. Depicted is the multi-list condition with explanations, where participants were presented with a reference recipe, as well as similar alternatives across five different and labeled lists.

3.3.2 Participants and Descriptive Statistics

General. In total ($N = 366$) participants fully completed the user experiment. The participants were sampled trough the survey platform *Prolific* ($N = 182$), and the crowd-sourcing platform *Amazon Mechanical Turk* ($N = 184$). Mean age for all participants was 34.24 with a standard deviation of 13.23, and 52 % of participants responded that they identified as male.

Recruitment Process. During the recruitment process on prolific, a filter on "No dietary restrictions" was used. Furthermore the experiment was also limited to participate only trough a desktop computer, as neither of the interfaces was optimized for any other device. On Amazon Mechanical Turk, the participants were filtered based on their qualification level, only allowing certified users of the platform, who have completed more than 500 HITs[2], Upon completion the participants were compensated with 1.25$ on Prolific, and 0.7$ on mTurk.

Cooking Proficiency and Healthiness. Apart from basic demographic questions, the introduction questionnaire included items that asked the participant about their self assessed cooking level, and also their dietary healthiness. 46.45% reported that they assess their cooking level as "Medium" good and 44 % reported that they consider they diet *healthy*, which was placed slightly over neutral. 36.8 % considered their diet as neutral. The rest of the responses on self assessed cooking proficiency can be seen in the Table 3.2, while responses for healthiness can be observed in 3.3.

Table 3.2: Self-assessed cooking proficiency.

Cooking Level	**Responses**
Very high	5.74 %
High	25.68 %
Medium	46.45 %
Low	5.74 %
Very Low	4.92 %

Table 3.3: Self-assessed dietary healthiness.

Dietary healthiness	**Responses**
Very healthy	7.3 %
Healthy	44.8 %
Neutral	36.8 %
Unhealthy	9.8 %
Very unhealthy	1.0 %

[2] A HIT represents a virtual task that a mturker can work and submit an answer to, and afterwards collect a reward for completing [4].

Health Goals. The participants could optionally input if they had any of four goals when they searched for recipes online, where multiple alternatives was available for selection. The majority of participants disclosed they had some type of goal when searching for recipes online ($N = 362$). The four goals available for selection can be separated in two different categories; *health-oriented eating goals* (Recipes should contain little fat or less calories or contain a lot of fiber) and the other category was *preference-oriented eating goals* (Recipes should be similar to what i usually like, or the recipe should fit my preferences). Out of all participants 40% disclosed reported health-oriented goals, and 76% also reported that they had preference-oriented goals.

Data Filtering. An attention check was included in the recommender system prototype, which was presented to participants on task 3 in the interface. The attention check was formulated in the form of a mathematical question on subtraction between two values. The results from the attention check showed 15 % of the users across both platforms failed to pass the attention check. However, a decision was made to perform analyses on both all users, while also reporting results when filtered non-passing participants are removed.

3.4 Measures

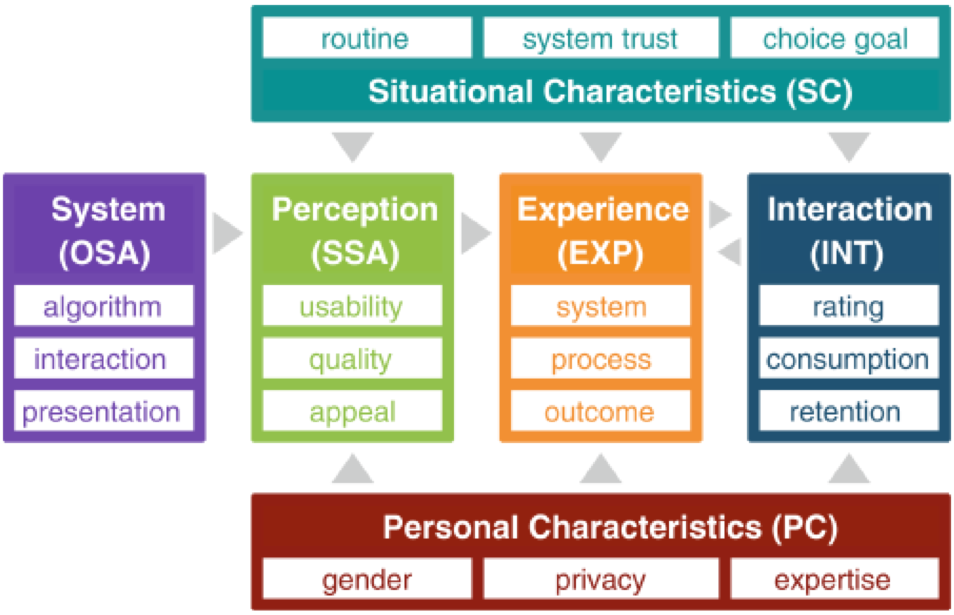


Figure 3.5: Overview of Knijnenburg et al. Framework [28]. The study investigated the relationship between changes in System and how it affects Perception, Experience and Interaction.

This section explains the several measurements that were measured in the study. To explain the measures used in the study, Knijnenburg et al. [28] framework for user-centric evaluauation of recommender systems will be used. A general overview of the framework is presented in 3.5. The measures can be explained according to different aspects of recommender systems, described in the framework. As the framework is often used for mediation analysis with structural equation modeling, it is important to note that this type of analysis is out of scope for this thesis. Instead the framework was used to explain the different measures used in the study, to explain relations between objective system aspects (OSA) and variables from the other boxes: perception (SSA), experience (EXP) and interaction (INT) of the system in terms of satisfaction, perceived diversity and what choices users made.

- **Objective system aspects (OSA):** Denotes what the system does, such as the underlying algorithms in the recommendation system, and also how the recommendations are presented to the user in a interface.

  The study investigated how changes to different OSAs affected users choices, and evaluation of the system. First, the different conditions altered between how many algorithms was used to produce recommendations for the participants, and also the recommendation set size, mainly five vs tventy-five. Another OSA that was investigated was how the presence or absence of explanatory labels affected user choice, and evaluation.

- **Subjective System Aspects (SSA) and Experience (EXP):** Firstly, SSA Denotes how the user perceive the system, such as the appeal of the recommender system and the usability and quality of the systems recommendations. Subjective system aspects is an important inclusion, as recommender systems often provide a personalized experience. The experience aspect in recommender systems according to Knijnenburg et al. signifies the users evaluation of the system and how the user perceives the interaction with the system. The experience aspect is measured using questionnaires, and is also further divided in system, process and outcome related evaluations.

  To investigate how participants evaluate multi-list interfaces with regards to food choices, versus a single list interface (RQ 1.1) the study included short questionnaires after completion of each task in a condition, where the participants could report on how satisfied they were with their chosen recipe, and the recommendation set in general. The final questionnaire also allowed for measurement of eventual choice difficulty the participants encountered, in order to answer whether explanations decreases choice difficulty (RQ2.1).

- **Interaction (INT):** Signifies the objective effect of using the system and is an observable behavior of the user. As figure 3.5 shows, the arrow between EXP and INT is double sided, meaning that there is a high interplay between these two aspects. A positive user experience can lead to a change in the interaction, but a interaction is also what caused the user experience in the first place. However, the interplay is not investigated in this study, as it is out of scope for this thesis.

  In order to answer whether a multi-list recommender interface supports healthy food choices to a larger extent than single-list interfaces (RQ1), the choices participants made in the interface was recorded, and a computed score, namely "FSA score" of each recipe was used as a measure for the healthiness of the recipes the participants chose. By looking at the amounts of macro nutrients; fat, saturated fat, sugar and salt per 100g in a recipe, the overall healthiness could then be computed by following the standards set by the Food Standards Agency (FSA). The score ranges from 4 to 12, where 4 is healthy and 12 unhealthy. Similarly, FSA scores were used as a measure to investigate to what extent explanations of recommendations support healthier food choices (RQ2). Since there are differences between conditions regarding recommendation set sizes (5 vs 25), where some are healthier than others, a comparison entirely based on the FSA score of the chosen recipe does not accurately depict if a choice was unhealthy or healthy. To combat this, an average FSA score for all presented recipes was computed, as well as a variable that allowed for relative comparison of the choice (FSA score of choice, minus average FSA of all recommendations).

- **Personal Characteristics (PC):** These are characteristics such as demographics of the user, trust the user has towards the system, the users expertise and knowledge in the domain and also perceived control. For example a new user will have low familiarity of the recommender system, but an older user has more familiarity, and thus this can affect the user experience and interaction in positive and negative ways.

