UNIVERSITY OF BERGEN

DEPARTMENT OF INFORMATION SCIENCE AND MEDIA STUDIES

MASTER'S THESIS

Nudging Healthy Choices in Food Search Through Front-of-Pack Nutrition Labels

Author: Jørgen Nyborg-Christensen
Supervisor: Prof. Dr. Christoph Trattner
Co-supervisor: Dr. Alain Starke

June 1, 2021

Acknowledgements

I would like to thank my supervisor Prof. Dr. Christoph Trattner for giving me a chance to undertake this research, and providing data, code and financial support. I would also like to thank my co-supervisor Dr. Alain Starke for his guidance, support, and scientific insights throughout this entire thesis project. I am grateful for the grant I received from the DARS Research Group (https://dars.uib.no/) and the Faculty of Social Sciences (UiB), which provided the financial means to run this study.

Jørgen Nyborg-Christensen
Bergen, Norway, May 2021

Abstract

Front-of-pack nutrition labels have been developed to help consumers make healthier food choices when shopping for groceries in supermarkets by making it easier to judge the healthiness of food products. The two most promising front-of-pack nutrition labels are the Multiple Traffic Lights label and the Nutri-Score label. Drawing upon research on front-of-pack nutrition labels and nudges, this thesis investigates whether healthy food choices can be supported in food search by depicting front-of-pack nutrition labels on the recipe card, as well as by re-ranking search results on health. We created a prototype and asked 728 users to search for recipes using predefined keywords and to select the recipes they liked the most. Our analyses revealed that users tended to choose a healthier recipe if either a Multiple Traffic Light or Nutri-Score label was depicted on the recipe card, relative to a no-label control. In addition to this, front-of-pack nutrition labels did not negatively impact user evaluation aspects such as choice satisfaction and choice difficulty. Furthermore, re-ranking recipes using a simple health ranking did not affect the healthiness of the chosen recipes.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Poor health due to a lack of understanding of nutrition is a major problem throughout the world. According to the World Health Organization, more than 1.9 billion adults were overweight in 2016, of these over 650 million were obese [73]. Overweight and obesity can cause many chronic diseases, and reduce a person's overall well-being [52].

Many of today's food choices are made online, often through recipe websites that allow users to explore new recipes to cook at home. Popular recipes on these websites tend to be unhealthy [64], and users find it difficult to judge the healthiness of online recipes [21]. Due to the complexity of food choices and the danger of living an unhealthy lifestyle, new approaches are needed to assist users of online recipe websites in making healthier food choices.

Front-of-Pack nutrition labels have been proposed as a possible solution to help consumers make healthier food choices, by making it easier to interpret and understand the nutritional content of food products. The most promising front-of-pack nutrition labels are the Nutri-Score and Multiple Traffic Light (MTL) labels. These labels have proven to be among the most efficient with regards to helping shoppers distinguish between healthy and unhealthy food products. MTL is a nutrient-specific label that incorporates traffic lights that indicate if the food has a high (red), medium (orange), or low (green) content of saturated fat, sugar, sodium and fat. Nutri-Score is a summary label indicating a product's healthiness with letters from A (healthiest) to E (least healthy) and colours from green to red indicating the healthiness of the product. The majority of studies indicate that these labels help shoppers to distinguish between healthy and less healthy foods [17, 23]. These labels are commonly used for food products, and have been examined in both offline and online supermarkets. However, no studies have applied front-of-pack nutrition labels in the context of online recipe retrieval.

Positioning effects have also been shown to be effective in nudging consumers towards healthier food products. Dayan and Bar-Hillel [14] found that placing healthier recipes at the top or bottom of a menu and less healthy recipes in their centre could potentially result in some increase in favour of healthier recipe choices. Results from Wansik and Hanks [69] suggest that by changing the presentation order of buffet foods and rearranging food order from healthiest to least healthy can nudge diners toward a healthier meal. To this end, we examine whether the same effect would apply to the choices made in a recipe website, by comparing a health ranking of recipes against a popular and random ranking of the same recipes.

1.2 Problem Statement

When searching for recipes on the internet it can be hard to judge the healthiness of these recipes and make healthy choices. Nutritional information is often not displayed on the recipe cards in the search results, as depicted in Figure 1.1. By adding a Nutri-Score or Multiple Traffic Light label to the recipe card we want to make it easier for users to make healthy recipe choices. Figure 1.2 depicts a recipe card with a Nutri-Score label used in the current study. Compared to the recipe in 1.1 the healthiness of this recipe can be easily judged with the Nutri-Score label.

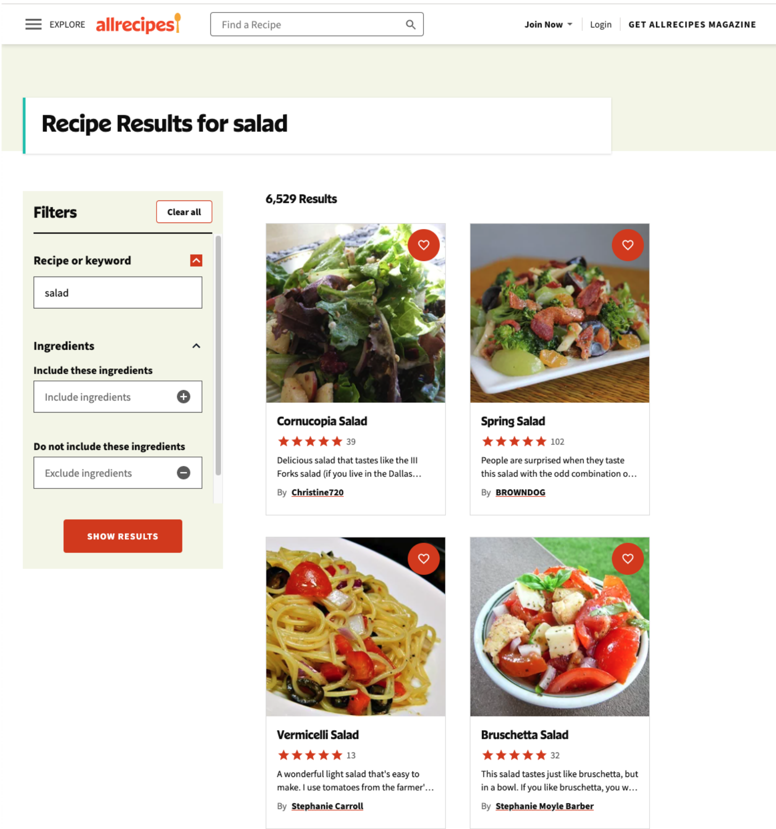


Figure 1.1: Partial search results for a query on salads on the website allrecipes.com. Depicted are summary recipe cards that do not display any nutritional information, only ratings and photos.

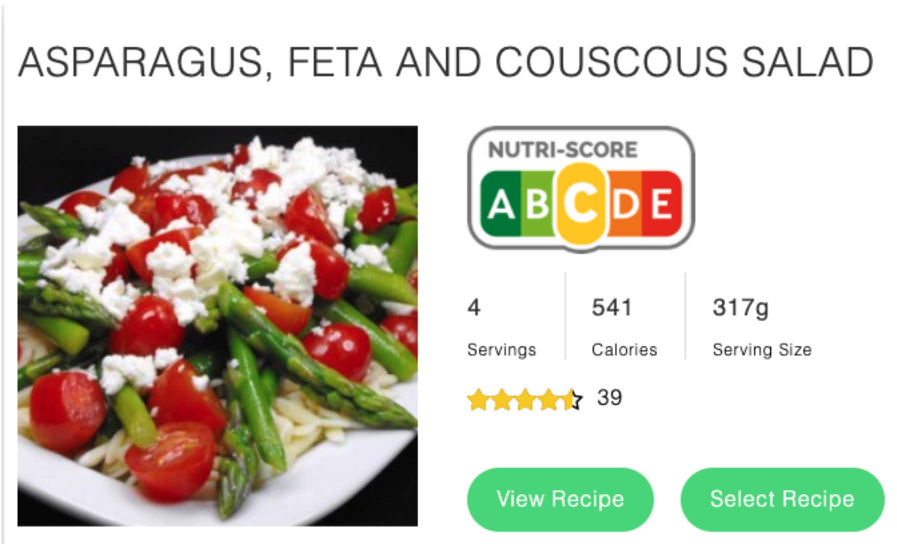


Figure 1.2: A recipe card with a Nutri-Score label using colour and letters (A-E) to signal a recipes healthiness. This label can make it easier for users to judge the healthiness of the recipes in the search results.

1.3 Research Questions

The primary goal of this master thesis is to examine if front-of-pack nutrition labels and a re-ranking of recipes on healthiness can be used to nudge users towards healthier recipes online. We also want to examine how these labels influence user evaluation aspects such as choice satisfaction and choice difficulty when combined with different re-rankings of recipes. To this end, the following research questions are addressed:

- **RQ1:** *To what extent can front-of-pack nutrition labels and re-ranking of search results be used to nudge users towards healthier recipes online?* Based on the literature, the Nutri-Score and Multiple Traffic Light label were selected. In section 4.1, analyses are conducted to examine the effect of these labels and re-ranking on the healthiness of the selected recipes.
- **RQ2:** *Can the presence of a front-of-pack nutrition label in combination with a health ranking decrease choice difficulty?* In section 4.2, analyses are conducted to examine the effect of front-of-pack nutrition labels and ranking on the participant's perceived choice difficulty when selecting recipes in an online food search context.
- **RQ3:** *To what extent can front-of-pack nutrition labels influence choice satisfaction?* In section 4.1, analyses are conducted to examine the effect of front-of-pack nutrition labels and ranking on the participant's perceived choice satisfaction when selecting recipes in an online food search context.

1.4 Thesis Outline

This master thesis is split into five chapters. This introduction chapter details the research question, motivation and contribution of this thesis. Chapter 2 reviews related work, such as food search, front-of-pack nutrition labels, nudging and food choices. Chapter 3 describes the materials and methods used in this thesis, such as the data being used, the development of a recipe search prototype, research design, procedure and measurements. Chapter 4 presents the results of the study, showing promising results for the use of front-of-pack nutrition to nudge people towards healthier recipes online. Chapter 5 discusses the results of the study, limitations, and future research directions.

Chapter 2

Background

This chapter gives an overview of previous work relevant to the context of this thesis and is divided into six sections.

- Section 2.1 describes the challenges of searching for healthy recipes and details common approaches.
- Section 2.2 gives an overview of nudging, and describes framing effects, positioning effects and popularity effects.
- Section 2.3 describes front-of-pack nutrition labels, specifically Nutri-Score and Multiple Traffic Lights label (MTL).
- Section 2.4 describes some of the challenges related to making healthy food choices online.
- Section 2.5 elaborates on the difference between previous work and the research discussed in this thesis.
- Section 2.6 concludes the chapter and details the contributions of this thesis.

2.1 Food Search

Searching for recipes online is a common task for many people in an increasingly digitalized society. Food choices are based on many different types of judgments, such as sensory information, beliefs about the healthiness of the food, and previous experiences when eating the food [13]. These judgments can be at odds with each other, making food choices complex. In an online context users typically considers features such as a recipe's title, ingredients, cooking directions and images [53]. Approaches to online food search can be divided into two categories, personalized and non-personalized.

2.1.1 Personalized Food Search

Personalized food search involves personalizing content without using search queries, such as recommending recipes that are similar to recipes selected in the past [62]. Personalized food search typically involves food recommender systems, which traditionally rely on two distinct approaches to deliver personalized recipe recommendations that generally cater to a user's past preferences. In content-based approaches, the system learns to recommend items that are similar to the items the user liked in the past [51]. For food recommender systems this has been used to tailor recommendations to user's individual tastes. The system recommends recipes to users by breaking recipes into individual ingredients, and scoring them according to the users rating of those recipes [24]. For example, if the user likes recipes with garlic, more recipes with garlic will be recommended to the user. Systems relying on the collaborative filtering approach recommend items to the user that users with similar tastes liked in the past. The similarity in taste is calculated based on the similarity with other users having similar preferences [27]

As visual properties of stimuli drive food choices [43, 75] images associated with recipes can be exploited, making some content-based approaches more applicable to the food domain than others. Image features such as brightness, colorfulness and sharpness can be used to predict user's food preferences [21]. These image features can also be used to algorithmically nudge users toward recipes with less fat, by selecting replacement recipes based on a predictive model [21].

In the food domain, recommendations that only consider user preferences fall short. When people make food choices they also want to take healthiness into account [28], [21]. This can be difficult as internet-sourced recipes tend to be high fat and calorie content [64]. Recent studies suggest that it is possible to optimize for healthier recipes, by incorporating health-related characteristics to generate personalized recipe suggestions [45]. However, algorithmic approaches that focus on changing what is recommended may fall short if users are not aware of the nutritional quality of the recommended items. To overcome this difficulty more work is needed on how to present and explain health-related attributes of recipe recommendations to users.

2.1.2 Non-personalized Food Search

Non-personalized approaches typically rely on the popularity of recipes, derived from user's ratings. A lot of research has been devoted to optimizing the retrieval of recipes, particularly with regards to increasing the healthiness of the recipes displayed to the user [21, 61, 62, 67]. Less attention is devoted to increasing user's understanding of the healthiness of the recipes in the search results. Pecune, Callebert and Marsella [47] investigated whether displaying a healthy tag on recipe cards in the search results would influence people's decision-making. The study suggests that by explicitly informing people how unhealthy some recipes are, it's possible to consciously change their eating habits and prevent them from picking unhealthy recipes [47]. In this thesis we examine to what extent adding front-of-pack nutrition labels to the recipe cards in the search results can nudge them towards healthier recipes.

2.2 Nudging

Nudging refers to strategic changes in a choice environment that are anticipated to alter people's behaviour in a predictable way, without forbidding any options or significantly changing their economic incentives [60, p.6]. In online settings, such as online recipe websites or mobile apps, digital nudging is the use of user-interface design elements to guide people's behaviour in digital choice environments [72]. Nudging builds on psychology and sociology theory that shows how our environment shapes and constrain behaviour. Nudging strategies may be used to promote healthy eating behaviour in both offline and online settings [39].

A widely accepted and cost-effective way to nudge consumer dietary behaviour is by providing nutrition information in the context of daily food selection [11]. Nudging includes a wide variety of strategies to altering social or physical environments to make certain behaviours more likely [42], such as choosing healthier food products. Li and Chapman [39] outlines 6 nudges that can be used to promote healthy behaviour, such as framing effects, default, implementation intentions, position effects, social norms, incentives, and emotions. These approaches to nudging consumers towards healthier food have been implemented in offline settings. For example, changing the size of dishware may reduce portion sizes leading to unconscious changes in actual food intake [57].