  The study measured user characteristics by prompting participants to answer questions in both the initial questionnaire. Each participant was asked to report their dietary preferences; carnivore, vegan, vegetarian, pescatarian or none. Furthermore, the users reported on simple demographic data, such as age and gender, and also their skill level when it comes to cooking. In addition they were inquired on if they had any health-related goals or not (i.e: recipes should be healthy and contain little fat), as well as any preference-realted goals (i.e: recipes should match my previous preferences). As demonstrated in previous chapters, users have different preferences for food, as well as goals which can affect their choices. Having specific goals because of health-related issues or the desire to eat healthier changes how a person looks for food items and what they can consume [29, 39]. By capturing whether participants have a health-related goal, an comparison can be conducted to see if users with health-related goals choose makes different choices trough the various tasks, compared to users without such goals and also it is possible to compare the level of choice-satisfaction between the two groups. Moreover, same comparisons can be conducted for participants with or without preference-related goals, to identify how having preferences affect user choices or satisfaction.

3.5 Statistical Analyses

The Knijnenburg et al. [28] framework mentioned in the measures section recommends a path model for evaluating recommender systems. In relation to this study the framework was not used beyond explaining the different measures in the study. Thus the statistical analyses conducted consisted primarily of comparing variances between the conditions from the experiment, and examining which changes to system aspects affected user choices and evaluation. The main method of comparison was utilizing two-way ANOVA tests, to examine variances and interaction in the 2x2 Research design (Multi-list x Explanations).

As the finishing questionnaire contained 11 items across three factors, an exploratory factor analysis was performed to remove items that did not load properly on their respective factor, and to reduce the total amount of items to three factors. A *Cronbachs alpha* test was conducted as a confirmatory step, to examine the internal consistency of the questionnaire. After a factor analysis model was fitted, a prediction was ran on values for the three factors across all individuals. Same approach was conducted for the shorter, post-task questionnaire with three items.

Chapter 4

Results

This chapter provides an overview of the results from the conducted analysis. The sections in the chapter are primarily organized by research questions, and the analysis conducted for answering them. Additional analyses were performed on users with eating goals versus without, in order to determine how having goals can affect the FSA score of a chosen recipe as well as satisfaction. Furthermore, reports on a conducted regression analysis, as well as other interesting findings is reported at the end of the chapter.

- Section 4.1 details results of analyses performed in order to evaluate if multi-list recommender systems leads to healthier food choices, as well as how having eating goals affects the healthiness of choices.
- Section 4.2 details results of analyses performed to determine whether multi-list interfaces increase satisfaction or not, as well as how having eating goals affects the evaluation of the recommender system.
- Section 4.3 details the results of analyses performed to evaluate the effectiveness of explanations in regards to healthier food choices.
- Section 4.4 details the results of the analysis performed to see if explanations can decrease choice difficulty.
- Section 4.5 Shows results of a fitted regression model used to examine which factors determines the FSA score of chosen recipes.
- Section 4.6 details the results of other, exploratory analyses conducted.

4.1 Multi-list For Supporting Healthier Food Choices (RQ1)

To answer RQ1, *To what extent can a multi-list recommender interface support healthy food choices, compared to a single list interface?* a two-way ANOVA test on the FSA score of the chosen recipe was performed across the four different conditions. A two-way ANOVA test was conducted to identify significant variances between the conditions, and the interaction between list conditions and explanation conditions.

FSA Deviation Between Recommendations and Choice. A two-way ANOVA test (Table 4.1) conducted on the *relative FSA score*, between multi-list and single-list groups shows that the average deviation was higher for participants using the multi-list interface ($M = 0.36$) compared to the single-list interface ($M = 0.03$): $F(1, 1826) = 25.07$, $p < 0.001$. An interaction effect was also found between explanations and list conditions, where the effect is apparent between the two single-list conditions; without-explanations ($M = 0.18$) and with-explanations ($M = -0.10$): $F(1, 1826) = 4.77$, $p < 0.05$. The marginal effects are depicted in figure 4.1 No significant differences were found when omitting users who did not pass the attention check.

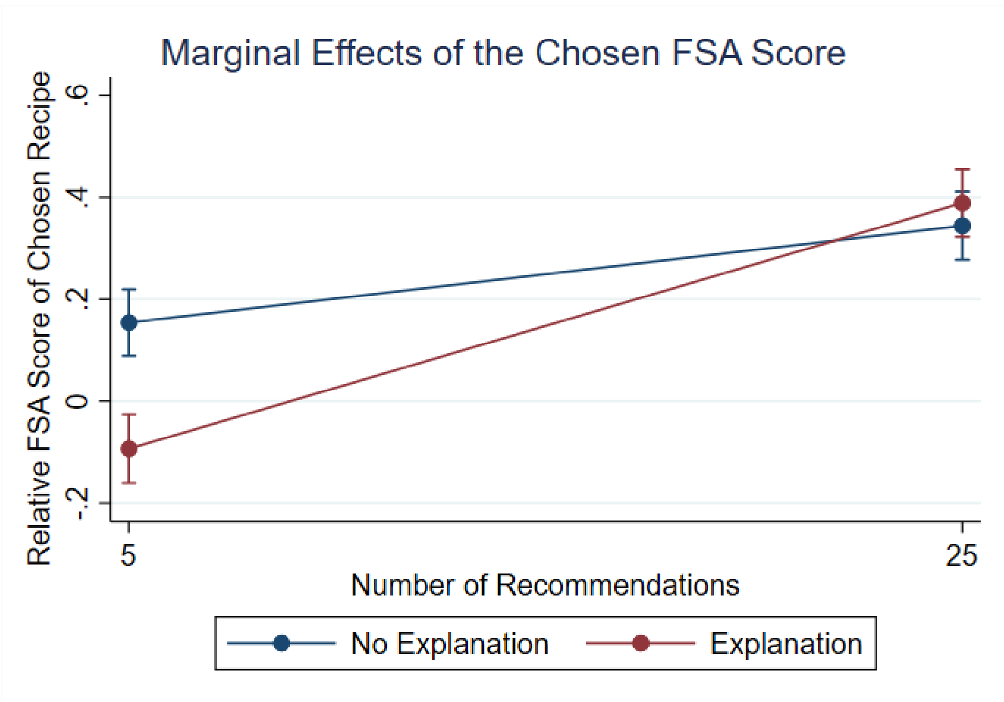


Figure 4.1: Marginal effects plot showing the interaction effect between the explanation and list condition, with S.E. Bars.

Table 4.1: Results of two-way ANOVA on the relative FSA score of chosen recipes.

Condition	**df**	**SS**	**MS**	F	p
Explanation	1	4.13	4.13	2.02	0.155
List	1	51.21	51.21	25.07	0.000
Explanation:List	1	9.75	9.75	4.77	0.029
Residual	1826	3729.45	2.04		

Considering Health-oriented Goals. Also analyzed was the relative FSA score while considering the different goals participants had while searching for recipes, trough a two-way ANOVA test. Results from the analysis, presented in Table 4.2 show that participants with health-oriented recipe search goals chose on average healthier recipes ($M = 0.08$) compared to those which did not have such goals ($M = 0.27$): $F(1,1828) = 7.33$, $p < 0.01$. No significant changes in the results were discovered after omitting participants who did not pass the attention check.

Considering Preference-oriented Goals. Apart from health-oriented goals, the participants could optionally inform if they have goals that a particular recipe should either match their preferences, or that they should be similar to what they usually like. Reported in Table 4.2, results from a one-way ANOVA showed that those who did not disclose any preference goal had higher FSA deviation ($M = 0.24$) compared to those with such goal ($M = 0.16$): $F(1,1828) = 1.47$, $p = 0.22$. The variation was however not found significant. No statistically significant results were found when participants who did not pass the attention check were omitted. Finally, results from a comparison between users with a goal for similar recipes, versus those without (Reported in Table 4.2) showed that having a similarity goal led to higher FSA deviation ($M = 0.26$), compared to not having such goal ($M = 0.13$): $F(1,1828) = 3.63$, $p = 0.057$. However, the variance was also not found significant. Omitting the non passing participants yielded the same outcome.

Table 4.2: Results of one-way ANOVA on the relative FSA score between users with goals.

Goal	**df**	**SS**	**MS**	*F*	*p*
Health-oriented goal	1	15.16	15.16	7.33	0.066
Residual	1828	3779.39	2.07		
Preference Goal	1	3.06	3.06	1.47	0.224
Residual	1828	3791.48	2.07		
Similar Goal	1	7.52	7.52	3.63	0.057
Residual	1828	3787.03	2.07		

4.2 Multi-List for Increasing Satisfaction (RQ1.1)

Table 4.3: Results of the factor analysis on user satisfaction with regards to their recipe choices, which was measured on 5-point scales. A single latent factor was eventually formed, labelled "User Satisfaction".

Aspect	**Item**	**Loading**
User Satisfaction	To what extent do you like the recipe you've chosen?	.8629
	How likely are you to actually prepare the recipe you've chosen?	.8428
	How much do you like the list of recommended similar recipes?	.7225

Similar approach to answering RQ1 was taken in order to answer RQ1.1: *To which extent do multi-list interfaces increase satisfaction regarding their food choices compared to a single-list interface?* In the experiment the participants were presented with a short questionnaire after each selection task, inquiring on satisfaction of the chosen recipe. All three items are presented in Table 4.4. The answers were captured on a 5-point Likert scale.

Factor Analysis. A factor analysis was conducted on the three questionnaire items asked after each recipe selection task, which can be seen with their loadings in Table 4.3. The results from the factor analysis show only one reliably identified factor, with proper loadings for each item above 0.5. To confirm internal consistency of the questionnaire, a Cronhbachs Alpha test was conducted showing an alpha score of $\alpha = 0.8$ which is an acceptable level of reliability. While a factor analysis was also conducted for the finishing questionnaire regarding choice difficulty, the two analyses were done separately. The decision was made due to the multi-levelness of responses on the short questionnaire, where each participant answered the questions five times in total, while the choice difficulty questionnaire was answered only once per participant.

Analyzing User Satisfaction. Using the single factor distinguished by the factor analysis above, a two-way ANOVA test reported in Table 4.4 showed higher satisfaction levels for multi-list users ($M = 0.11$), compared to single-list users ($M = -0.11$): $F(1,1826) = 28.10$, $p < 0.001$. No interaction effect was found between explanations and list condition: $F(1,1826) = 2.89$, $p = 0.08$.

Whereas the multi-list interface was more favorable on the dataset with all users, the single-list interface becomes more favorable when omitting users that did not pass the attention check, as reported in Table 4.5. For multi-list users the satisfaction score was lower ($M = -0.11$). In comparison single-list users showed higher satisfaction levels ($M = 0.11$): $F(1,1551) = 22.71$, $p < 0.001$. An interaction effect between list and explanation conditions was also found, showing differences between single-list with explanations ($M = 0.03$) and single-list without explanations ($M = 0.18$), as well as multi-list without explanations ($M = -0.15$) and with explanations ($M = -0.07$): $F(1,1551) = 5.95$, $p < 0.05$. The marginal effects are illustrated in figure 4.2.

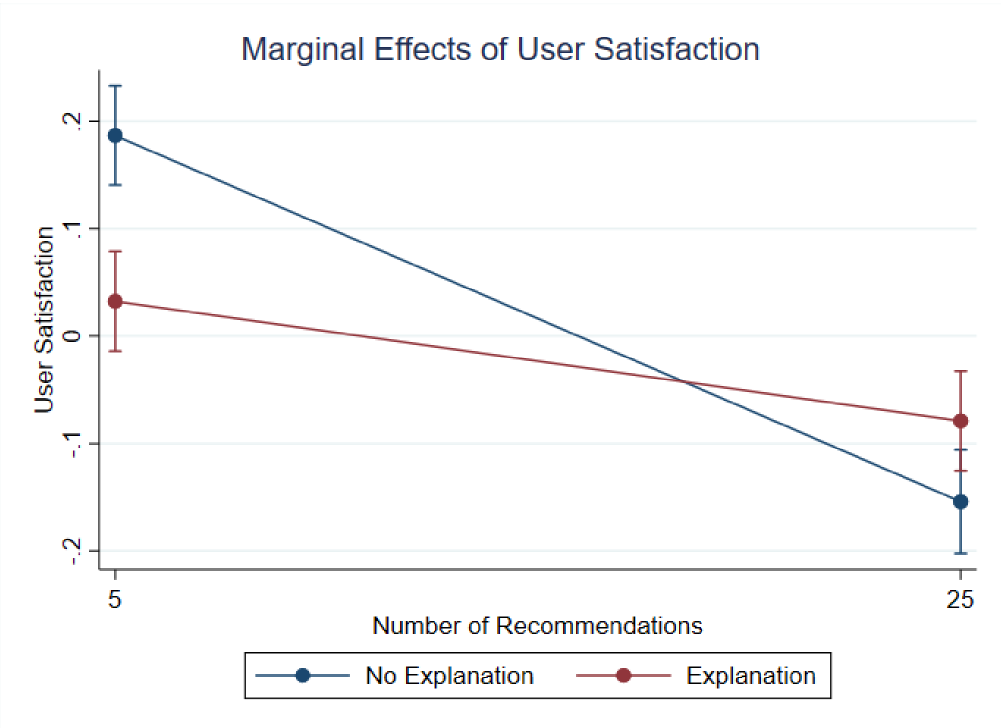


Figure 4.2: Marginal effects plot showing the interaction effect between the explanation and list condition, when omitting participants that did not pass the attention check, with S.E. Bars.

Table 4.4: Results of ANOVA on the user satisfaction.

Condition	**df**	**SS**	**MS**	*F*	*p*
Explanation	1	1.3	1.3	1.51	0.218
List	1	23.96	23.96	28.10	0.000
List:Explanation	1	2.46	2.46	2.89	0.090
Residual	1826	1556.91	2.92		

To identify the reason for discrepancy when omitting non-passing users, an analysis was conducted on *only* the participants that did not pass the attention check. Results, as reported in Table 4.6 show that users which did not pass the attention check rated the multi-list interface very favorably ($M = 0.10$), whereas quite lower values were found for single-list with explanations ($M = -0.12$): $F(1,271) = 3.89$, $p < 0.05$. Taking these values into consideration may explain the discrepancy in satisfaction when omitting participants that did not pass the attention check.

Table 4.5: Results of two-way ANOVA on user satisfaction, after omitting users who did not pass the attention check.

Condition	**df**	**SS**	**MS**	*F*	*p*
Explanation	1	0.87	0.87	1.02	0.313
List	1	19.50	19.50	22.71	0.000
List:Explanation	1	5.11	5.11	5.95	0.015
Residual	1551	1331.83	0.86		

Table 4.6: Results of two-way ANOVA on the 55 participants that did not pass the attention check.

Condition	**df**	**SS**	**MS**	*F*	*p*
Explanation	1	2.78	2.78	3.41	0.066
List	1	3.18	3.18	3.89	0.050
List:Explanation	1	2.36	2.32	2.84	0.093
Residual	271	221.24	0.82		

Considering Health-oriented Goals. Results from a one-way ANOVA reported in table 4.7 showed that participants without any health oriented goal showed lower satisfaction with their choices ($M = -0.11$) compared to those who informed that they had any of the two health oriented goals ($M = 0.16$): $F(1,1828) = 38.57$, $p < 0.001$. Differences were found when omitting users that did not pass the attention check. Users with health oriented goals were *less* satisfied with their choice ($M = -0.14$) compared to those without such goals ($M = 0.08$): $F(1,1553) = 22.48$, $p < 0.001$.

Considering Preference-oriented Goals. Also explored was the relationship between preference oriented goals that participants had. The results from the ANOAVA are presented in Table 4.7 showed that participants which disclosed that the recipes should match their preferences were less satisfied with their choice ($M = -0.07$), compared to those without such goal ($M = 0.08$): $F(1,1826) = F12.64$, $p < 0.001$. Different tendencies were found when non-passing users on the attention check were omitted. Users who reported having a preference goal reported higher satisfaction with their choice ($M = 0.04$) compared to users without preference goal ($M = -0.06$): $F(1,1553) = 5.59$, $p < 0.05$.

Table 4.7: Results of one-way ANOVA test on user satisfaction between users with and without goals.

Condition	**df**	**SS**	**MS**	F	p
Health goal	1	32.74	32.74	38.57	0.000
Residual	1828	1551.88	0.85		
Preference goal	1	10.89	10.89	12.648	0.000
Residual	1828	1573.73	0.86		

Table 4.8: Results of one-way ANOVA test on user satisfaction between users with and without goals, after omitting users who did not pass the attention check.

Condition	**df**	**SS**	**MS**	F	p
Health goal	1	19.37	19.37	22.48	0.000
Residual	1553	1337.94	0.86		
Preference goal	1	4.87	4.87	5.59	0.018
Residual	1553	1352.44	0.87		

4.3 Recommendation Explanations for Supporting Healthier Choices (RQ2)

In line with the analysis performed for RQ1, RQ2: *To what extent do recommendation explanations support healthier food choices?* was examined by performing a two-way ANOVA test on the relative FSA score variable.

FSA Deviation Between Recommendations and Choice. Results from a two-way ANOVA test reported in Table 4.1 showed no evidence for explanations decreasing the relative FSA score. Even though the deviation was higher in the conditions with explanations ($M = 0.24$) versus without explanations ($M = 0.15$): $F(1, 1826) = 2.02$, $p = 0.15$ The variance can be explained due to the slight existence of interaction effect between the list condition and explanation condition, which is apparent between single-list with explanations ($M = 0.15$) and without explanations ($M = -0.09$): $F(1, 1826) = 4.77$, $p < 0.05$. The marginal effects of the interaction can be seen in figure 4.1. No significant changes were found in the results when omitting users who did not pass the attention check. The tendencies of unhealthier choices by users in non-explanation conditions still persist, however the results are not significant.