Food positioning might also influence food choice. Studies have shown that people eat more unhealthy foods if they are located prominently. With buffet foods, the first ones seen are on average the ones that are selected the most often [69]. Placing healthy foods next to the cash register desk nearly doubled the sales of these foods [38]. However, it is unclear whether minor changes in food position, which are not accompanied by changes in effort, also promote healthier food choices [14].

Predicting the most effective nudges involves trade-offs because predicting the consequences of implementing nudges is not always possible [56]. Historically, most nudges have been developed and researched in offline environments, and may not be directly transferable to a digital context. On the other hand, some nudges such as varying the number of alternatives, the use of defaults, partitioning of options, visual cues, and the ordering of attributes are easily implemented in digital choice environments such as websites. However, since digital nudges rarely lead to changes in effort, they might be less effective than their physical counterparts.

Helping users make better choices is one way of enhancing the user experience in interactive systems. Research shows that the effect sizes of the nudges increase as the focus of the nudges increase from cognition, to affect to behaviour [6]. This suggests that behavioural nudges, such as changing portion sizes (behavioural) are more effective than nutrition labelling (cognition). In the current study, nudging is implemented through visual cues (FOP nutrition labels) and by re-ranking of recipes displayed to the user.

2.2.1 Framing Effect

Different descriptions of the same options can lead to different choices, this is a well-known decision bias known as the framing effect. The idea is to put the emphasis of a persuasive message on the positive or negative consequence of adopting or failing to adopt a particular message [54]. Regarding healthy eating, messages can be framed to highlight either the benefits of eating healthy (a gain-frame), or the consequences of unhealthy eating (a loss-frame)[25]. Nearly all health-related information can be framed in terms of either benefits or costs.

FOP nutrition labels can be seen as a form of visual framing, where the nutritional content of a food product is framed as being either healthy or unhealthy. By summarizing key information about the nutritional information using colors and health scores FOP labels can be used to frame a particular food product as being either healthy or unhealthy. Research show that negatively framed health messages are more persuasive in terms of their effect on intentions and behaviour [26].

2.2.2 Positioning Effect

Serial positioning effects are basic memory phenomena and are among the most robust results in psychology. In a typical study, participants hear or read a list of words and are then asked to recall them. Typically results show that words placed at the beginning and/or end of the ordered list are the easiest to recall [44].

Position effects have been examined with regards to food choice, and Dayan and Bar-

Hillel [14] examined the effect of manipulating the position of different foods on a restaurant menu. They found that items placed at the beginning or the end of the list of their category options were up to twice as popular as when they were placed in the centre of the list. Given this effect, placing healthier recipes at the top or bottom of item lists and less healthy recipes in their centre could potentially result in some increase in favour of healthier recipe choices.

With buffet foods, the first food a buffeteer sees are the ones most selected. Over 75% of diners selected the first food they saw, and the first three foods a person encountered in the buffet comprised 66% of all the foods they took [69]. The same study also found that serving the less healthy foods first led diners to take 31% more total food items. Which foods the buffeteers chose were dramatically determined by the presentation order of the food. Using this approach and rearranging food order from healthiest to least healthy can nudge unknowing or even resistant diners toward a healthier meal, helping make them slim by design [69].

Based on previous work with regards to positioning effects, we included a health ranking in our experiment, where the healthiest recipes as measured by the FSA score or Nutri-Score were placed at the top of the list of search results. The health ranking was compared to both a random ranking and a popularity ranking. The random ranking was used as a baseline.

2.2.3 Popularity Effect

The popularity effect is the consequence of how people use popularity information to make decisions [34]. When people are presented with limited information, they tend to follow what others are doing (popularity) instead of using their own judgement [3]. This is also true for food choices, and displaying popularity information of restaurant dishes can increase people's choice of that dish [7].

Popularity ranking is a common feature of most recipe websites. Search results are often ranked according to their popularity in terms of user ratings and number of ratings. In the food domain, the popularity effect can have detrimental consequences, as popular recipes are often found to be unhealthy [65], and ranking recipes by popularity might nudge users towards less healthy recipes. We want to examine whether front-of-pack nutrition labels in combination with a popularity ranking will make users more aware of the healthiness of the recipes, and not just the popularity.

2.3 Front-of-Pack Nutrition Labels

Interpreting the nutrition content of food products is typically not easy. Back-of-Package (BOP) nutrition labels on food products are difficult to read and understand for many consumers, and most of these consumers don't consider nutritional information while making food choices [8]. The self-reported label use among the general population in the EU is 47% [16]. However, observational in-store studies have found that less than 10% actually use labels when shopping [30].

Different FOP labels have been developed to deal with these issues, and applied to pre-packaged food products to increase consumer awareness of the nutritional quality of food and improve consumer information [20]. FOP labelling has proved successful in helping consumers identify healthier food products [71]. Front-of-Pack nutrition labels are simple graphical labels providing at-a-glance nutritional quality on the primary display panel of foods and beverages, and complement the more detailed nutrient declarations on the back-of-pack label. Front-of-pack (FOP) labelling has been proposed as a potential strategy to improve diet quality and to encourage healthier food choices by making it easier to integrate nutrition into food choices [46].

The various formats of FOP nutrition labels can be organized into two main categories: nutrient-specific and summary indicators. Nutrient-specific labels are either numeric (such as reference Intake Format) and colour coded (such as Multiple Traffic Lights). Summary labels can also be divided into two categories: endorsement schemes (such as Green Keyhole schemes) indicating higher nutritional quality in a given food category and graded indicators, such as the Nutri-Score label depicted in Figure 2.3 [10].

FOP labelling increases the ability of consumers to rank food products according to their healthiness and nutritional quality [23], and enables them to make healthier choices [17]. The health goals of consumers increase attention to and use of nutrition labels, especially when these health goals concern the consumption of specific nutrients [50, 66]. Compared to regular nutrient labels FOP labels enhance healthy food choices, even when consumers are put under time pressure [66].

The most promising FOP labels are Nutri-Score and the Multiple Traffic Lights label. Both of these labels have shown to be the most effective in nudging shoppers towards healthier products, compared against no label control conditions or against other FOP labels in previous studies [41]. The presence of colour in these labels is probably an additional reason for their effectiveness [40]. These labels are generally helpful at enabling shoppers to identify which foods are more healthy and which are less healthy. However, there is little hard evidence that this enhanced knowledge has a significant

Figure 2.1: The Nutri-Score label

impact on actual shopping behaviour [59].

2.3.1 Nutri-Score Label

Nutri-Score is a five-coloured label developed by the French Nutritional Epidemiology Research Team [10] to nudge consumers towards healthier food choices at the point of purchase and as an incentive for manufacturers to create healthier products. The label presents a single summary score representing the overall healthiness of a food/beverage on a five-point color-coded scale from green (best) to red (worst). Nutri-Score relies on the computation of a nutrient profiling score derived from the Food Standards Agency nutrient profiling system (FSA-NPS) [10]. This score is computed using nutrient content per 100 g, and allocates positive points (0-10) for energy (kj), total sugar (g), saturated fat (g) and sodium (mg) content. Negative points (0-5) are allocated for fruit, vegetables and nuts, fibre and protein content. The score is based on a discrete continuous scale from -15 (most healthy) to +40 (least healthy) [10].

In a study comparing four types of labels [18] the Nutri-Score label was considered the easiest to identify, and the most likely to be found easy and quick to understand. Due to the simplicity of the Nutri-Score label, and people need significantly less time to evaluate simpler FOP labels as compared to more complex formats [22]. The Nutri-Score label also had the highest support in the population, and particularly in subjects with low adherence to nutritional recommendations [36].

The Nutri-Score label has not yet been applied to online recipes. However, studies show that the Nutri-Score label increases a consumer's ability to classify products as healthy or unhealthy [23], this is also a challenge with regards to online recipes, where users are unable to judge the healthiness of the recipes [62]. By incorporating nutrition labels into the recipe interface, we can make it easier for people to judge the healthiness of online recipes and make healthier food choices.

2.3.2 Multiple Traffic Light Label

The Multiple Traffic Lights label depicted in Figure 2.2 is a nutrient-specific colour coded label developed in the UK by the Food Standards Agency (FSA) [1]. The label provides an evaluation of the content of energy (kj), fat (g), saturated fat (g), total sugar (g) and salt (g). Colour coding is used to highlight the content of these four nutrients as either low (green), medium (amber), or high (red), according to reference values defined by the UK Food Standards Agency [1]. Table 2.2 shows the nutrition criteria for the colour coding of MTL.

Text	LOW	MEDIUM	HIGH	
Colour code	Green	Amber	Red	
Fat	≤ 3.0g/100g	> 3.0g to ≤ 17.5g/100g	> 17.5g/100g	> 21g/portion
Saturates	≤ 1.5g/100g	> 1.5g to ≤ 5.0g/100g	> 5.0g/100g	> 6.0g/portion
(Total) Sugars	≤ 5.0g/100g	> 5.0g and ≤ 22.5g /100g	> 22.5g/100g	> 27g/portion
Salt	≤ 0.3g/100g	> 0.3g to ≤ 1.5g/100g	>1.5g/100g	>1.8g/portion

Figure 2.2: The Multiple Traffic Light nutrition criteria.

Unlike the Nutri-Score label MTL does not account for positive nutrients or ingredients. However, the MTL label might indicate energy content (which is not evaluated by a colour), and provides numerical information about the nutrient content. The design of MTL is comparable to Nutri-Score in that key information is summarized using colours and easy to understand [59]. Research suggests that MTL may perform better than Nutri-Score if the goal is to reduce total energy intake, because calories are displayed on the label [23]. However, the Nutri-Score label may be prefered if the goal is to improve overall diet quality [23].

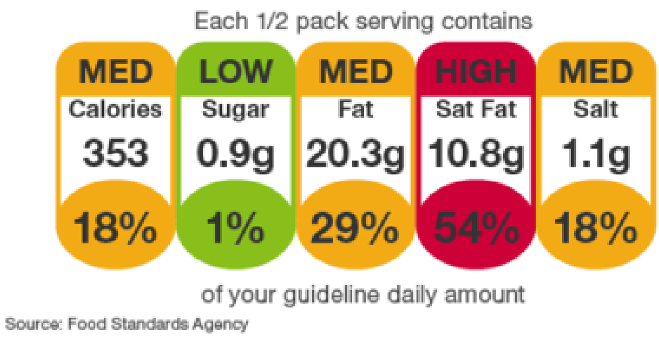


Figure 2.3: Example of a Multiple Traffic Light label combined with GDA information.

2.4 Nutritional Knowledge impact on Food Choice

For people high in nutritional knowledge, the macronutrient content of food is a strong source of evaluative information [13]. And for these individuals the micronutrient composition affects their sensory (taste) or cognitive evaluation (healthiness) of a particular food [13]. For example, people find the sensory qualities of fat appealing, and at the same time believe that eating too much fat is unhealthy [55]. This example highlights the importance of nutritional knowledge, because one might have a very positive sensory evaluation of high-fat food because they like the feel, taste and smell, while at the same time have a negative health evaluation because they know that high-fat foods are relatively unhealthy. One must have some knowledge of food composition and the health implications of this for macronutrient composition to impact health evaluations [13]. Most recipe websites have access to data about the macronutrient content of different foods, but we also need to display this data to the users in a way that helps them make healthy food decisions. Users who are high in food knowledge might be able to judge the healthiness of recipes accurately, but for users with low food knowledge FOP labels can help them judge the healthiness of recipes.

Research shows that individuals who frequently read nutrition labels both value healthy eating and engage in healthy eating behaviour more than individuals who read labels infrequently [4, 29, 37]. Wardle, Parmenter and Waller [70] investigated the relationship between knowledge and intake of fat, fruit and vegetables and found that respondents in the highest quintile for knowledge were almost 25 times more likely to meet current recommendations for fruit, vegetable and fat intake than those in the lowest quintile.

Students who reported to read nutritional labels ate less fast food and added sugar and more fibre, fruits and vegetable compared to student who rarely read labels [29]. Frequent label use might be one way individuals valuing healthy eating translate their intentions into healthy eating behaviour. For students who didn't believe in the importance of preparing healthy meals, "frequent nutritional label use was significantly associated with healthier dietary intake, suggesting that label use may operate independently of nutrition-related attitude in contributing to healthful diet" [29]. These findings indicate the importance of displaying the nutritional contents of food, and the positive effect this can have on healthy eating behaviour. Food recommender system not only needs to recommend healthy recipes, but also show the users why a particular recipe is healthier than another, making it easier to support users in making healthy food choices.

Individuals with a health motivation view the nutritional labels on food products for longer and more often than those who have a taste motivation [68]. Being motivated to eat healthy also stimulated deeper processing of the nutritional information [68]. Another factor influencing whether people pay attention to nutritional information is the design, if the nutrition information is hard to find due to poor design, a health motivated individual will use greater effort to localize it.

Research suggests that processing nutritional labels might not be as straightforward as one might think. Black and Rayner [5] revealed that when consumers are presented with several nutrients simultaneously, they were unable to process information. The Nutri-Score label is designed to help solve this problem, as nutrient information is summarised into a health score from A to E.

There are several decisions consumers need to be able to make using nutrition information of products and websites. They have to identify the amount of a specific nutrient a recipe contains, and assess what counts as a low or high amount of this nutrient. They must be able to compare specific nutrient content/the overall nutrient content of a recipe with other recipes. Lastly, they must be able to assess the recipe in the context of their total daily intake [12]. Front-of-pack nutrition labels can make it easier for users to deal with some of these challenges, by providing summarized nutrient information.

2.5 Summary and Differences

Not much attention has been devoted to help make it easier for users of recipe websites to judge the healthiness of the recipes they are evaluating. Front-of-pack (FOP) labelling has been proposed as a potential strategy to improve diet quality and to encourage healthier food choices by making it easier to integrate nutritional information into these choices. Whereas limited use of FOP labels has emerged on websites with food products, these labels have not yet been investigated in the context of online recipe choices. By displaying the Nutri-Score and MTL labels on the recipe card we believe we can help users make healthier food choices by helping them more accurately judge the healthiness of these recipes. Based on previous work with regards to positioning effects, we also want to examine whether these effects can be used in the context of online food search through a simple re-ranking on health.