4.4 Explanations for User Evaluation (RQ2.1)

After the participants completed the five selection tasks, they were presented with an final questionnaire containing 11 items. The questionnaire inquired on three main factors: Choice difficulty, recommendation variety and understandability of the recommendations and explanations.

Factor Analysis. Firstly, an exploratory factor analysis was conducted in order to remove items that did not load properly on their respective factor. The results from the factor analysis can be seen in Table 4.9 with loadings, as well as the removed items. At last a Cronbachs Alpha test was conducted after the factor analysis to check the internal consistency of the questionnaire, with results showing $\alpha = 0.5$ before removal of items that did not load properly, and $\alpha = 0.61$ after removal, which is an acceptable level of reliability. The factor analysis allowed for prediction of values for each participant across the three factors, which were in turn used in a two-way ANOVA between the conditions.

Table 4.9: Results of the exploratory factory analysis on on user evaluation aspects, which participants were inquired on after completing the five selection tasks. Items in grey and without factor loading were omitted from the analysis.

Aspect	Item	Loading
Choice Difficulty	I changed my mind several times before choosing a recipe.	.7051
	I think I selected the most attractive recipe from each list.	
	I was in doubt between multiple recipes.	.7255
	The task of choosing a recipe was overwhelming.	.5674
Perceived Diversity	The lists of recommended recipes were varied.	.5755
	The recommendation lists included recipes from many different categories.	.6433
	Several recipes in each list differed strongly from each other.	.5299
	Most recipes were of the same type.	.5926
Understandability	I understood why recipes were recommended to me.	6623
	The explanations of recipes, such as 'similar recipes', were clear to me.	.7565
	I did not understand the presented explanations.	

Perceived Choice Difficulty. One of the factors measured if users experienced choice difficulty when selecting a recipe from the recommendation sets. A two-way ANOVA test, reported in Table 4.10 between the conditions of with and without explanations showed that explanations did not effectively reduce choice difficulty for participants. Participants assigned conditions with explanations did report higher choice difficulty ($M = 0.04$) compared to participants without explanations ($M = -0.04$): $F(1,362) = 0.89$, $p = 0.34$, however the variance was not found significant. Furthermore, no significance was found on the interaction between list-condition and explanation-condition: $F(1,362) = 0.01$, $p = 0.92$. Even though single list users reported less choice difficulty compared to multi-list users, the variance is not explained by the inclusion of explanations, but rather multi-list condition creating higher choice difficulty.

Perceived Diversity of Recommendations. Furthermore, there was no significant variance found (Table 4.11) on perceived diversity of recommendations between having explanations ($M = 0.05$), versus not having any explanations ($M = -0.5$): $F(1,362) = 0.58$, $p = 0.44$. No interaction effect was found between list and explanation conditions: $F(1,362) = 1.79$, $p = 0.18$.

Table 4.10: Results of the two-way ANOVA on the choice difficulty variable.

Condition	**df**	**SS**	**MS**	*F*	*p*
Explanation	1	0.65	0.65	0.89	0.346
List	1	15.15	15.15	20.75	0.000
List:Explanation	1	0.01	0.01	0.01	0.924
Residual	362	264.29	0.73		

Table 4.11: Results of two-way ANOVA on the diveristy variable.

Condition	**df**	**SS**	**MS**	*F*	*p*
Explanation	1	0.41	0.41	0.59	0.445
List	1	1.05	1.05	1.49	0.223
List:Explanation	1	1.26	1.26	1.79	0.182
Residual	362	255.59	0.71		

Perceived Understandability. The last factor measured how well the participants understood both the explanations and recommendations. Results from a one-way ANOVA test reported in Table 4.12 showed no significant difference for explanations increasing the understandability of recommendations. Although variance in means exists between having explanations ($M = 0.07$) compared to not having any explanations ($M = -0,07$): $F(1,362) = 3.08$, $p = 0.07$ conditions, the results are not statistically significant. This entails that participants that were served explanations did not have higher understanding of the recommendations and explanations compared to those who were not served any explanation further than *Similar recipes*. Equal analyses for each of the factors were performed on the dataset after omitting users which did not pass the attention check, without significant changes in the results.

Table 4.12: Results of two-way ANOVA on the understandability variable.

Condition	**df**	**SS**	**MS**	*F*	*p*
Explanation	1	2.10	2.10	3.08	0.080
List	1	0.04	0.04	0.05	0.818
List:Explanation	1	0.59	0.59	0.86	0.353
Residual	362	246.27	0.68		

4.5 Regression Analysis

To examine which factors determine the FSA score of chosen recipes, a regression analysis was conducted on the dependent variable *FSA score*, based on several independent variables on personal characteristics, as well as the two condition variables. The variables that were used in the regression analysis stems from the introductory questionnaire, and can be categorized in four different categories, all variables can be seen in the entire regression Table 4.13. Variables under demographics include age, education and gender, while the category "eating goals" are concerned whether the participant disclosed any of the four eating goals as explained earlier. Furthermore, the variables multi-list and explanation were used in the regression analysis, as well as personal factors such as if participants had any allergies, what level of cooking experience they had as well as the healthiness of their diets.

The dataset was examined for potential multicollinearity before conducting the regression analysis. No traces of multicollinearity was found. The pairwise correlations are depicted in figure 4.3.

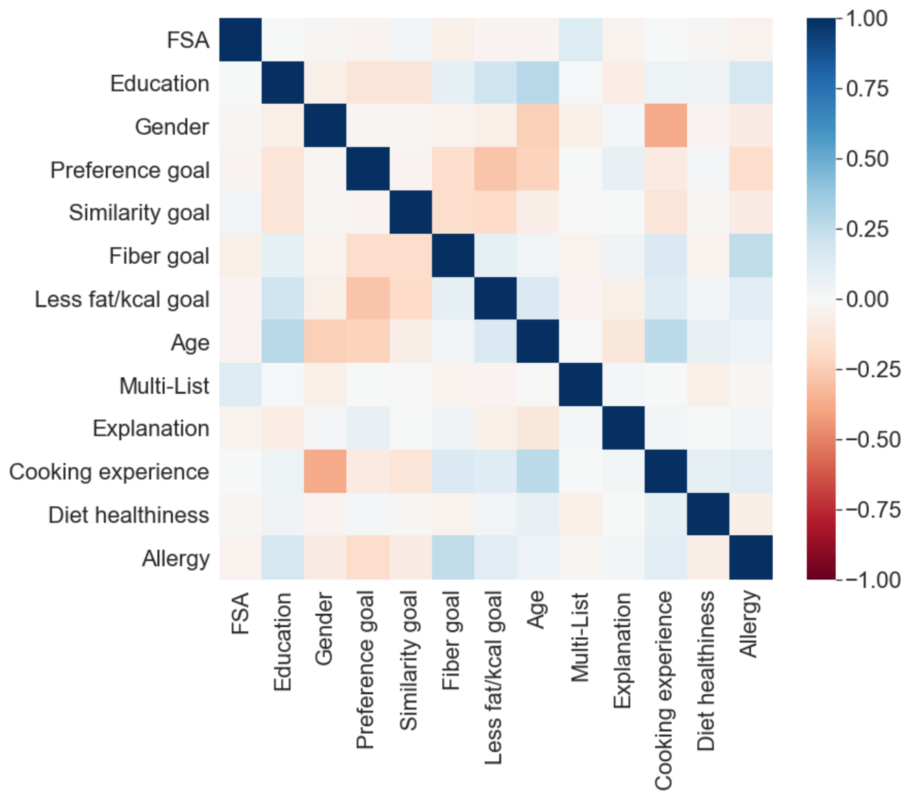


Figure 4.3: Heatmap depicting pairwise correlations, which were used to examine multicollinearity.

Demographics. The intro-survey inquired users on three different demographic variables, namely age, education and gender. In turn, these variables were used to examine how they affect the FSA score of the chosen recipe. However, as shown in Table 4.13 no statistically significant results were found for either of the three variables.

Eating goals. Responses on whether participants had any of the four eating goals were also used as variables in the regression analysis, to identify if having eating goals determines the FSA score of a chosen recipe. As shown in Table 4.13, two statistically significant results were found, namely having the eating goal *recipes should be healthy and contain a lot of fiber* ($\beta = -0.245$, $p < 0.05$), and *the recipe should fit my preferences* ($\beta = -0.196$, $p < 0.05$). Having either of these goals affected the FSA score positively, meaning lower FSA scores. The goal, *Recipes should contain less fat and fewer calories* also affected the FSA score positively ($\beta = -0.114$, $p = 0.23$), however this result was not statistically significant. The last goal, *recipes should be similar to what i usually like* impacted the FSA score negatively, as well as having much smaller impact. The results was however not statistically significant ($\beta = 0.071$, $p = 0.4$).