2.6 Contribution

Helping users select healthier recipes online is a complex problem. Front-of-pack nutrition labels have been developed and researched for use on pre-packaged food products in supermarkets. Research on front-of-pack nutrition labels has revealed the potential of these labels, and how they can assist consumers in making healthier choices. This thesis extends the research on front-of-pack nutrition labels, by applying them to online recipes. Currently, no studies exist where front-of-pack labels have been applied to nudge users towards healthier recipes online. The contributions are as follows:

- The main contribution of this thesis is to examine whether front-of-pack nutrition labels can be used to nudge users towards healthier recipes in online recipe search. Results show that participants in our study selected healthier recipes when a Multiple Traffic Light label or a Nutri-Score label was displayed on the recipe cards in the search results, suggesting that these labels are effective in nudging users towards healthier recipes online.
- We also provide insights into how front-of-pack labels affect user evaluation aspects such as choice difficulty and choice satisfaction. No work exists that has measured the Nutri-Score and Multiple Traffic Light labels on these metrics. We found that adding a Nutri-Score or Multiple Traffic Light label to the recipe card in the search results did not negatively affect perceived choice difficulty or choice satisfaction.
- In addition to this we also provide insights into how front-of-pack labels can be combined with different re-rankings of recipes on popularity, health, and random ranking of recipes. Results show that a simple health ranking did not nudge users toward healthier recipes. Participants in our study choose healthier recipes when the Multiple Traffic Light and Nutri-Score label was combined with all re-rankings, suggesting that these labels are efficient no matter how search results are ranked.

Chapter 3

Methodology

To answer the research questions we designed an experiment investigating whether the presence of front-of-pack nutrition labels in combination with a health re-ranking can be used to nudge users towards healthier recipes online. This chapter describes the data and methods used in the current study and is split into six sections. Section 3.1 describes the dataset used in the study, and how recipes were selected. Section 3.2 describes the web-based prototype used in the user study. It begins by describing the search functionality, recipe interface, and how the labels for this study were applied to the recipe cards. Section 3.3 elaborates on the research design, procedure, participants recruited for the study, and which variables were measured in the study. Section 3.5 describes the statistical methods used to analyze the data from the user study. Lastly, section 3.6 describes how the data from the user study was processed before running the statistical analysis.

3.1 Dataset

We used a dataset containing recipes from Allrecipes.com, also used by [21, 58, 64]. As of August 2020, Allrecipes.com is the most visited recipe website on the web, receiving an estimated 25 million unique monthly visitors [19]. The dataset consists of 58,263 main dish recipes, along with information about popularity ratings, macronutrients per 100g, and cooking directions. The current study used features regarding the nutritional content of recipes (fat, sodium, sugar, calories, fibre, protein), image link, title, list of ingredients, number of ingredients, preparation steps, preparation and cooking times, number of serving, average ratings and the number of ratings.

Recipe Selection

From the full dataset, a subset of 60 recipes were selected based on keyword matching for one out of three keywords: “Chicken”, “Pasta” and “Salad”. For each keyword we wrote SQL queries extracting all recipes where the keyword was in the recipe title, the average rating was higher than 3.5 and the total number of ratings was higher than 30. From this subset, we picked a total of 20 recipes for each keyword (60 recipes in total) that were similar on most attributes but differed in terms of healthiness. We wanted to keep the recipes the same across conditions and ensure that the recipes all the participants interacted with would be the same. Recipes that did not include all the relevant metadata (e.g., title, image, macronutrients) were excluded.

Table 3.1 shows the recipe selected for this study, including FSA score, Nutri-Score, fat, saturated fat, sugar, sodium, and the average ratings. The mean FSA score of all the recipes was 6.71, the healthiest recipes had an FSA score of 4, and the unhealthiest recipe an FSA score of 10.

Calculating Recipe Healthiness

Similar to [58, 62], we used the "traffic lights" system of the UK Food Standard Agency to compute a recipes healthiness. The FSA provides standard ranges for low content (green), medium content (orange), or high content (red) of fat, saturates, sugar and sodium. To calculate a health score, we assign points for each of the fat, saturates, sugar and sodium elements, one point if the element’s quantity is within the low range, 2 for the medium range, and 3 for the high range. The health score therefore ranges from 4 (best) to 12 (worst).

The recipes in our dataset are not too unhealthy: the health score ranges from 4 to 10, and 31.7% of the recipes have a health score of 8, 9 or 10. 21.6% of the recipes have a health score of 4 or 5, and 46.7% of recipes a health score of 6 or 7.

We also calculated the Nutri-Score which relies on the computation of a nutrient profiling score derived from the Food Standards Agency nutrient profiling system (FSA-NPS) [10]. This score is computed using nutrient content per 100 g, and allocates positive points (0-10) for energy (kj), total sugar (g), saturated fat (g), and sodium (mg) content. Negative points (0-5) are allocated for fruit, vegetables and nuts, fibre and protein content. The score is based on a discrete continuous scale from -15 (most healthy) to +40 (least healthy) [10]. Because the amount of fruit, vegetables, and nuts were difficult to extract from the recipes, we only gave negative points for fibre and protein content.

Table 3.1: Selection of recipes from Allrecipes.com used in the current study. All nutrient values are per 100g.

Recipe Title	FSA Score	Nutri-Score	Fat	Saturated Fat	Sugar	Sodium	Average Rating
Chinese Chicken Salad	7	A	11.49	1.89	5.05	0.27	4.56
Easy Honey Mustard Mozzarella Chicken	8	D	4.56	1.85	19.60	0.31	4.43
Easy Garlic Broiled Chicken	9	E	19.98	9.87	0.15	0.48	4.48
Easy Chicken Rice Casserole	4	A	2.91	0.98	1.44	0.24	3.89
General Tsao's Chicken	7	D	9.30	1.33	6.06	0.47	4.06
Chicken Tikka Masala	9	E	9.68	5.66	2.37	1.50	4.43
Chicken Breasts with Herb Basting Sauce	6	C	9.06	1.94	0.04	0.16	4.43
Cheddar Baked Chicken	8	E	16.49	9.70	1.04	0.64	4.62
Hawaiian Chicken Salad	7	B	13.70	4.32	7.06	0.26	4.34
Incredibly Easy Chicken and Noodles	6	D	3.93	1.01	0.86	0.30	4.40
Scrumptious Chicken Vegetable Stew	5	B	3.07	0.83	1.20	0.07	4.66
Buffalo Chicken Burgers with Blue Cheese Dressing	6	C	3.18	1.01	1.04	0.40	4.54
Best Baked Chicken	8	E	11.11	6.31	1	0.73	4.46
Thai Red Chicken Curry	6	C	4.86	1.96	1.10	0.04	3.90
Baked Lemon Chicken with Mushroom Sauce	6	C	6.42	2.96	0.34	0.09	4.16
Oven Fried Chicken IV	6	C	6.16	2.42	1.43	0.13	3.96
Amy's Garlic Egg Chicken	8	E	14.16	7.90	1.24	0.77	4.49
Amazing Italian Lemon Butter Chicken	7	D	16.15	9.13	0.79	0.22	4.24
Restaurant Style Chicken Nachos	7	B	14.01	4.41	1.30	0.36	4.57
Famous Butter Chicken	8	D	16.65	8.86	0.58	0.34	4.55
Lemon Pepper Pasta	5	B	6.35	0.90	2.57	0	3.87
Mizithra Browned Butter Pasta	10	E	25.76	15.03	7.04	1.11	4.71
Greek Pasta Salad I	7	D	13.54	3.59	1.30	0.30	4.54
Greek Pasta Salad	6	D	16.31	3.59	1.93	0.17	4.54
Onion Pasta	6	C	11.71	3.26	1.66	0.04	4.40
Pizza Pasta	7	D	9.11	3.75	4.01	0.36	4.42
Tomato Basil Pasta	6	D	13.49	4.67	2.99	0.26	4.40
Pasta Con Broccoli	8	D	24.26	14.91	1.32	0.11	4.10
Pasta Melanzana	6	C	10.46	3.28	1.87	0.15	4.40
Pasta e Fagioli a la Chez Ivano	5	C	3.81	1.10	0.79	0.20	4.73
Cheesy Sausage Pasta	8	D	12.51	5.82	2.03	0.33	4.59
Champagne Shrimp and Pasta	6	C	7.44	3.74	0.82	0.11	4.66
Spinach and Pasta Shells	5	B	4.97	0.81	2.23	0.03	3.82
Spinach Pasta Salad	7	D	8.50	1.74	2.51	0.59	4.66
Teena's Spicy Pesto Chicken and Pasta	6	C	10.97	2.81	0.86	0.15	4.62
Tomato and Garlic Pasta	4	C	2.26	0.46	2.50	0.07	4.36
Asiago Sun-Dried Tomato Pasta	7	D	17.39	9.82	2.10	0.19	4.65
Angel's Pasta	5	C	3.51	1.22	2.23	0.10	4.35
Angel Hair Pasta Chicken	5	C	4.56	0.87	1.30	0.16	4.05
Rich Pasta for the Poor Kitchen	9	E	30.32	18.70	1.43	1.08	4.55
Seafood Pasta Salad	8	D	13.25	2.08	5.09	0.55	4.39
Shrimply Delicious Shrimp Salad	6	C	14.70	2.46	1.20	0.26	4.47
Strawberry Pretzel Salad	8	E	12.55	8.40	20.64	0.23	4.57
Pineapple Pretzel Salad	8	E	13.83	9.28	19.13	0.25	4
Amazing Brown Rice Salad	5	C	4.48	0.56	4.31	0.15	4.04
Garden Pasta Salad	8	D	13.30	2.38	5.14	0.74	4.44
Panzanella Salad	7	D	13.73	3.03	2.21	0.34	4.60
Mandarin Almond Salad	6	D	10.28	1.29	10.22	0	4.60
Fruit Salad in Seconds	5	D	0.51	0.17	12.67	0.01	4.50
Egg Salad I	6	C	10.17	2.78	1.11	0.27	3.89
Chickpea and Quinoa Salad with Lemon and Tahini	5	B	6.91	0.87	2.35	0.067	4.60
Champagne Salad	8	E	12.42	7.59	21	0.03	4.55
Caribbean Sweet Potato Salad	5	C	4.62	0.44	2.64	0.10	4.25
Broccoli Salad II	10	E	26.35	6.38	14.06	0.85	4.85
Broccoli and Ramen Noodle Salad	7	C	15.04	1.74	9.75	0.15	4.56
Beet Salad with Goat Cheese	7	D	14.51	2.78	11.51	0.05	4.64
Balsamic Bleu Cheese Salad	9	E	30	9.45	1.42	0.57	3.88
Asparagus, Feta and Couscous Salad	6	C	5.25	2.36	1.73	0.15	4.48
Pasta Chickpea Salad	5	A	5.19	1.11	1.71	0.24	3.50
Watergate Salad	8	E	10.96	4.15	26.53	0.26	4.67

Due to the recipes being ranked by health using different health scores (FSA and Nutri-Score), we also calculated the WHO score to compare the healthiness of recipes between the health ranking variants for the Nutri-Score and MTL conditions. The WHO score was calculated following the approach of Howard et al. [33].

3.2 Prototype

For the purpose of our study, we developed a search prototype. The prototype consists of a questionnaire inquiring about demographics, pages where participants can search for and browse recipes, and surveys about the recipe the participant chose. The web prototype was made using JavaScript, Node, Express, CSS, HTML, and PostgreSQL and hosted on Heroku. This section will detail the decisions and considerations that went into the development of this prototype.

Three different prototypes were created, each prototype displaying a different label or no label on the recipe card. Condition 1 displayed recipes with a Nutri-Score label [1], condition 2 with a MTL label [2], and condition 3 with no-label [3].

3.2.1 Recipe Search

Users could search for recipes by submitting a predefined keyword in the search bar. If users typed the wrong keyword an error message would appear prompting the user to read the instructions more carefully and enter the correct keyword. When users pressed "Enter", or clicked the search icon 10 recipes would appear in a vertical list below the search bar.

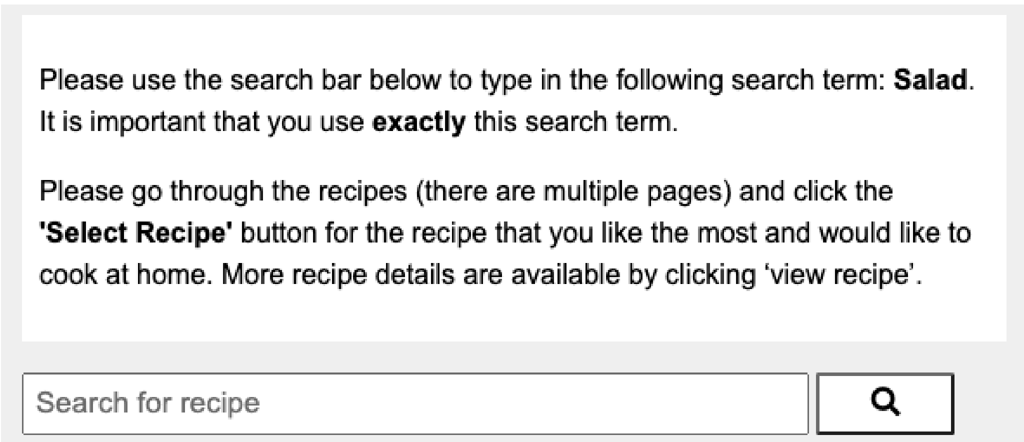


Figure 3.1: The search bar used in the prototype.

We created an Application Programming Interface (API) using Node.js, Express.js, and PostgreSQL to fetch recipes from the database. We created four different endpoints, one to get recipes in random order, one to get recipes ordered by average rating, and two endpoints to get recipes ordered by the FSA and Nutri score. We used SQL ORDER BY

[1] http://recipe-search-1.herokuapp.com/
[2] http://recipe-search-2.herokuapp.com/
[3] http://recipe-search-3.herokuapp.com/

RANDOM () to get the random ordering of the recipes. In conditions 2 and 3 we created an API endpoint that fetched recipes by FSA score in ascending order (from healthiest to least healthy). In condition 1 (Nutri-Score) we fetched recipes by the Nutri-Score and ordered them in ascending order (A-E). We also created an endpoint where we got recipes by popularity, ordered by the average user rating from Allrecipes.com.