Condition Variables. Both condition variables were used in the regression analysis; whether the participants was assigned multi-list condition or not, as well as if the interface presented explanations or not. The findings were consistent with what was discovered in the ANOVA tests reported earlier. Multi-list condition showed negative impact on the FSA score, meaning unhealthier recipes ($\beta = 0.451$, $p < 0.001$). Explanations showed positive impact on the FSA score, however the results was not found significant ($\beta = -0.118$, $p = 0.15$).

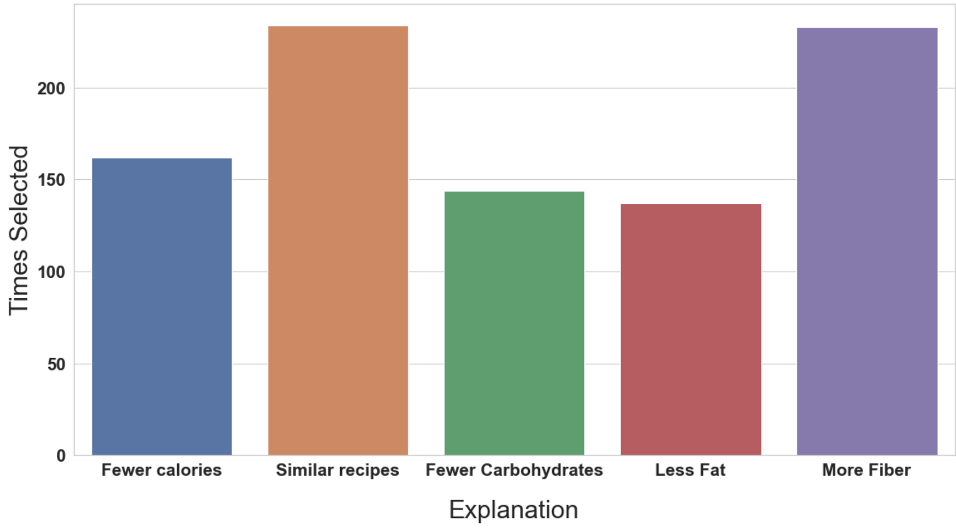
Personal Factors. Three additional personal factors were also used in the regression analysis, which are concerned with having any eating allergies, what level of self-assesed cooking experience the participant had, as well as how the participant assessed the healthiness of their diet. While the variable concerned with allergies showed positive impact on the FSA score ($\beta = -0.156$, $p = 0.10$), neither of the results showed statistical significance.

Table 4.13: Results of regression analysis on the dependent variable FSA score of the chosen recipe.

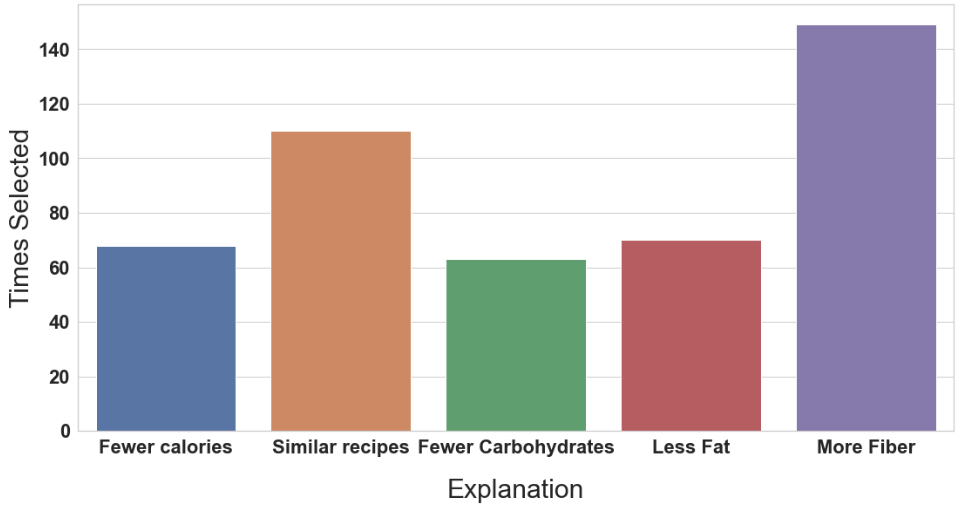
Aspect	**Variable**	β	$Std.Err$	p
Demographics	Age	-0.006	0.003	0.061
	Education	0.057	0.051	0.264
	Gender	-0.057	0.088	0.523
Eating goals	Fiber Goal	-0.245	0.113	0.030
	Less Fat/Cal Goal.	-0.114	0.094	0.225
	Pref. Goal	-0.196	0.088	0.026
	Sim. Goal	0.071	0.084	0.396
Condition	Explanation	-0.118	0.081	0.146
	Multi-List	0.451	0.080	0.000
Personal	Allergy	-0.156	0.095	0.102
	Cooking Experience	0.045	0.055	0.409
	Diet Health.	-0.006	0.044	0.899

4.6 Other Findings

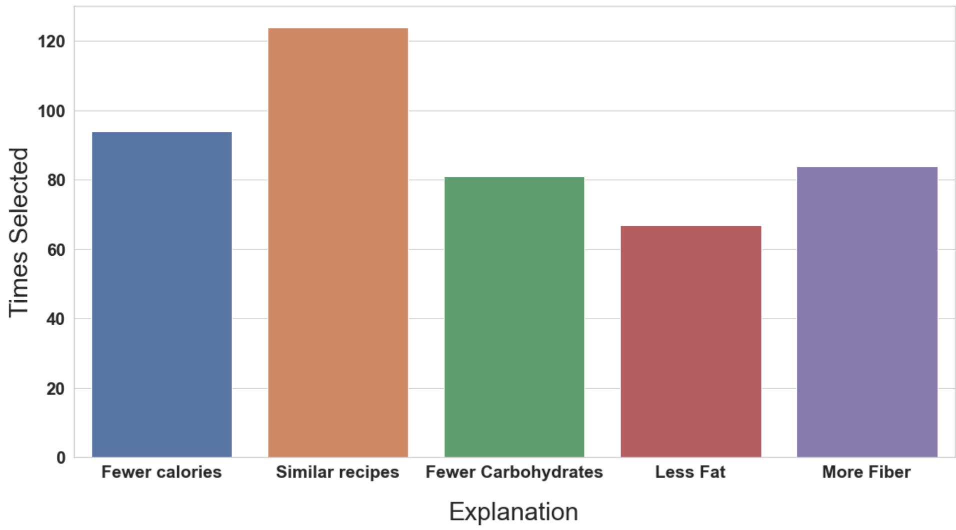
Other analyses were also conducted, in order to identify potential interesting results regarding favorable lists. While all single-list users made one choice for each list, the multi-list users had the ability to select from multiple selections for each list. A simple count between participants in the multi-list condition show that "similar recipes" list was selected from most (233 times), with "more fiber" coming in second place (233 times), which is depicted in figure 4.4, subfigure A. However, comparing multi-list with explanation users to multi-list without explanations it became evident that participants assigned the multi-list condition with explanations selected more from "recipes with more fiber" list, as opposed to users without explanation where "Similar recipes" was more favorable. The difference in distributions are depicted in figure 4.4, under subfigure B and C.



(a) Both multi-list conditions



(b) Multi-list with explanations



(c) Multi-list without explanations

Figure 4.4: Distribution of which lists recipes were chosen from, in the Multi-List condition only.

Chapter 5

Discussion

The goal of this thesis was to investigate whether a multi-list food recommender interface with explanations could lead to healthier food choices, as well as exploring how multi-list and explanations affect how users evaluate the recommender system in terms of reducing choice difficulty, and increase perception of diversity and understandability. Multi-list recommender interfaces and explanations have been both utilized in commercial settings, but limited attention has been paid to effects such interfaces have in recommender systems, not only limited to the food domain. Furthermore, considering health in food recommender systems has only been focused in recent years, which is a challenge with regards to providing healthier meals for users. Multi-list interfaces have however been explored in research where in [11, 40], the authors examined satisfaction levels, trust and intention to return. In [38] the authors looked at the persuasive abilities of structured recommendation sets, versus unstructured, and [26] explored different multi-list approaches for similarity-based recommendations. However, research on multi-list interfaces have yet to be addressed in the food recommender system domain.

This thesis has addressed the mentioned shortcomings by conducting an online experiment with a food recommender prototype. This, in turn allowed for several comparisons on healthiness and satisfaction between a multi-list interface and more traditional single-list interface. The work conducted in the thesis also compared potential effects of explanations on both the healthiness of user choices, as well as their evaluation of the recommender systems expressed in choice difficulty, perceived diversity and understandability, as well as satisfaction. On one hand, there has been some recent work for health in food recommender systems, even with explanations, such as Cataldo et al. [36] which explored natural language justifications based on user characteristics and knowledge about the food items. Explanations in the form of nutritional values were explored in [22, 15, 31]. On the other hand explanations in conjunction with multi-lists have been utilized and researched previously in multiple domains such as Netflix, however their effects have not been evaluated thoroughly in a user-centric way.

Multi-list For Healthier Food Choies (RQ1). With regards to multi-list for healthier choices, it was expected that a multi-list interface where several lists optimized for different healthiness factors would lead to users taking healthier food choices (RQ1). However, results from the analysis show the contrary; multi-list interface led to users taking unhealthier choices, compared to the single-list counterparts. Not only did users in single-list conditions have healthier choices compared to multi-list users, they also actively chose healthier alternatives relative to what they were presented. As the single-list interface in the study had somewhat healthier recommendations, the larger availability of healthier alternatives may have led to healthier choices as well. The experiment did not specifically prompt users to select healthy recipes, but rather what they find satisfactory. While no previous works have been conducted on multi-lists in food recommender systems regarding healthiness of choices, looking at previous findings with regards online food preferences tendencies are that preferences towards unhealthier food items are more prevalent [16]. These findings are likewise supported by the study conducted by Musto et al. [37] which found that popular recipes in their dataset were most likely to be chosen if they contained more carbs and saturated fat. Results in this thesis further supports the notion of unhealthy recipes being selected more frequently as well as being evaluated more satisfactory. Meaning that the multi-list recommender system was not able to conquer users preferences towards unhealthier alternatives.

Multi-list and Evaluation (RQ1.1). It was examined whether multi-list recommender systems could affect how users evaluates the recommender system, because of previous findings in [25, 38], which links larger diversity of choices with higher satisfaction. From the analysis detailed in section 4.2 findings show that users were indeed more satisfied with multi-list interfaces compared to those that used the single-list interface. This entails that results from the study are similar to previous studies which have showed how increasing the amount of options also increases the choice satisfaction [25, 38]. Furthermore, user evaluation was measured on both their recipe choices, as well as the entire set of recommendations, and perceived diversity. Users expressed more satisfaction with their choices in the multi-list interface compared to single-list users, which is consistent with previous studies linking organization of recommendations with satisfaction. Chen & Pu found that organization based interfaces left users more satisfied [11], as well as the users built more trust to the system and had higher intention to return [40]. Surprisingly, the tests done on perceived diversity did however not find statistical significant evidence that users had the impression of a more diverse recommendation set in a multi-list interface, which is in contrast to Jannach et al. [26] who found that multi-list interfaces left the user with the impression of more diversified set of recommendations.

Explanations for Healthier Food Choices (RQ2). Also investigated was the addition of explanations to determine if this could affect food choices users made towards healthier alternatives, based on previous studies that explanations may have persuasive abilities [50, 51]. One factor for humans selecting unhealthy food online can be explained by the general preference for unhealthy food, as well as not being able to identify what is healthy or not. By explaining each set of recommendations it was hypothesized that this could mitigate the lack of knowledge and lead to higher appeal for less similar, but healthier alternatives. The analysis however showed that explanations did not have any significant effect on the healthiness of chosen recipes. These results confirms previous findings regarding persuasion and explanations, which have shown that users can be hard to persuade, even with explanations [52]. Previous studies have demonstrated that it may be challenging for consumers and users to understand nutritional information, and deciding if food items are healthy or not [42, 16]. While the explanations informed the user about nutritional content in the recipes, limited understanding on the subject of nutrients may have led to inefficient nudging towards healthier recipes.

Eating goals. Further analyses showed that personal factors could serve as potential factors for selecting healthier recipes. The experiment inquired on whether participants had any specific goals when searching for recipes online, both health-oriented goals, as well as preference-oriented goals. The analysis showed significant difference between users with health-oriented goals compared to those without any. Users who disclosed that they had a health-oriented goal selected healthier recipes. Similar findings where found in a fitted regression model which showed significant results between the health-oriented goal *Recipes should be healthy and have lot of fiber* and a positive effect on the FSA score (healthier recipes), further confirming that users that have health-oriented goals select healthier alternatives. This can be seen up against findings in previous literature from other domains showing that people with healthier diets are more aware of nutritional values in products, and what constitutes healthy meals or not, as well as consumers with diet related health problems utilizes nutrition labeling in order to identify food that they can consume [29, 39]. This could entail that participants which had health-oriented goals had greater prerequisites to identify healthier recipes and also have initial preference for selecting such recipes in the first place, as they will constitute a satisfactory choice.

As previous findings show that humans generally have a preference towards unhealthy food [16], analyses were conducted for both preference-oriented goals to see how having such goals affects the FSA score of a chosen recipe. The findings showed variance in FSA score, between goal-having users and those without, meaning that participants with the goal *Recipes should be similar to what i usually like* chose recipes with higher FSA score. However, the results were not found statistically significant in a two-way ANOVA test. For the goal *Recipes should fit my preferences* the participants who reported having this goal chose more healthier recipes, but yet no statistical significance was found. Analyses were also conducted to measure any correlation between personal factors and satisfaction, similar to what was done for healthiness. Findings from the analysis showed significant difference between users who had health-oriented goals as opposed to users without. Those who reported having health-oriented goals were more satisfied with their choices and recommendations, whereas those without reported slight dissatisfaction, which could imply that the recommender system to some extent supported users with health-goals in finding satisfying choices, similar to how Musto et al. User Holistic Model supported participants who seeked general healthy eating habits [37].

Explanations and Evaluation of Recommender System(RQ2.1) Finally it was hypothesized that addition of explanations to the recommendations would lead to decreased choice difficulty and increase perception of diversity and understandability. Findings from previous studies suggest that explanations of recommendations makes it more comfortable to users to use a recommender system [40], as well as explanations can help users understand the recommendations better thus making it easier for them to make a choice [58]. Surprisingly, the findings detailed in section 4.9 found no significant difference for explanations reducing choice overload, or increasing users understandability of what was presented to them. Higher prevalence of choice overload was found by users using the multi-list system, which are consistent with findings in previous studies done on the subject of choice overload [25], and explanations did not successfully mitigate this issue.

5.1 Limitations

Some limitations can be listed for this thesis. Firstly, the developed recommender system did not offer participants the ability to search for particular recipes themselves, and explore the interface in an entirely naturalistic setting. Hosting the study on crowd-sourcing platforms such as mTurk and Prolific also introduces some limitations, such as some participants may not have been entirely interested in food. However, hosting on crowd-sourcing platforms allows for a greater sample size. Furthermore, even though food preferences are subjective no recommendations were personalized to each specific user in this experiment, as various studies have personalized food recommendations with success earlier. Thus, this thesis only focused on analyzing a multi-list approach with recommendations, leaving personalized recommendations to future work in multi-list food recommender systems.

The experiment included an attention check midway in the tasks, where the goal was to identify potential users who submitted random responses both on the selection task and in the questionnaires. The total amount of participants that did not pass the attention check was 55 users. A decision was made to analyze an the data and report the findings both including all users, as well as omitting the users who did not pass the attention check. Making this decision was supported by two factors: Comparisons showed that omitting non-passing users did not make significant difference on the main analyses regarding healthiness of choices and choice difficulty. Differences were however found regarding choice satisfaction, where it was shown that multi-list were most favorable between all users, but when omitting non-passing users single-list interfaces were more preferred. The discrepancy was caused by non-passing users who favored multi-list interfaces highly, compared to single-list interfaces. Furthermore, even if attention checks are common practice in online experiments, studies have shown that attention checks can change the behavior of the participant for the remainder of the study [23].

Even though participants in online surveys and experiments such as the one conducted here are prompted to pay attention and give honest answers, the fact that the experiment is conducted on ones personal computer cannot be overseen. This meaning that notifications, other applications on the computer and social media can steal a users attention while participating in the experiment. While this lack of control of setting can in some ways harm the collected data, introducing any control on this will further distance the study from a realistic setting.