3.2.2 Recipe Interface

One aspect of choice architecture especially important for list interfaces is the question of how many options to present. Johnson et al. [35] stress the need to balance two criteria, that more options increase the chance of offering a preferred match to the user, and second that more options lead to a greater cognitive burden through the need to evaluating options. The answer to the question of how many choices to present is also contingent on individual characteristics, further complicating the issue. However Johnson et al. [35] outline some general guidelines: one wants the fewest number of options that will encourage a reasoned consideration of trade-offs amongst conflicting values, while not overwhelming the decision-maker. Yet too few options may generate context-specific preferences, where the presence or absence of one option influences what is chosen. A recommendation to balance these somewhat conflicting considerations "is that four or five non-dominating options may represent reasonable initial values for the choice architecture given these trade-offs" [35]. One could also start with a limited choice set, and allow the decision-maker to increase the number of options if desired. To prevent choice overload, 10 recipes were displayed per page, and the user had the option of switching between pages to see more recipes. The recipe interface is depicted in Figure 3.2.

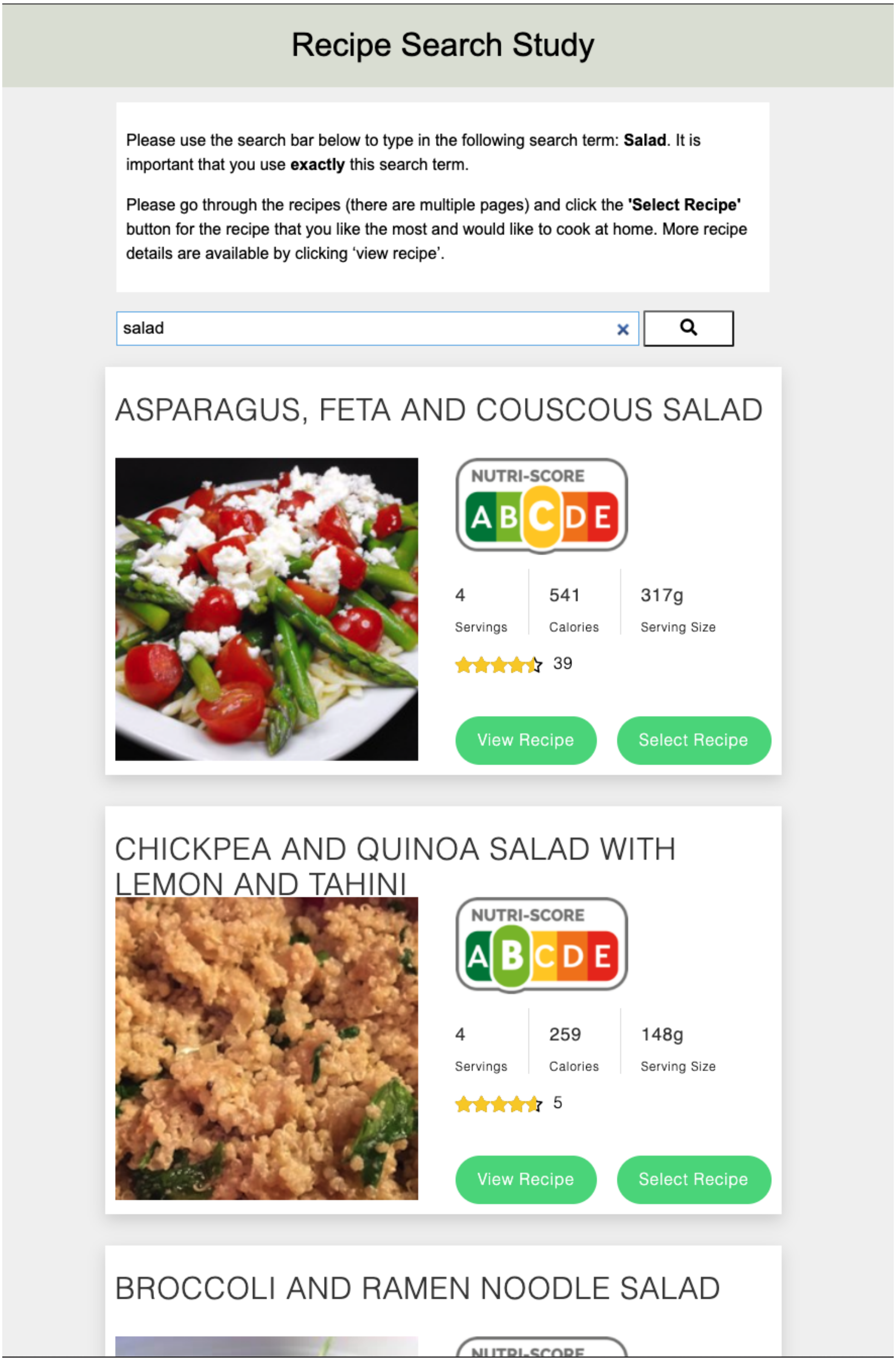


Figure 3.2: Partial screenshot of the prototype/search interface used in the user study. Each query produced a list of 10 recipes. Depicted here are 2.5 recipes in the Nutri-Score condition with a random ranking.

3.2.3 Recipe Labels

The recipes displayed in the search results in our prototype contained either a Nutri-Score, Multiple Traffic Light, or a no-label baseline. Except for the labels, all the recipe cards display the same information: title, image, number of servings, calories per serving, serving size (g), and a button that flips the recipe card allowing the user to view cooking directions and ingredients. The different labels were displayed in the same location on both recipe cards.

Nutri-Score Label

A recipe card with the Nutri-Score label is depicted in 3.3. The images used for the Nutri-Score label were found on the Colruyt Group´s Nutri-Score calculator [4]. We created a JavaScript function that displayed the correct Nutri-Score label image based on the Nutri-Score of each recipe.

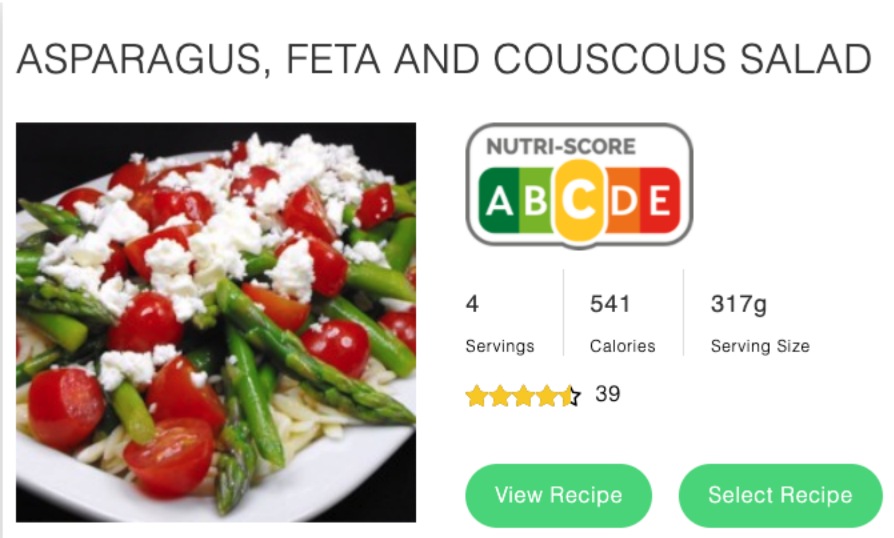


Figure 3.3: Recipe card with a Nutri-Score label.

Multiple Traffic Lights Label

A recipe card with the Multiple Traffic Lights label is depicted in Figure 3.4. We created a MTL label that displayed information about sugar, fat, saturated fat, and salt per 100g. The label had a red, amber and green colour coding, together with the text "HIGH", "MED" and "LOW", to indicated whether a recipe is high in fat, saturated fat, sugar or salt. The more green on the label, the healthier the recipe is.

[4] https://nutriscore.colruytgroup.com/colruytgroup/en/nutri-score-calculator

Figure 3.4: Recipe card with a Multiple Traffic Light label.

No-label

We also included a no-label baseline to be able to examine whether front-of-pack nutrition labels actually lead to healthier recipe choices. A recipe card with no label is depicted in 3.5. This recipe card only display information about the title, image, calories and serving size (g).

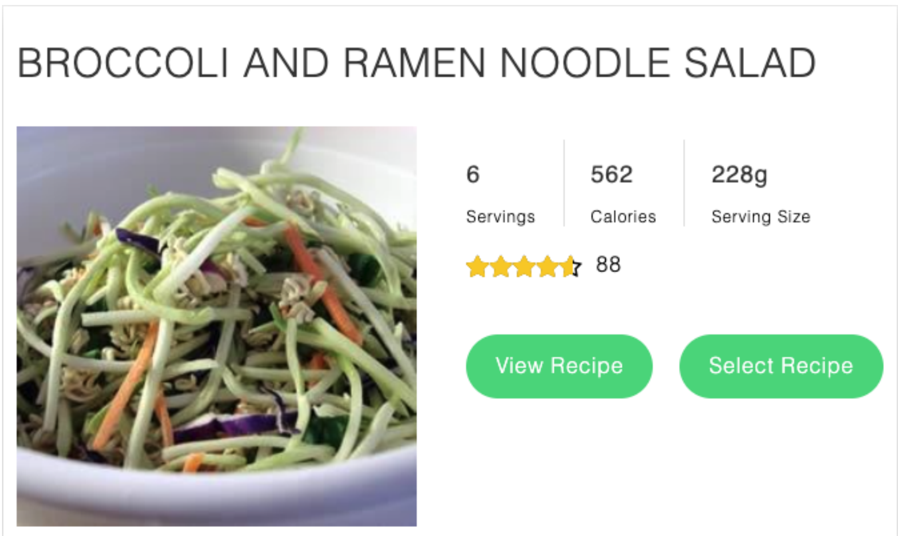


Figure 3.5: Recipe card with no label.

3.3 Research Design

Using the prototype described in the previous section, we performed an online user study to examine whether FOP nutrition labels and health ranking can support healthy food choices.

We identified two different independent variables. The first one represents the label displayed on the recipe card, as a between-subject independent variable, and has three levels: Nutri-Score, Multiple Traffic Lights (MTL), and a no-label baseline. We choose to include the no-label baseline as it is not yet clear from the literature whether front-or-pack style nutrition labels actually nudge users towards healthier recipes.

The second within-subject variable represents how the recipes are displayed to the user and have three levels: a healthy-ranking level in which recipes are ranked according to their healthiness (FSA score or Nutri-Score), a random ranking level and a popularity level in which recipes are ranked according to the rating of the recipe (average All-recipes.com rating). the random ranking level was used as a baseline. Our experiment has a 3x3 mixed between- and within-subjects design with the label as a between-subject variable, and ranking as a within-subject variable.

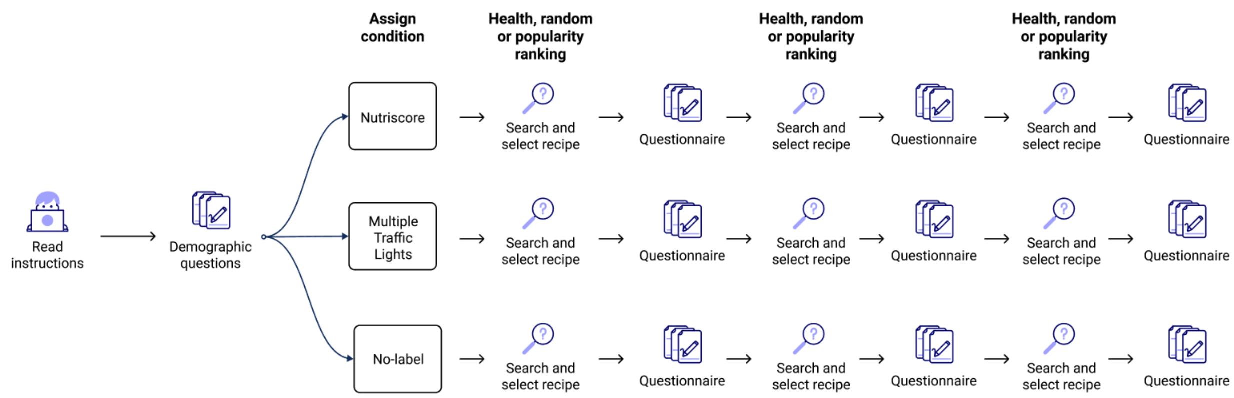


Figure 3.6: The full procedure of the user study, including the within-subject research design. We recruited users to one condition at a time. Participants could only participate in one of the conditions. After each user completed the demographic questionnaire, they were randomly assigned a sequence of variants combining both manipulations ('Label + Rank') and a random list of three keywords to search for. Each search task was followed by a short user experience questionnaire. After the last user experience questionnaire, users were asked which recipe attributes most influence their choices.

3.3.1 Conditions

We created one condition for each of the between-subject independent variables (Nutri-Score, MTL, No-label). This was done to make the manipulation less obvious for the participants and to make it easier to compare the effect the FOP labels had on the selected recipes. Within each condition, users had to search for and select three recipes, each time a new ranking (within-subject variable) was applied to the search results. The ranking is more subtle than the labels, and we did not expect users to become aware of the ranking manipulation between searches (as there was a different keyword for each of the three queries). In each of the three conditions, participants followed the same procedure.

3.4 Procedure

We recruited participants to the study one condition at a time, and participants could only participate in one of the three conditions. The study procedure can be divided into five steps. These steps are instructions, demographic questionnaire, search, a survey about the chosen recipe and completing an end survey. Screenshots from the prototype used in the study can be found in the appendix 5.3. Searching for and selecting a recipe and completing the survey about the chosen recipe was done three times in succession. This section details each of these steps.

3.4.1 Instructions

Participants were presented with a short description of the task: "You will be asked to search for three different recipes using different predefined keywords. Each search will result in a total of 20 recipes, you will be asked to choose the recipe that you like the most and would like to cook at home. In addition, you will be asked to answer a few questions after each recipe you have selected". These instructions were identical across all conditions. The instructions did not include any explanations about how to interpret the Multiple Traffic Light or Nutri-Score label. After having read the instructions participants continued to the demographics questionnaire.

3.4.2 Demographics Questionnaire

To begin with each participant completed a questionnaire, where we inquired about their gender, age, level of education, recipe website usage (5-point scale), cooking experience (5-point scale), food knowledge (5-point scale), dietary goals (i.e., lose/gain weight) and dietary restrictions (i.e., allergies/ vegan). The questions and the scale for the demographics questions are depicted in 3.2. After the participants completed this questionnaire, a random order of the recipe keywords (chicken, salad, pasta) and the order of variants (health, random, popular ranking) was generated and the participants continued to the search task.

3.4.3 Search Phase

All participants had to search for, and select three different recipes during the study. Participants were given the following instructions: "Please use the search bar below to type in the following search term: Pasta. It is important that you use exactly this search term. Please go through the recipes (there are multiple pages) and click the 'Select Recipe' button for the recipe that you like the most and would like to cook at home. More recipe details are available by clicking 'view recipe'". Participants had to search for either chicken, pasta or salad and select 1 of 20 recipes. If the participant used the wrong keyword, an error message would appear prompting the user to read the instructions more carefully and to use the correct keyword.