5.2 Future Work

Several aspects could be addressed in future research. Addressing the limitations, future studies should be conducted in a more naturalistic setting, and where users have more control over the search aspect. For example allowing for custom searches on an established recipe website, where users can input a search query and get recommendations presented in a multi-list interface with explanations - as opposed to the approach taken in this thesis with a reference recipe. Furthermore, food-search is a highly subjective aspect, loaded with constructed personal preferences. Accounting for personal preferences trough personalizing recommendations should be investigated, similar to commercial applications in other domains such as Netflix, where several lists in the multi-list interface are personalized on different aspects. This can be done by adapting the multi-list interface with an addition of one or more lists that serve participants recommendations directly based on different factors. Such factors can be specific preferences, like favorable cuisine or ingredients. Furthermore, goals and dietary restrictions can be catered to, by providing lists that recommends recipes that suits a persons diet, as well as eliminating alternatives that are not suitable for the person. Following the personalization of recommendations, personalized explanations should also be explored and investigated if these have a greater potential for persuasiveness.

Investigating larger changes in the choice architecture could yield interesting results. This can be done by investigating how list placement affects user choices and evaluation, as well as emphasizing particular aspects trough making specific lists larger than others. Future research should also pay more attention to *how much* information should be displayed to the user at once, and *how*. As multi-list interfaces introduce choice difficulty because of the larger recommendation set, changes to how information such as ingredients and ingredients are presented should be investigated to see if this can reduce some of the choice difficulty that is being introduced.

Bibliography

[1] G. Agapito, B. Calabrese, P. H. Guzzi, M. Cannataro, M. Simeoni, I. Care, T. Lamprinoudi, G. Fuiano, and A. Pujia. "DIETOS: A recommender system for adaptive diet monitoring and personalized food suggestion". In: *International Conference on Wireless and Mobile Computing, Networking and Communications* (2016). ISSN: 21619654. DOI: 10.1109/WiMOB.2016.7763190.

[2] Muhammad Aljukhadar, Sylvain Senecal, and Charles Etienne Daoust. "Using recommendation agents to cope with information overload". In: *International Journal of Electronic Commerce* 17.2 (2012), pp. 41–70. ISSN: 10864415. DOI: 10.2753/JEC1086-4415170202.

[3] *Allrecipes Press Portal | Find out more about Allrecipes.com*. URL: http://press.allrecipes.com/.

[4] Amazon. *Amazon Mechanlical Turk*. URL: https://www.mturk.com/.

[5] Corey H. Basch, William D. Kernan, and Anthony Menafro. "Presence of Candy and Snack Food at Checkout in Chain Stores: Results of a Pilot Study". In: *Journal of Community Health* 41.5 (2016), pp. 1090–1093. ISSN: 15733610. DOI: 10.1007/s10900-016-0193-7.

[6] James R. Bettman, Mary Frances Luce, and John W. Payne. "Constructive consumer choice processes". In: *Journal of Consumer Research* 25.3 (1998), pp. 187–217. ISSN: 00935301. DOI: 10.1086/209535.

[7] Dirk Bollen, Bart P. Knijnenburg, Martijn C. Willemsen, and Mark Graus. "Understanding choice overload in recommender systems". In: *RecSys'10 - Proceedings of the 4th ACM Conference on Recommender Systems* (2010), pp. 63–70. DOI: 10.1145/1864708.1864724.

[8] Renata Bracale and Concetta M. Vaccaro. "Changes in food choice following restrictive measures due to Covid-19". In: *Nutrition, Metabolism and Cardiovascular Diseases* 30.9 (2020), pp. 1423–1426. ISSN: 15903729. DOI: 10.1016/j.numecd.2020.05.027. URL: https://doi.org/10.1016/j.numecd.2020.05.027.

[9] *Cardiovascular diseases (CVDs)*. URL: https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[10] *Check the label | Food Standards Agency*. URL: https://www.food.gov.uk/safety-hygiene/check-the-label?navref=main.

[11] Li Chen and Pearl Pu. "Eye-tracking study of user behavior in recommender interfaces". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6075 LNCS.June 2010 (2010), pp. 375–380. ISSN: 03029743. DOI: 10.1007/978-3-642-13470-8{_}35.

[12] Paolo Cremonesi, Antonio Donatacci, Franca Garzotto, and Roberto Turrin. "Decision-making in recommender systems: The role of user's goals and bounded resources". In: *CEUR Workshop Proceedings* 893.December 2016 (2012), pp. 1–7. ISSN: 16130073.

[13] Kristin Diehl and Cait Poynor. "Great expectations?! Assortment size, expectations, and satisfaction". In: *Journal of Marketing Research* 47.2 (2010), pp. 312–322. ISSN: 00222437. DOI: 10.1509/jmkr.47.2.312.

[14] Helen Dixon, Maree Scully, and Kristiina Parkinson. "Pester power: Snackfoods displayed at supermarket checkouts in Melbourne, Australia". In: *Health Promotion Journal of Australia* 17.2 (2006), pp. 124–127. ISSN: 10361073. DOI: 10.1071/he06124.

[15] Mehdi Elahi, Mouzhi Ge, Francesco Ricci, David Massimo, and Shlomo Berkovsky. "Interactive food recommendation for groups". In: *CEUR Workshop Proceedings* 1247 (2014), pp. 6–7. ISSN: 16130073.

[16] David Elsweiler, Christoph Trattner, and Morgan Harvey. "Exploiting food choice biases for healthier recipe recommendation". In: *SIGIR 2017 - Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. IEEE Computer Society, May 2017, pp. 575–584. ISBN: 9781450350228. DOI: 10.1145/3077136.3080826.

[17] EUFIC. *Global update on nutrition labelling executive summary*. Tech. rep. February. 2016, pp. 1–10.

[18] Mouzhi Ge, Francesco Ricci, and David Massimo. "Health-aware food recommender system". In: *RecSys 2015 - Proceedings of the 9th ACM Conference on Recommender Systems* (2015), pp. 333–334. DOI: 10.1145/2792838.2796554.

[19] Carlos A. Gomez-Uribe and Neil Hunt. "The netflix recommender system: Algorithms, business value, and innovation". In: *ACM Transactions on Management Information Systems* 6.4 (2015). ISSN: 21586578. DOI: 10.1145/2843948.

[20] Elizabeth Gorbonos, Yang Liu, and Chinh T. Hoang. "NutRec: Nutrition Oriented Online Recipe Recommender". In: *Proceedings - 2018 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2018* (2019), pp. 25–32. DOI: 10.1109/WI.2018.0-111.

[21] F. Maxwell Harper and Joseph A. Konstan. "The movielens datasets: History and context". In: *ACM Transactions on Interactive Intelligent Systems* 5.4 (2015), pp. 1–19. ISSN: 21606463. DOI: 10.1145/2827872.

[22] Morgan Harvey and David Elsweiler. "Automated recommendation of healthy, personalised meal plans". In: *RecSys 2015 - Proceedings of the 9th ACM Conference on Recommender Systems* (2015), pp. 327–328. DOI: 10.1145/2792838.2796551.

[23] David J. Hauser and Norbert Schwarz. "It's a Trap! Instructional Manipulation Checks Prompt Systematic Thinking on "Tricky" Tasks". In: *SAGE Open* 5.2 (2015). ISSN: 21582440. DOI: 10.1177/2158244015584617.

[24] J. L. Herlocker, J. A. Konstan, and J. Riedl. "Explaining collaborative filtering recommendations". In: *Proceedings of the ACM Conference on Computer Supported Cooperative Work* January (2000), pp. 241–250. DOI: 10.1145/358916.358995.

[25] Sheena S. Iyengar and Mark R. Lepper. "When choice is demotivating: Can one desire too much of a good thing?" In: *Journal of Personality and Social Psychology* 79.6 (2000), pp. 995–1006. ISSN: 00223514. DOI: 10.1037/0022-3514.79.6.995.

[26] 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 (UMAP '21), June 21-25, 2021, Utrecht, Netherlands* (2021).

[27] Eric J. Johnson et al. “Beyond nudges: Tools of a choice architecture”. In: *Marketing Letters* 23.2 (2012), pp. 487–504. ISSN: 09230645. DOI: 10.1007/s11002-012-9186-1.

[28] Bart P. Knijnenburg and Martijn C. Willemsen. “Evaluating Recommender Systems with User Experiments”. In: *Recommender Systems Handbook, Second Edition*. 2015. Chap. 9, pp. 309–352. ISBN: 9781489976376. DOI: 10.1007/978-1-4899-7637-6.

[29] Matthew Kreuter W and Laura K Brenan. “Do Nutrition Label Readers Eat Healthier Diets? Behavioral Correlates of Addultls Use of Food Labels”. In: *American Journal of preventive Medicine* 13.4 (1997), pp. 277–283.

[30] Floor M. Kroese, David R. Marchiori, and Denise T.D. De Ridder. “Nudging healthy food choices: A field experiment at the train station”. In: *Journal of Public Health (United Kingdom)* 38.2 (2016), e133–e137. ISSN: 17413850. DOI: 10.1093/pubmed/fdv096.