3.4.4 Survey About Chosen Recipe

Each choice task was followed by a short questionnaire, regarding how satisfied they were with their choice (choice satisfaction) and how easy it was to make this choice (choice difficulty). This questionnaire consisted of six questions, divided into two groups, one for choice satisfaction and one for choice difficulty. The answers for these questions were 5-point Likert items (anchors: -2 = strongly disagree, 2 = strongly agree). The questions for this survey is depicted in 3.2 under "Choice Questionnaire".

3.4.5 End Survey

After completing the last choice questionnaire, participants had to complete a questionnaire with three questions inquiring about how easy it was to judge the healthiness of the recipes, the information provided and which factors influenced their choices the most. These questions and their scales are depicted in 3.2. Question number two was included to measure information sufficiency, which denoted the ability of the user interface to display information of an item to help users with making a decision. The question was adapted from Pu et. al. 2011 [48].

3.4.6 Participants

We recruited participant from both Amazon Mechanical Turk (Mturk)[5] and Prolific[6]. We recruited 363 participants from both platforms, resulting in a total of 726 participants. We recruited an equal amount of participants for each of the three conditions (242). By recruiting participants for both platforms we also able to analyse if there were any difference in FOP nutrition label usage between Americans and Europeans.

[5] https://www.mturk.com/

[6] https://www.prolific.co/

Table 3.2: Questions used in the user study.

Question	Scale
Demographic Questions	
1. What is your age?	Numeric Input
2. What is your highest completed education?	Scale shown in Figure 3.8
3. What is your nationality?	Text Input
4. What is your gender?	Scale shown in Figure 3.7
5. Do you have any dietary restrictions?	Options shown in Figure 5.3 or text input
6. Do you have any dietary goals?	Options shown in Figure 5.3 or text input
7. I consider my cooking experience to be:	Scale shown in Figure 3.9
8. I consider my eating habits to be:	Scale shown in Figure 3.10
Choice Questionnaire	
Choice satisfaction:	
1. I would recommend the chosen recipe to others	Likert scale 1 (Strongly disagree) - 5 (Strongly agree)
2. My chosen recipe could become one of my favorites	Likert scale 1 (Strongly disagree) - 5 (Strongly agree)
3. I think I would enjoy eating the chosen recipe	Likert scale 1 (Strongly disagree) - 5 (Strongly agree)
Choice difficulty:	
4. I changed my mind several times before making a decision	Likert scale 1 (Strongly disagree) - 5 (Strongly agree)
5. It was easy to make this choice	Likert scale 1 (Strongly disagree) - 5 (Strongly agree)
6. Making a choice was overwhelming	Likert scale 1 (Strongly disagree) - 5 (Strongly agree)
End Questionnaire	
1. It was easy to judge the healthiness of the recipes	Likert scale 1 (Strongly disagree) - 5 (Strongly agree)
2. The information provided was sufficient for me to make a decision	Likert scale 1 (Strongly disagree) - 5 (Strongly agree)
3. Which factors influenced your recipe choices the most?	Scale shown in Figures 5.8, 5.9 and 5.10

349 of the participants were Americans, while 318 were from European countries, the rest were from other parts of the world.

To make sure we recruited high-quality workers we set the number of HITs approved above 500 and the HIT approval rate above 95%. Using these qualifications is one of the most effective strategies for optimizing the quality of results from Mturk [2]. Second, the Web application included an attention check. The attention check appeared randomly in any of the three post-choice questionnaires, and a new question appeared asking "What is two plus two?". Only 3 participants did not pass the attention check. The estimated time to complete the user study was 5-7 minutes, and the reimbursement was set to £1 for completing the study.

To join the study participants had to be fluent in English, and participants were discouraged to join the study if they were on a vegetarian or vegan diet. We also excluded participants with allergies from participating in the study. 368 of the participants were male, 355 female, and 5 replied other, as depicted in Figure 3.7 . The mean education was a bachelor's degree, also depicted in Figure 3.8. Figure 3.9 depicts the self-reported cooking experience, and Figure 3.10 depicts a count plot of participants self-reported eating habits. The mean age of the participants was 35.3 years.

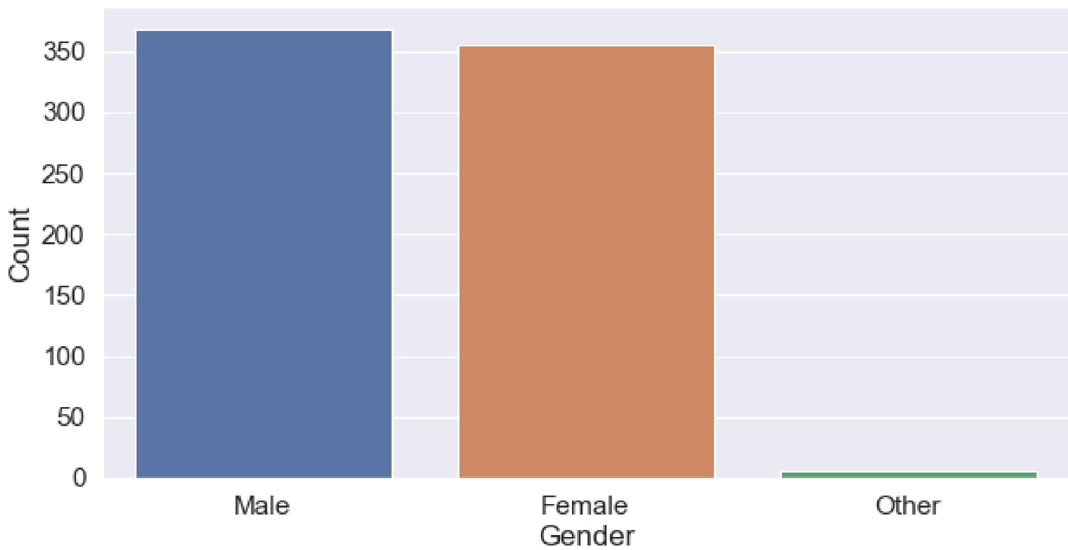


Figure 3.7: Frequency plot of the gender of the participants in the study.

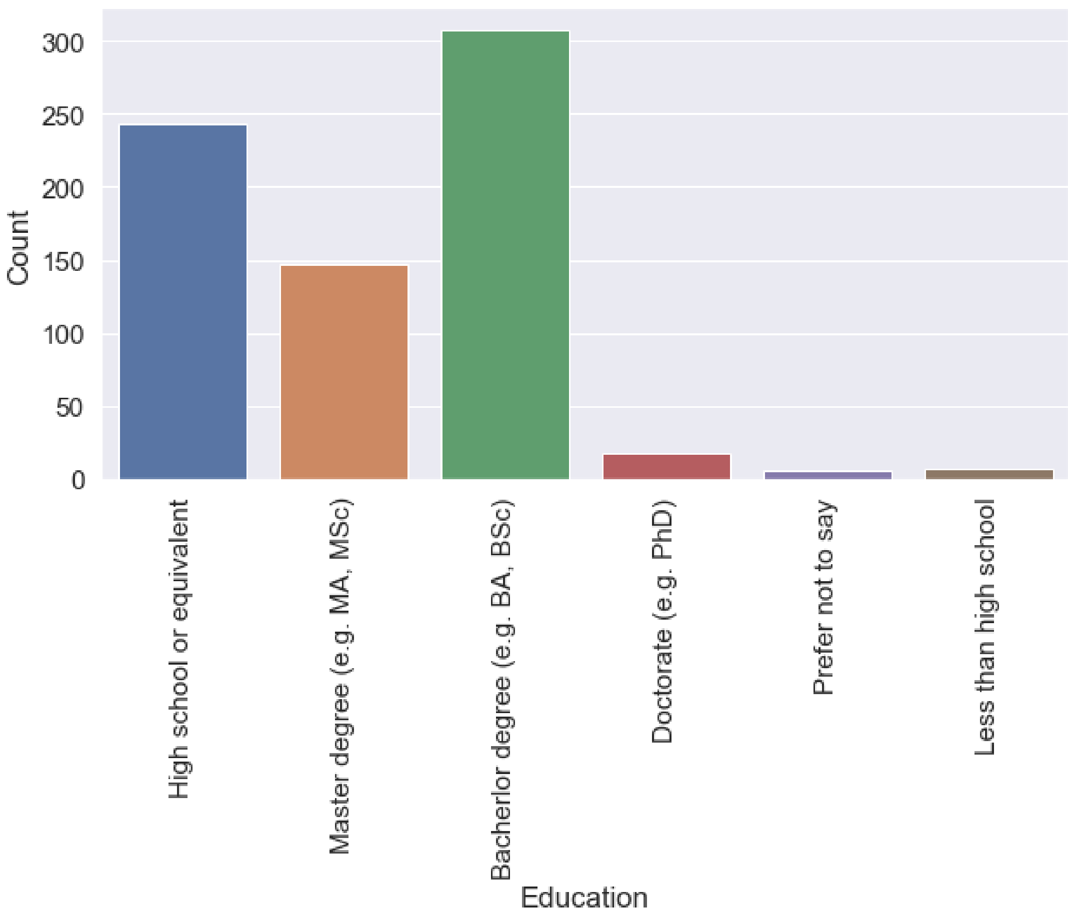


Figure 3.8: Frequency plot of the educational background of the participants in the study.

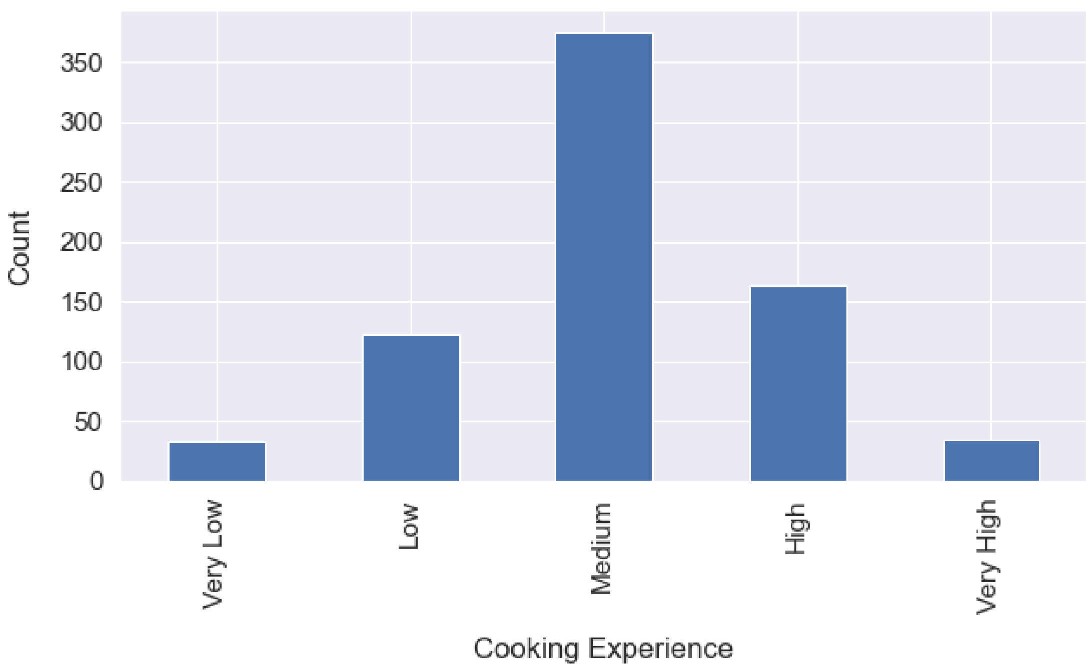


Figure 3.9: Frequency plot of the self-reported cooking experience of the participants in the study.

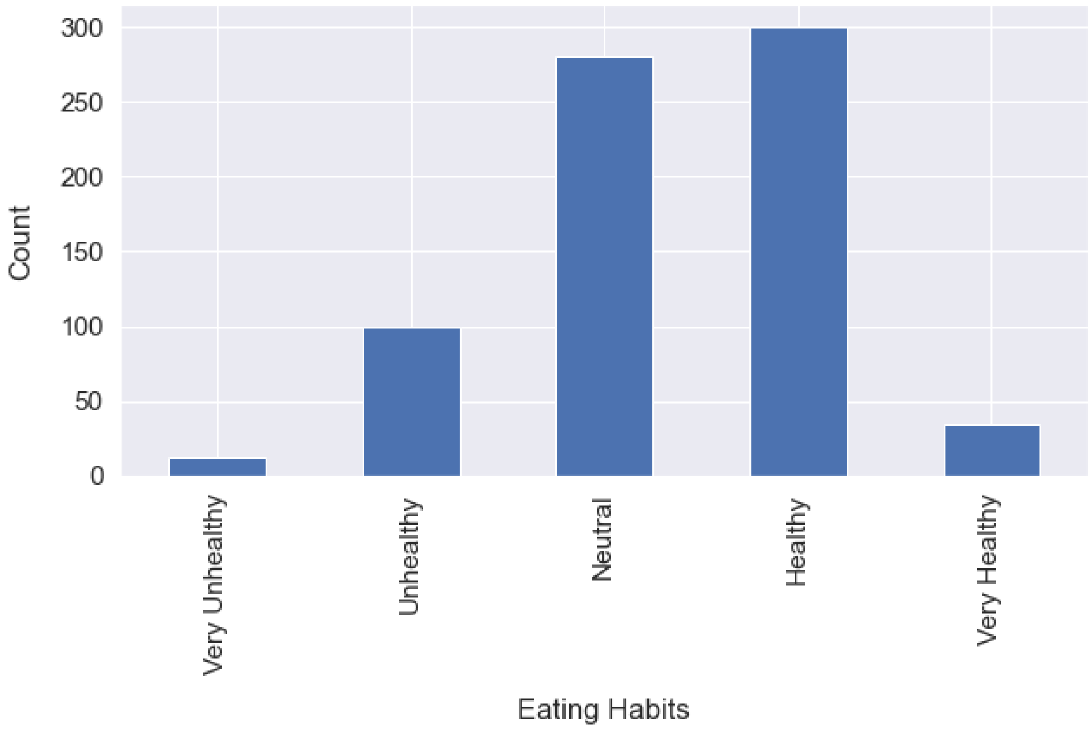


Figure 3.10: Frequency plot of the self-reported eating habits of the participants in the study.

3.4.7 Measures

FSA score

The main variable under investigation was the healthiness of the recipes, as measured by the FSA score. We used it to examine whether different labels or the ranking of the recipes led to changes in the healthiness of the chosen recipe. Features underlying these scores describe a recipe's nutritional content, such as fat, saturated fat, sugar and salt per 100 g.

WHO score

Since recipes were ranked by healthiness, we also needed to be able to compare these recipes. In the MTL and No-label condition recipes were ranked by the FSA score, while in the Nutri-Score condition recipes were ranked by the Nutri-Score. We needed to calculate another measure to be able to compare the different conditions in the health ranking variant. We decided to use the WHO score, also used in other studies [63]. The WHO score is calculated using the approach of Howard et al. [33] who chose the 7 most important (i.e. proteins, carbohydrates, sugars, sodium, fats, saturated fats, and fibres) and their corresponding ranges to determine a so-called WHO health score. The scale ranges from 0 - 7 (0 meaning none of the WHO ranges are fulfilled and 7 meaning all ranges are met). A recipe or meal plan with a WHO score of 7 is interpreted as being very healthy whereas a score of 0 is seen as very unhealthy.

User Characteristics

In the first questionnaire, we inquired about a participant's age, gender, level of education and nationality. We also asked if participants had any dietary restrictions such as vegetarianism, allergies and gluten-free. Participants with allergies or vegans were discouraged to participate in the study. Besides this, we also inquired whether participants had any dietary goals (i.e., eat more protein, lose weight). Lastly, we asked about their self-reported eating habits on a 5-point scale (i.e., very unhealthy to very healthy), as well their cooking experience on a 5-point scale (i.e., very low to very high).

List Characteristics

We also stored data considering the position of each recipe that was displayed in the list of search results. The variable 'Position' referred to the position of a given recipe, with values ranging from 0 to 19, where 0 was the first recipe presented. Since the search results were divided into two pages the participants could navigate between, position 0-9 refer to the recipes in the first list, and 9-19 refers to the recipes on the second page. We also created a variable 'Selected', where the selected recipe had a value of 1 and all the other unselected recipes in the list had a value of 0. We also recorded data about whether the participant pressed the "Load More Recipe" button, but due to a technical error this data was only collected from the first condition of our study.

User Evaluation Aspects

To address RQ2 and RQ3, we inquired about the participant's evaluation of the presented search results in each trial using shot questionnaires consisting of six propositions. Participants were asked to indicate on a 5-point Likert scale (i.e, strongly disagree to strongly agree) to what extent they agreed with each proposition. Three items were designed to capture a user's perceived choice satisfaction, and the other three items were designed to capture perceived choice difficulty. We also inquired about whether participants found the information provided sufficient to make a decision, based on earlier work in [48] designed to inquire about information sufficiency. Lastly, we asked participants about which factors influenced their recipe choice the most (i.e., healthiness, calories, high ratings).

3.5 Statistical Analysis

We used statistical analysis to look for statistically significant variances between conditions and variants. To this end, two main statistical methods are used. The two-way ANOVA was used to analyze how two independent variables (Ranking and Label), in combination, affected the dependent variables (FSA score, choice difficulty and choice satisfaction). For the ANOVA test that returns a statistically significant value ($p < 0.05$), the post-hoc Tukey's Honestly-Significant-Difference (TukeyHSD) test is used to check which groups are different from one another. This section details the statistical methods used to answer each of the research questions, and whether the data used met the assumption for these models.

RQ1: To what extent can front-of-pack nutrition labels and re-ranking of search results be used to nudge users towards healthier recipes online?

We conducted a two-way ANOVA to analyze how two independent variables (Ranking and Label), in combination, affected the dependent variable (FSA score of the chosen recipes). We used the Levene's test to check the homogeneity of variances, the p-value is not less than the significance level of 0.05 ($p = 0.158$). The Shapiro-Wilk test on the ANOVA residuals ($W = 0.96, p = 0.00$) suggests that normality is violated, however, ANOVA test results are often robust to violations of this assumption. The ANOVA test returned a statistically significant value ($p < 0.05$), and a TukeyHSD test was used to check which groups are different from one another. This analysis was also performed without users who reported to have an allergy ($N = 5$) and users who reported that they were vegan ($N = 53$), but this does not significantly affect the results.

To be able to compare the two labels (Nutri-Score and MTL), another two-way ANOVA was conducted to analyze how Ranking and Label, in combination, affected the dependent variable (WHO score of the chosen recipes). In the health ranking variant, these labels were not both ranked by the FSA score since the Nutri-Score condition was ranked by the Nutri-Score (A-E).

We also examined whether personal characteristics in combination with labels had any effect on the FSA score. We used a two-way ANOVA to analyze how two independent variables (Personal characteristics and Label), affected the dependent variable (FSA score of the chosen recipes).

We also created a multiple linear regression model to examine if other features and user characteristics predicted the FSA score of the chosen recipes.

RQ2: Can the presence of a front-of-pack nutrition label in combination with a health ranking decrease choice difficulty?

We conducted a two-way ANOVA to analyze how two independent variables (Ranking and Label), affected the dependent variable (choice difficulty). We used the Brown-Forsythe test to check the homogeneity of variances, the p-value is less than the significance level of 0.05 ($p = 0.005$). The Shapiro-Wilk test on the ANOVA residuals ($W = 0.96, p = 0.00$) suggests that there is a possible violation of normality, however, when plotting the residuals the points fall approximately along the reference line. The post-hoc TukeyHSD test was used to check which groups are different from one another.

RQ3: To what extent can front-of-pack nutrition labels influence choice satisfaction?

We conducted a two-way ANOVA to analyze how two independent variables (Ranking and Label), affected the dependent variable (choice satisfaction). We used the Brown-Forsythe test to check the homogeneity of variances, the p-value is not less than the significance level of 0.05 ($p = 0.056$). The Shapiro-Wilk test on the ANOVA residuals ($W = 0.96, p = 0.00$) suggests that normality is violated, however, when plotting the residuals the points fall approximately along the reference line. The post-hoc TukeyHSD test was used to check which groups are different from one another.

3.6 Data Filtering

We recruited a total of 726 participants in our study. Participants who did not complete the attention check (answer the questions "What is two plus two?" in the second choice survey) were omitted from the data analysis ($N = 6$). We also omitted participants who did not complete the study or selected more or less than 3 recipes ($N = 26$). This might have been caused by a technical error that caused the application to crash during the user study. A total of 694 participants is included in the final analysis.

Chapter 4

Results

We conducted a user study where the participants had to search for and select three recipes they would like to cook at home using different predefined keywords. After each search task participants were asked to answer six questions inquiring about the perceived choice difficulty and choice satisfaction. We examined to what extent front-of-pack nutrition labels, popularity, and positioning effects led to healthier choices, without decreasing user satisfaction or increasing choice difficulty. We recruited 726 participants from Prolific and Amazon Mechanical Turk, of which 694 were included in the final analysis. The following chapter describes the data analysis conducted to answer the research questions.

4.1 RQ1: To what extent can front-of-pack nutrition labels and re-ranking of search results be used to nudge users towards healthier recipes online?

To examine RQ1 a two-way ANOVA was conducted to examine the effect of front-of-pack nutrition labels and ranking on the healthiness of recipes chosen across three different trials. The factors are label (with 3 levels) and ranking (with 3 levels), the factors were ordered as follows: label(no-label, MTL, Nutri-Score) and ranking(random, popular, health). The results of the ANOVA run on 694 participants are reported in table 4.1 .

As reported in Table 4.1 we found no statistically significant interaction between both manipulations on the FSA score of the chosen recipe (i.e., 'Label*Ranking') ($F(4,2073) = 0.25$, $p = .908$). Regarding main effects, the results in our ANOVA supported RQ1, and there was a statistically significant effect for Label ($F(2,2073) = 8.25$,

$p = .0003$) on the healthiness of the chosen recipes. A Tukey post hoc test depicted in Figure 4.2 revealed that there were statistically significant differences for both labels when compared to the baseline no-label condition (MTL $p = 0.001$ and Nutri-Score $p = 0.024$). The results provide support for RQ1, and suggest that FOP labels can be used to nudge users towards healthier recipes. There was no significant main effect for Ranking ($F(2, 2073) = 1.57$, $p = .208$), suggesting that re-ranking recipes on health do not lead to healthier recipes being chosen.

Table 4.1: Result of 2x2 ANOVA on the FSA score of the chosen recipes (three per user).

	Df	Sum Sq	Mean Sq	F value	p
Rank	2	5.84	2.92	1.57	0.2084
Label	2	30.72	15.36	8.25	0.0003
Rank:Label	4	1.88	0.47	0.25	0.9082
Residuals	2073	3859.43	1.86		

Table 4.2: Result of Tukey post-hoc test comparing FSA scores for the different label conditions (no-label, Nutri-Score, MTL).

	Term	Contrast	Null	Estimate	Conf. Low	Conf. High	p
1	Label	No-label-MTL	0.001	0.29	0.12	0.46	0.001
2	Label	Nutriscore-MTL	0.001	0.10	-0.07	0.27	0.349
3	Label	Nutriscore-No-label	0.001	-0.19	-0.36	-0.02	0.024

To determine if the labels actually nudged users toward healthier recipes we have to look at the mean FSA scores for the different conditions (lower FSA score equals healthier recipe choices). The healthiest recipes were selected in the MTL condition ($M = 6.35$), the Nutri-Score condition was second ($M = 6.45$) and the least healthy recipes was selected in the No-label condition ($M = 6.64$). This indicates that FOP labels actually nudged users towards healthier recipes.

We conducted a post-hoc Tukey test on the interaction 'Label*Ranking' to determine if a health rank in combination with a label nudged users towards healthier recipes. Table 4.3 depicts the post-hoc Tukey tests for different combinations of health ranking and labels. There were no statistically significant differences when comparing the combination of labels and health ranking, indicating that a simple health ranking in combination with FOP labels did not nudge users towards healthier recipes.

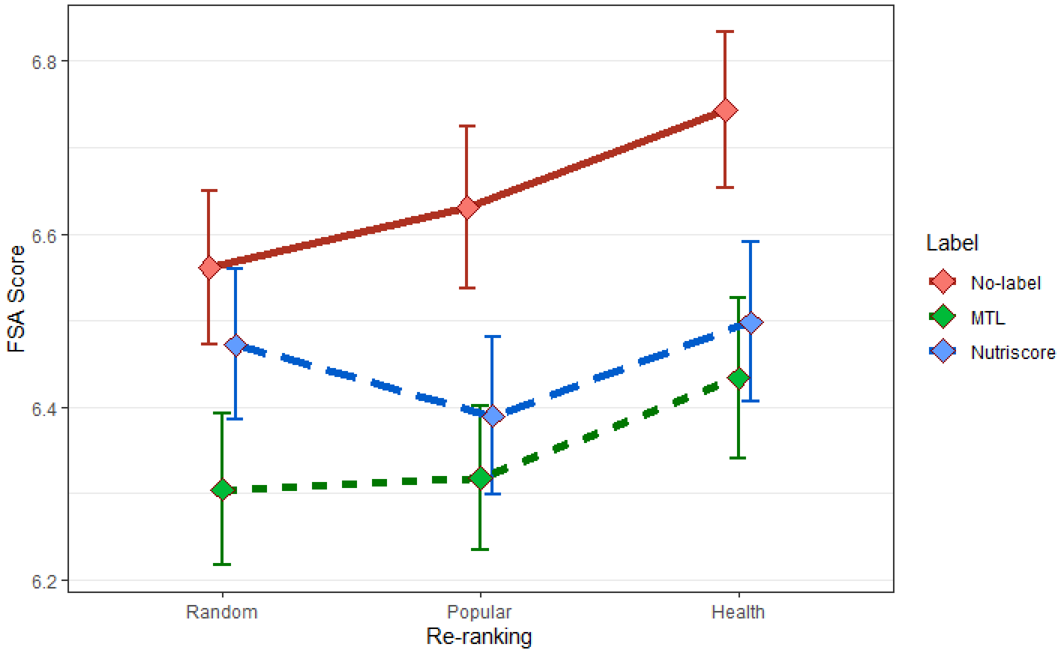


Figure 4.1: Marginal effects plot with error bars (depicting 1 S.E.), between re-ranking of search results and different labels regarding the healthiness (mean FSA score) of the chosen recipes. The least healthy recipes were selected in the no-label condition, and participants selected healthier recipes when a front-of-pack nutrition label was displayed on the recipe card. Ranking recipes by health did not nudge participants towards healthier recipes choices.

*Table 4.3: Post-hos Tukey test on the interaction (Rank*Label)*

	Term	Contrast	Null	Estimate	Conf. Low	Conf. High	*p*
1	Rank:Label	Health:No-label-Random:No-label	0.001	0.18	-0.21	0.58	0.88
2	Rank:Label	Health:MTL-Random:No-label	0.001	-0.13	-0.52	0.27	0.99
3	Rank:Label	Health:Nutriscore-Random:No-label	0.00	-0.06	-0.46	0.33	1.00
4	Rank:Label	Health:No-label-Popular:No-label	0.001	0.11	-0.28	0.51	0.99
5	Rank:Label	Health:MTL-Popular:No-label	0.001	-0.20	-0.59	0.20	0.83
6	Rank:Label	Health:Nutriscore-Popular:No-label	0.00	-0.13	-0.53	0.26	0.98
7	Rank:Label	Health:MTL-Health:No-label	0.001	-0.31	-0.70	0.08	0.26
8	Rank:Label	Health:Nutriscore-Health:No-label	0.00	-0.25	-0.64	0.15	0.59
9	Rank:Label	Health:MTL-Random:MTL	0.00	0.13	-0.26	0.52	0.98
10	Rank:Label	Health:Nutriscore-Random:MTL	0.001	0.19	-0.20	0.59	0.84
11	Rank:Label	Health:MTL-Popular:MTL	0.001	0.12	-0.28	0.51	0.99
12	Rank:Label	Health:Nutriscore-Popular:MTL	0.001	0.18	-0.21	0.57	0.89
13	Rank:Label	Health:Nutriscore-Health:MTL	0.001	0.06	-0.33	0.46	1.00
14	Rank:Label	Health:Nutriscore-Random:Nutriscore	0.001	0.03	-0.37	0.42	1.00
15	Rank:Label	Health:Nutriscore-Popular:Nutriscore	0.001	0.11	-0.29	0.50	1.00

Table 4.4 depicts mean FSA scores for different combinations of ranking and labels. Looking at the means we see that the unhealthiest recipes according to the FSA score was selected in the health ranking condition.