[31] Nadja Leipold, Mira Madenach, Hanna Schäfer, Martin Lurz, Nada Terzimehić, Georg Groh, Markus Böhm, Kurt Gedrich, and Helmut Krcmar. “Nutrilize a personalized nutrition recommender system: An enable study”. In: *CEUR Workshop Proceedings* 2216 (2018), pp. 24–29. ISSN: 16130073.

[32] Sean M. McNee, John Riedl, and Joseph A. Konstan. “Being accurate is not enough: How accuracy metrics have hurt recommender systems”. In: *Conference on Human Factors in Computing Systems - Proceedings* August (2006), pp. 1097–1101. DOI: 10.1145/1125451.1125659.

[33] Susanna Mills, Heather Brown, Wendy Wrieden, Martin White, and Jean Adams. “Frequency of eating home cooked meals and potential benefits for diet and health: Cross-sectional analysis of a population-based cohort study”. In: *International Journal of Behavioral Nutrition and Physical Activity* 14.1 (2017), pp. 1–11. ISSN: 14795868. DOI: 10.1186/s12966-017-0567-y.

[34] MouselabWEB. *MouselabWEB*. 2021. URL: https://www.mouselabweb.org/index.html.

[35] Blain Murphy, Tony Benson, Amanda McCloat, Elaine Mooney, Chris Elliott, Moira Dean, and Fiona Lavelle. “Changes in consumers’ food practices during the covid-19 lockdown, implications for diet quality and the food system: A cross-continental

comparison". In: *Nutrients* 13.1 (2021), pp. 1–14. ISSN: 20726643. DOI: 10.3390/nu13010020.

[36] Cataldo Musto and Bari Aldo. "Exploring the Effects of Natural Language Justifications in Food Recommender Systems". In: *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21), June 21â•fi25, 2021, Utrecht, Netherlands* 1.1 (2021), pp. 1–16. DOI: 10.1145/3450613.3456827.

[37] Cataldo Musto, Christoph Trattner, Alain Starke, and Giovanni Semeraro. "Towards a Knowledge-aware Food Recommender System Exploiting Holistic User Models". In: *UMAP 2020 - Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (2020), pp. 333–337. DOI: 10.1145/3340631.3394880.

[38] Theodora Nanou, George Lekakos, and Konstantinos Fouskas. "The effects of recommendations' presentation on persuasion and satisfaction in a movie recommender system". In: *Multimedia Systems* 16.4-5 (2010), pp. 219–230. ISSN: 09424962. DOI: 10.1007/s00530-010-0190-0.

[39] Rodolfo M. Nayga, Daria Lipinski, and Nitin Savur. "Consumers' use of nutritional labels while food shopping and at home". In: *Journal of Consumer Affairs* 32.1 (1998), pp. 106–120. ISSN: 00220078. DOI: 10.1111/j.1745-6606.1998.tb00402.x.

[40] Pearl Pu and Li Chen. "Trust-inspiring explanation interfaces for recommender systems". In: *Knowledge-Based Systems* 20.6 (2007), pp. 542–556. ISSN: 09507051. DOI: 10.1016/j.knosys.2007.04.004.

[41] Francesco Ricci, Lior Rokach, and Bracha Shapira. *Recommender Systems Handbook*. 2011, pp. 1–35. ISBN: 9780387858203. DOI: 10.1007/978-0-387-85820-3.

[42] Russell L. Rothman, Ryan Housam, Hilary Weiss, Dianne Davis, Rebecca Gregory, Tebeb Gebretsadik, Ayumi Shintani, and Tom A. Elasy. "Patient Understanding of Food Labels. The Role of Literacy and Numeracy". In: *American Journal of Preventive Medicine* 31.5 (2006), pp. 391–398. ISSN: 07493797. DOI: 10.1016/j.amepre.2006.07.025.

[43] J Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. "Collaborative Filtering Recommender Systems". In: *The Adaptive Web: Methods and Strategies of Web Personalization*. Ed. by Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 291–324. ISBN: 978-3-540-72079-9.

DOI: 10.1007/978-3-540-72079-9{_}9. URL: https://doi.org/10.1007/978-3-540-72079-9_9.

[44] Barry Schwartz, Andrew Ward, Sonja Lyubomirsky, John Monterosso, Katherine White, and Darrin R. Lehman. "Maximizing versus satisficing: Happiness is a matter of choice". In: *Journal of Personality and Social Psychology* 83.5 (2002), pp. 1178–1197. ISSN: 00223514. DOI: 10.1037/0022-3514.83.5.1178.

[45] Angela Shine, Seamus O'Reilly, and Kathleen O'Sullivan. "Consumer use of nutrition labels". In: *British Food Journal* 99.8 (1997), pp. 290–296. ISSN: 0007070X. DOI: 10.1108/00070709710188390.

[46] Alain D Starke and Christoph Trattner. "Promoting Healthy Food Choices Online : A Case for Multi-List Recommender Systems". In: (), pp. 7–9.

[47] Alain D Starke, Martijn C Willemsen, and Christoph Trattner. "Nudging Healthy Choices in Food Search Through Visual Attractiveness". In: *Frontiers in Artificial Intelligence* 4 (2021), p. 20. ISSN: 2624-8212. DOI: 10.3389/frai.2021.621743. URL: https://www.frontiersin.org/article/10.3389/frai.2021.621743.

[48] Panagiotis Symeonidis, Alexandros Nanopoulos, and Yannis Manolopoulos. "MoviExplain: A recommender system with explanations". In: *RecSys'09 - Proceedings of the 3rd ACM Conference on Recommender Systems* (2009), pp. 317–320. DOI: 10.1145/1639714.1639777.

[49] Richard H. Thaler and Cass R. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. 2008. ISBN: 9780300122237. DOI: 10.1016/s1477-3880(15)30073-6.

[50] Nava Tintarev and Judith Masthoff. "A survey of explanations in recommender systems". In: *Proceedings - International Conference on Data Engineering* May (2007), pp. 801–810. ISSN: 10844627. DOI: 10.1109/ICDEW.2007.4401070.

[51] Nava Tintarev and Judith Masthoff. *Designing and evaluating explanations*. 2011, pp. 479–510. ISBN: 9780387858203. DOI: 10.1007/978-0-387-85820-3.

[52] Nava Tintarev and Judith Masthoff. "Evaluating the effectiveness of explanations for recommender systems: Methodological issues and empirical studies on the impact of personalization". In: *User Modeling and User-Adapted Interaction* 22.4-5 (2012), pp. 399–439. ISSN: 09241868. DOI: 10.1007/s11257-011-9117-5.

[53] Christoph Trattner and David Elsweiler. “An evaluation of recommendation algorithms for online recipe portals”. In: *CEUR Workshop Proceedings* 2439 (2019), pp. 24–28. ISSN: 16130073.

[54] Christoph Trattner and David Elsweiler. “Food Recommender Systems: Important Contributions, Challenges and Future Research Directions”. In: (2017), pp. 1–16. URL: http://arxiv.org/abs/1711.02760.

[55] Christoph Trattner and David Elsweiler. “Investigating the healthiness of internet-sourced recipes implications for meal planning and recommender systems”. In: *26th International World Wide Web Conference, WWW 2017* (2017), pp. 489–498. DOI: 10.1145/3038912.3052573.

[56] Christoph Trattner and Dietmar Jannach. “Learning to recommend similar items from human judgments”. In: *User Modeling and User-Adapted Interaction* D (2019). ISSN: 15731391. DOI: 10.1007/s11257-019-09245-4.

[57] L. C. Van Gestel, F. M. Kroese, and D. T.D. De Ridder. “Nudging at the checkout counter–A longitudinal study of the effect of a food repositioning nudge on healthy food choice”. In: *Psychology and Health* 33.6 (2018), pp. 800–809. ISSN: 14768321. DOI: 10.1080/08870446.2017.1416116. URL: https://doi.org/10.1080/08870446.2017.1416116.

[58] Jesse Vig, Shilad Sen, and John Riedl. “Tagsplanations: Explaining recommendations using tags”. In: *International Conference on Intelligent User Interfaces, Proceedings IUI* (2009), pp. 47–56. DOI: 10.1145/1502650.1502661.

[59] WHO. *Global Report on Diabetes*. Tech. rep. 2016, pp. 6–86. URL: http://www.who.int/about/licensing/copyright_form/index.html%0Ahttp://www.who.int/about/licensing/copyright_form/index.html%0Ahttps://apps.who.int/iris/handle/10665/204871%0Ahttp://www.who.int/about/licensing/.

[60] Raciel Yera Toledo, Ahmad A. Alzahrani, and Luis Martinez. “A food recommender system considering nutritional information and user preferences”. In: *IEEE Access* 7 (2019), pp. 96695–96711. ISSN: 21693536. DOI: 10.1109/ACCESS.2019.2929413.