*Table 4.4: Mean FSA score for different combinations of Labels and Ranking (Rank*Label)*

	Label	Rank	N	Mean	SD
1	MTL	Health	233.00	6.43	1.42
2	MTL	Popular	233.00	6.32	1.27
3	MTL	Random	233.00	6.30	1.34
4	No-label	Health	230.00	6.74	1.36
5	No-label	Popular	230.00	6.63	1.42
6	No-label	Random	230.00	6.56	1.35
7	Nutriscore	Health	231.00	6.50	1.40
8	Nutriscore	Popular	231.00	6.39	1.38
9	Nutriscore	Random	231.00	6.47	1.32

Which label is the best with regards to the healthiness's of the chosen recipe?

To figure out which label performed best (if any) we calculated the WHO score for the recipes we used in our experiment. The reason for doing this was that in the Health ranking scenario the recipes were ordered differently, MTL is ranked by FSA (4-12), while Nutri-Score is ranked by (A-E). We needed to have the same health score to compare the two conditions. A recipe with a WHO score of 7 is interpreted as being very healthy whereas a score of 0 is seen as very unhealthy. We conducted a two-way ANOVA to compare the influence of labels and ranking on the WHO score. We performed a post-hoc Tukey to compare the different labels, but the result show that there are no statistically significant difference in WHO score between MTL and Nutri-Score ($p = 0.682$) suggesting that both labels were effective in nudging users towards healthier recipes.

Personal Factors

We also examined whether personal characteristics such as education, cooking experience and food goals in combination with a front-of-pack nutrition label had any effect on the FSA score of the chosen recipe. Previous work suggests that nutrition knowledge has a strong effect on label use [15]. We assume that people with high cooking experience have more knowledge about nutrition. Previous studied also suggests that higher education tend to have higher nutritional knowledge [9]. Multiple two-way ANOVAs was carried out examining the effect of personal characteristics and labels on FSA score.

A two way ANOVA was carried out on FSA score by cooking experience and label, the results are described in Table 4.5. Cooking experience was divided into three groups (Low, Medium, High). There was no statistically significant interaction between the effects of cooking experience and label on the FSA score of the chosen recipes ($p = 0.218$).

Table 4.5: Result of 2x2 ANOVA on the FSA score of the chosen recipes, examining cooking experience and label use.

	Df	Sum Sq	Mean Sq	F value	p
Label	2	30.23	15.12	8.13	0.001
Cooking experience	2	0.05	0.03	0.01	0.986
Label:Cooking experience	4	10.71	2.68	1.44	0.218
Residuals	2052	3813.89	1.86		

For goals, we put participants into 1 of 3 groups, more (participants who want to eat more protein, fruit, vegetables or gain weight), no goal, and reduce (participants who want to lose weight or eat less salt). There was no statistically significant difference for the interaction Goal*Label ($p = 0.403$).

For education, we created three groups for education level and ran a two-way ANOVA. There were no statistically significant differences for the interaction Education*Label ($p = 0.582$). We examined if self-reported eating habits had an effect. There was no statistically significant difference for the interaction Eating habits*Label ($p = 0.320$). We also compared Americans and Europeans. We expected Europeans to be more familiar with the labels and therefore use them to make healthier decisions. We ran a one-way ANOVA comparing the mean FSA score of Americans and Europeans, there was no statistically significant difference between the two groups ($p = 0.125$).

Can we predict the FSA scores of the chosen recipes?

We also explored to what extent other features and user characteristics predicted the chosen FSA score. To this end, we created a linear regression model to examine if personal characteristics or list features affected the FSA score of the chosen recipes across all conditions. The results are depicted in 4.6. Results suggests that the average rating ($p = 0.001$) and position ($p = 0.001$) have an effect on the FSA score for the chosen recipe. However, with an R^2 of 0.129 the model does not explain much of the variance in the data.

Table 4.6: Linear regression model predicting the FSA score of the chosen recipes.

	Term	Estimate	Std Error	Statistic	p
1	(Intercept)	-1	0.464	-1.442	0.150
2	Average Rating	1	0.097	15.144	0.001
3	Position	0	0.005	11.588	0.001
4	Gender	0	0.056	1.184	0.237
5	Age	-0	0.002	-0.013	0.990
6	Education	0	0.034	1.532	0.126
7	Experience	-0	0.027	-0.699	0.485
8	Eating Habits	0	0.031	1.849	0.065

4.2 RQ2: Can the presence of a front-of-pack nutrition label in combination with a health ranking decrease choice difficulty?

To examine RQ2 a two-way ANOVA was conducted to examine the effect of front-of-pack nutrition labels and ranking on a participant's perceived choice difficulty level across three different trials. The results of the ANOVA run on 694 participants are reported in Table 4.9. To determine if we could use the mean of the answers to the three different choice difficulty question's as a single aspect we conducted a factor analysis, for which the items and factor loading are reported in Table 4.7. The KMO values of these three questions ($KMO = 0.7$) are considered acceptable [32]. Based on the results in Table 4.7 we omitted question 5 from the analysis, because it had a low factor loading. Doing so Cronbach's alpha increased from -0.60 to 0.62. Participant's responses for the two remaining questions were averaged and used in the analysis.

Table 4.7: Factor loading's from the factor analysis conducted on the three choice difficulty questions.

Questions	Loading
4. I changed my mind several times before making a decision	0.67
~~5. It was easy to make this choice~~	
6. Making a choice was overwhelming	0.67

The results of the 2x2 ANOVA, indicates that there is no statistically significant interaction between Rank and Label on the participants choice difficulty ($F(4,2073) = 1.04$, $p = 0.385$). The main effect of Rank is not statistically significant ($F(2,2073) = 0.79$, $p = 0.456$). This suggested that a participant's choice difficulty did not decrease due to changing how recipes were ranked.

Table 4.8: Result of 2x2 ANOVA on the perceived choice difficulty of the chosen recipes (three per user).

	Df	Sum Sq	Mean Sq	F value	p
Rank	2	1.45	0.73	0.79	0.456
Label	2	5.89	2.94	3.18	0.041
Rank:Label	4	3.85	0.96	1.04	0.385
Residuals	2073	1919.75	0.93		

The main effect of Label is statistically significant and very small ($F(2,2073) = 3.18$, $p = 0.042$). A post hoc Tukey test depicted in Table 4.9 showed that the Nutri-Score and MTL label differed significantly ($p = .043$). Looking at the mean choice difficulty for the two labels depicted in Table 4.10, we see that MTL has decreased choice difficulty the most (-0.587), and that the choice difficulty increases when a recipe is presented with a Nutri-Score label (-0.463). This might be an indication that the Nutri-Score label causes user to assess the healthiness of the recipes more thoroughly, making it harder to make a choice (a user might have to consider an attractive image vs. a negative Nutri-Score of E).

Table 4.9: Result of Post-hoc Tukey comparing choice difficulty between different conditions.

Contrast	Null	Estimate	Conf. Low	Conf. High	p
MTL-No-label	0.001	-0.03	-0.15	0.09	0.85
Nutriscore-No-label	0.001	0.10	-0.03	0.22	0.15
Nutriscore-MTL	0.001	0.12	0.001	0.25	0.04

Figure 4.2 depicts the interaction plot for the 2x2 ANOVA in Table 4.9. Looking at the Random ranking variant, we can see that the Nutri-Score label in combination with a Random ranking increased choice difficulty the most. This can probably be explained

Table 4.10: Mean choice difficulty for all the different conditions and variant. The higher the value, the more choice difficulty increased. Choice difficulty increased the most when the Nutri-Score label was combined with a random re-ranking of search results.

Label	Rank	N	Mean	SD
No-label	Random	230.00	-0.57	0.96
No-label	Popular	230.00	-0.60	0.83
No-label	Health	230.00	-0.51	0.97
MTL	Random	233.00	-0.60	1.01
MTL	Popular	233.00	-0.59	0.86
MTL	Health	233.00	-0.57	0.98
Nutriscore	Random	231.00	-0.35	1.07
Nutriscore	Popular	231.00	-0.54	0.95
Nutriscore	Health	231.00	-0.50	1.01

by the fact that the Nutri-Score label has an inherent ranking (A-E) that does not work well when the search results are ordered randomly. We can also see that combining FOP labels with a health ranking does not increase choice difficulty when compared to a no-label baseline, indicating that FOP labels in combination with a health ranking have no detrimental effect on choice difficulty.

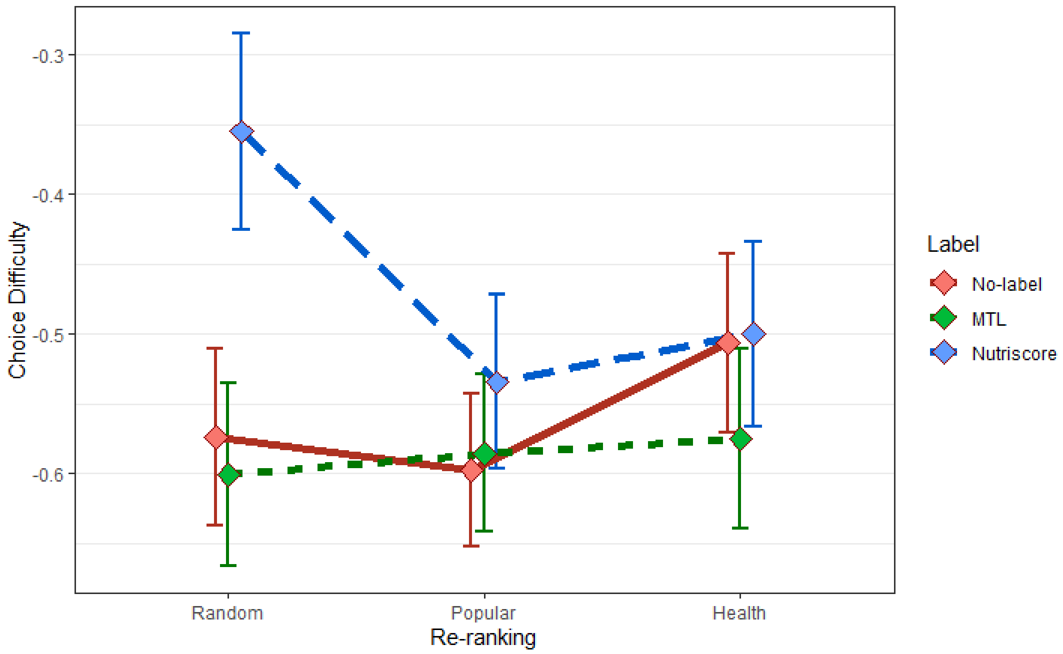


Figure 4.2: Marginal effects plot with error bars (depicting 1 S.E.), between re-ranking of search results and different labels regarding the perceived choice difficulty. Displaying a Nutri-Score label on the recipe card in combination with a random re-ranking of the search result increased perceived choice difficulty the most.

We also ran the analysis accounting for several personal factors, such as excluding vegans and participants with allergies, comparing participants with different goals, education, nationality, and cooking experience, but we found no significant statistical difference in choice difficulty.

4.3 RQ3: To what extent can front-of-pack nutrition labels influence choice satisfaction?

A two-way ANOVA was also performed to examine the effects of front-of-pack nutrition labels and ranking on a participant's perceived choice satisfaction level across three different trials. The results of the ANOVA run on 694 participants are reported in Table 4.13. To determine if we could use the mean of the answers to the three different choice satisfaction question's as a single aspect we used a confirmatory factor analysis, for which the items and factor loading are reported in Table 4.11. The KMO values of these three questions $(KMO = 0.7)$ are considered acceptable [32]. The α coefficient for the three items is .805, a questionnaire with an α of 0.8 is considered reliable [32]. Participant's responses for the three remaining questions were averaged and used in the analysis.

Table 4.11: Factor loading's of the factor analysis. We took the mean of the three choice satisfaction questions to measure choice satisfaction.

Questions	Loading
1. I would recommend the chosen recipe to others	-0.730
2. My chosen recipe could become one of my favourites	-0.859
3. I think I would enjoy eating the chosen recipe	-0.709

The results of the 2x2 ANOVA, indicates that there is no statistically significant interaction between Rank and Label on the participants choice satisfaction $(F(4,2073) = 0.84$, $p = 0.499)$. The main effect of Rank is statistically not significant $(F(2,2073) = 1.20$, $p = 0.303)$. The main effect of Label is statistically not significant $(F(2,2073) = 2.18$, $p = 0.113)$. This suggested that a participants choice satisfaction did not decrease or increase in any significant way due to changing how recipes were ranked and with different labels. Figure 4.3 depicts the interaction plot (Rank*Label) for the 2x2 ANOVA model in Table 4.13.

Table 4.12: Result of 2x2 ANOVA on the perceived choice satisfaction of the chosen recipes (three per user).

	Df	Sum Sq	Mean Sq	F value	*p*
Rank	2	0.73	0.36	1.20	0.302
Label	2	1.33	0.66	2.18	0.113
Rank:Label	4	1.02	0.26	0.84	0.499
Residuals	2073	631.22	0.30		

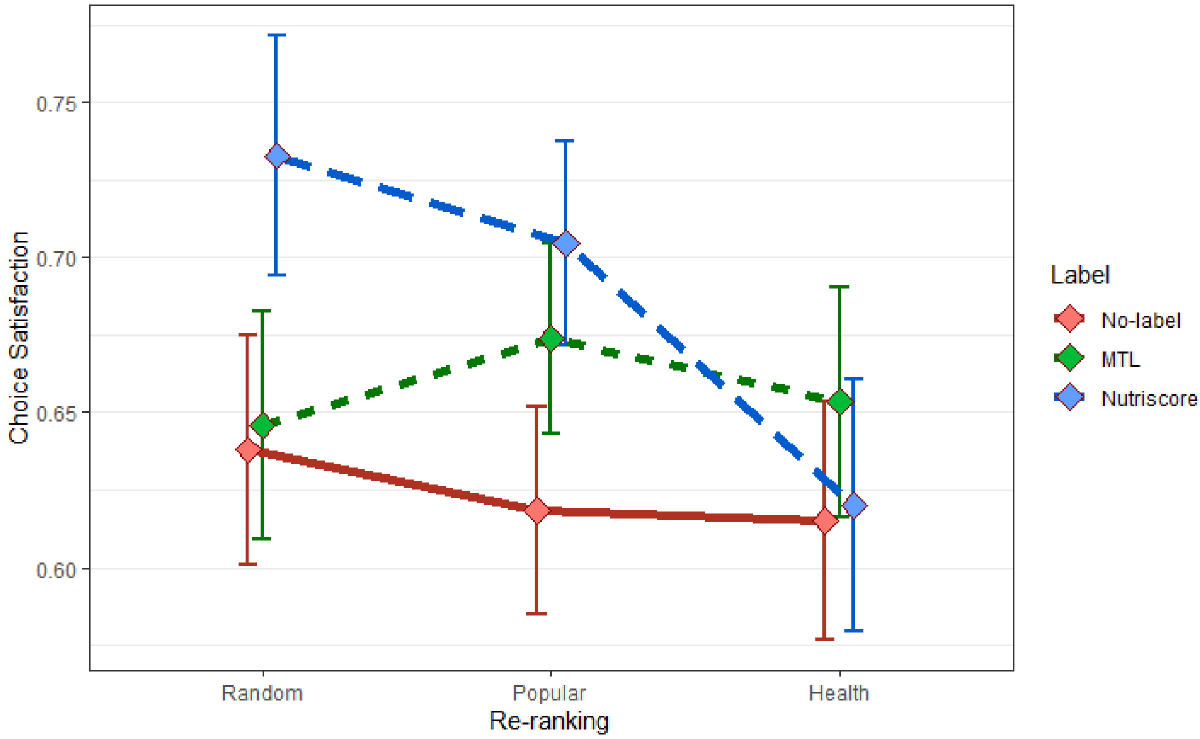


Figure 4.3: Marginal effects plot with error bars (depicting 1 S.E.), between re-ranking of search results and different labels regarding the perceived choice satisfaction. Higher values indicates higher choice satisfaction.

We also conducted a post hoc Tukey test on Label, to further examine the effect the different labels had on choice satisfaction. The results show no statistically significant difference between the labels. However, it is interesting to note that choice satisfaction is highest when Nutri-Score is combined with a random ranking.

Table 4.13: Result of post-hov Tukey HSD test to examine the effect different labels had on choice satisfaction.

Contrast	Null	Estimate	Conf. Low	Conf. High	*p*
MTL-No-label	0.001	0.03	-0.04	0.10	0.49
Nutriscore-No-label	0.001	0.06	-0.01	0.13	0.09
Nutriscore-MTL	0.001	0.03	-0.04	0.10	0.61

Chapter 5

Conclusions and Future Work

This chapter concludes the thesis by summarizing the findings from the study, possible limitations of the approach, and provides possible directions going forward.

5.1 Discussion

Choosing recipes to prepare is a complex process. Many people have difficulties understanding the nutritional quality of food [12]. In addition, it is hard to judge the healthiness of internet recipes [63]. Popular internet recipes are often high in specific nutrients, such as fat and sodium [65]. Recipes are often ranked by popularity in food search interfaces, and the unhealthiness of these recipes promotes unhealthy eating habits. To alleviate this, we have examined whether front-of-pack nutrition labels and ranking recipes based on health scores can support healthy food choices, without increasing choice difficulty and decreasing choice satisfaction. This thesis is the first to examine the use of front-of-pack nutrition labels in information retrieval interfaces for internet recipes.

Our results show that through the use of front-of-pack nutrition labels search results could be re-designed to support healthier food choices. By explicitly informing users how unhealthy some recipes are might change their eating habits by nudging them towards healthier recipes. The findings of this master thesis can be summarized as follows:

RQ1: To what extent can front-of-pack nutrition labels and ranking be used to nudge users towards healthier recipes online?

To answer this research question a two-way ANOVA with post-hoc Tukey HSD tests were conducted. The result of the analysis indicated that there was a difference between both the labels (MTL and Nutri-Score) and the baseline (no-label). This suggests that front-of-pack nutrition labels can nudge users towards healthier recipes and support healthy food choices online. Moreover, the analysis revealed no statistically significant difference between the FSA and WHO score for the two labels, suggesting that both labels were equally effective in nudging users towards healthier recipes. These findings are in line with those from a recent study showing an improvement in the nutritional quality of food items selected in an online grocery store when items were presented with a Nutri-Score and Multiple Traffic Lights label and compared against a no label control condition [23].

By using a simple re-rank on health (i.e., a recipe's FSA score or Nutri-Score) we found no improvement in the healthiness of the chosen recipes. This suggests that a simple health ranking, where recipes are ranked on health scores from healthy to unhealthy is not enough to nudge people towards healthier recipes. This finding was not consistent with Starke et. al. [58], which indicated that a health-based ranking led to healthier choices compared to a popularity-based ranking. The results might be explained by the presence of the FOP labels in the search result. The labels which might have caused a "healthy food = less tasty" effect which describes how people associate healthy food with being less tasty than unhealthy food [49]. This might have caused participants to skip the first recipes in the health-based ranking scenario. There was a significant difference in FSA score between MTL and No-label in the random ranking scenario, suggesting that the MTL label is more effective at nudging users towards healthier recipes when recipes are ranked randomly.

The impact of the two different front-of-pack nutrition labels was similar across demographic subgroups, and for people with different food knowledge and health goals. These findings are in line with previous studies that examined different front-of-pack nutrition labels [22, 23]. The lack of differential effects by personal factors is encouraging as it suggests all users of recipe websites may equally benefit from front-of-pack nutrition labels being applied to the recipe cards.

RQ2: Can the presence of a front-of-pack nutrition label in combination with a health ranking decrease choice difficulty?

To answer this research question a two-way ANOVA with post-hoc Tukey HSD tests were conducted. The results suggest that the use of front-of-pack nutrition labels did not increase choice difficulty when compared with a no-label baseline. However, there was a statistically significant difference in choice difficulty between the two labels. The choice difficulty was highest in the Nutri-Score condition and lowest in the Multiple Traffic Lights condition. Previous work suggests that the Nutri-Score label has the greatest effect on perceived healthiness [31]. This might indicate that users become more aware of recipe healthiness when presented with a Nutri-Score label, making it more difficult to choose a recipe because healthiness is considered more closely, in addition to other recipe attributes (i.e., image, title, ingredients).

We also found that combining a Nutri-Score label with a random ranking of the recipes increased choice difficulty the most. This can probably be explained by the fact that the Nutri-Score label has an inherent ranking (A-E) that does not work well when the search results are ordered randomly. Looking at labels in combination with a simple health ranking neither increased nor decreased perceived choice difficulty. This suggests that combining a health ranking with front-of-pack nutrition labels does not have any detrimental effects on perceived choice difficulty. Furthermore, combining front-of-pack nutrition labels with a popular re-ranking of search results did not negatively impact choice difficulty either, which suggests that front-of-pack nutrition labels can be applied to the recipe card without having a negative impact on the user experience.

RQ3: To what extent can front-of-pack nutrition labels influence choice satisfaction?

To answer this research question a two-way ANOVA with post-hoc Tukey HSD tests were conducted. The results suggest that there is no statistically significant decrease in participant's perceived choice satisfaction either by labels or ranking. It is interesting to note that both choice satisfaction and choice difficulty are highest when Nutri-Score is combined with a random ranking. However, the effect is not statistically significant. This might indicate that the Nutri-Score label has the greatest effect of the two labels, in that it makes it harder to make a choice (perhaps because healthiness is more closely considered) and that the users are more satisfied with the recipes they have chosen.

This study is the first to measure how front-of-pack nutrition labels affect choice difficulty and choice satisfaction, the results show that the Multiple Traffic Light label can be applied to online recipes without negatively impacting these user evaluation aspects. Front-of-pack nutrition labels neither increase choice difficulty nor decrease choice satisfaction when presented on the recipes card in a search interface. This has positive implications for implementing front-of-pack nutrition labels in recipe search interfaces, as the labels do not negatively impact user evaluation in terms of choice difficulty or choice satisfaction. These findings are consistent with Pecune and Marsella [47], which indicates that the presence of a healthy tag on the recipe card increases choice satisfaction.

5.2 Limitations and Future Research

There are a few shortcomings to the work in this thesis. We choose a relatively small set of recipes to gain more control over the different recipe features but did not account for the attractiveness of the images. The attractiveness of recipe images has been shown to be an important factor when nudging users towards healthier recipe choices [21, 58], as people are attracted to tasty-looking images. In our study, we asked participants which factors influenced their recipe choices the most, and the option "Attractive images" was the most important factor across all three conditions, the top eight answers to this question are depicted in Figures 5.8, 5.9 and 5.10. By also accounting for image attractiveness when choosing which recipes to include in our search results we believe the result of the study could have been more robust.

By only including a simple health ranking we were not able to take full advantage of the positioning effects. We know from the literature that items placed both at the top and bottom of a list tend to be chosen more often[14]. Instead of a simple health ranking where the healthiest recipes are placed at the top of the page, a more advanced ranking solution could have been developed, where the healthiest recipes are placed both at the top and bottom of the recipe list. This would allow us to take full advantage of the positioning effect, and examine if this would lead to even more healthy recipes being chosen, as compared to only placing healthy recipes at the top of the list.

Such an approach could work well with the Multiple Traffic Light label, as there is no inherent ranking implied in the label, which is the case for the Nutri-Score label where healthiness is expressed from A (healthy) to E (unhealthy). Overall, the healthiest recipes were selected when we combined Multiple Traffic Light labels with a random ranking. This indicates that combining the Multiple Traffic Light label with a more advanced health ranking could be advantageous, and should in theory increase the chance of healthy recipes being chosen more often.

By using a more advanced health ranking we would also increase the diversity of the set of recipe items in the search results. The result from Willemsen, Graus and Knijnenburg [74] suggests that selecting from a small, diversified set of items reduces choice difficulty and objective effort while at the same time maintaining similar or higher levels of satisfaction than choosing from larger, non-diverse sets. We believe that investigating more advanced health ranking approaches could be an interesting avenue for future research.

Using a vertical list where users only saw 2/3 recipes at a time might have lessened the positioning effect we were aiming for. Positioning effects are more pronounced when all items are assessed at the same time, such as in a restaurant menu [14]. Therefore, it might be interesting to examine whether placing healthy recipes both at the beginning and end of a horizontal list or in a multi-list interface might be more effective.

The study also highlights another challenge with regards to which measure should be used to determine the healthiness of recipes. In the current study, we examine the WHO, FSA and Nutri-Score to assess the healthiness of recipes. As can be seen in table 4.2, these measures vary widely, one recipe is judged as being healthy by the Nutri-Score and unhealthy by the FSA score. To be able to actually recommend healthier recipes we need a robust measure to accurately describe what a healthy recipe actually is.

An interesting avenue for future research could be to develop more advanced nutrition labels designed specifically for online recipe search interfaces. Currently, all front-of-pack nutrition labels are designed to be used on pre-packaged food products. The primary tool used in both the Nutri-Score label and the Multiple Traffic Light label to grab people's attention and signal the healthiness of a product is the colours on the labels. In a web-based user interface, more tools are available to grab users attention, such as animations. Another avenue for future research would be to integrate front-of-pack nutrition labels with personalized approaches to food search such as food recommender systems.

5.3 Open Science

To make this study reproducible, the code for the prototype, the data from the user study, and the code used to analyze it is shared with the scientific community. All resources used in this study are available in a GitHub repository [1]. The repository includes the source code for the prototype, with one folder for each of the three conditions (Nutri-Score, Multiple Traffic Lights and No-label). The folder "Data Analysis" contains a folder with the cleaned data from the user study "Cleaned data" and a folder with the unprocessed results "Raw data". The "Data Analysis" folder also contains one folder for each research question (i.e., "RQ1_Results") with the R and Python code used to analyze the data.

[1] https://github.com/JorgenNyborgChristensen/FrontOfPackFoodSearchMasterThesis

Bibliography

[1] F. S. Agency. Front-of-pack traffic light signpost labelling: Technical guidance, 2007. 2.3.2

[2] Amazon. Qualifications and Worker Task Quality . https://blog.mturk.com/qualifications-and-worker-task-quality-best-practices-886f1f4e03fc, 2019. Online last accessed 23 November 2020]. 3.4.6

[3] A. V. Banerjee. A simple model of herd behavior. *The quarterly journal of economics*, 107(3):797–817, 1992. 2.2.3

[4] J. Barreiro-Hurlé, A. Gracia, and T. De-Magistris. Does nutrition information on food products lead to healthier food choices? *Food Policy*, 35(3):221–229, 2010. 2.4

[5] A. Black and M. Rayner. *Just read the label: understanding nutrition information in numeric, verbal and graphic formats*. HM Stationery Office London, 1992. 2.4

[6] R. Cadario and P. Chandon. Which healthy eating nudges work best? a meta-analysis of field experiments. *Marketing Science*, 39(3):465–486, 2020. 2.2

[7] H. Cai, Y. Chen, and H. Fang. Observational learning: Evidence from a randomized natural field experiment. *American Economic Review*, 99(3):864–82, 2009. 2.2.3

[8] S. Campos, J. Doxey, and D. Hammond. Nutrition labels on pre-packaged foods: a systematic review. *Public health nutrition*, 14(8):1496–1506, 2011. 2.3

[9] K. Cannoosamy, P. Pugo-Gunsam, and R. Jeewon. Consumer knowledge and attitudes toward nutritional labels. *Journal of nutrition education and behavior*, 46(5):334–340, 2014. 4.1

[10] J. Chantal, S. Hercberg, W. H. Organization, et al. Development of a new front-of-pack nutrition label in france: the five-colour nutri-score. *Public Health Panorama*, 3(04):712–725, 2017. 2.3, 2.3.1, 3.1

[11] W.-Y. Chao and Z. Hass. Choice-based user interface design of a smart healthy food recommender system for nudging eating behavior of older adult patients with newly diagnosed type ii diabetes. In *International Conference on Human-Computer Interaction*, pages 221–234. Springer, 2020. 2.2

[12] G. Cowburn and L. Stockley. Consumer understanding and use of nutrition labelling: a systematic review. *Public health nutrition*, 8(1):21–28, 2005. 2.4, 5.1

[13] S. L. Crites and S. N. Aikman. Impact of nutrition knowledge on food evaluations. *European journal of clinical nutrition*, 59(10):1191–1200, 2005. 2.1, 2.4

[14] E. Dayan, M. Bar-Hillel, et al. Nudge to nobesity ii: Menu positions influence food orders. *Judgment and Decision making*, 6(4):333–342, 2011. 1.1, 2.2, 2.2.2, 5.2

[15] A. C. Drichoutis, P. Lazaridis, and R. M. Nayga. Nutrition knowledge and consumer use of nutritional food labels. *European Review of Agricultural Economics*, 32(1):93–118, 2005. 4.1

[16] A. C. Drichoutis, P. Lazaridis, R. M. Nayga, M. Kapsokefalou, and G. Chryssochoidis. A theoretical and empirical investigation of nutritional label use. *The European Journal of Health Economics*, 9(3):293–304, 2008. 2.3

[17] P. Ducrot, C. Julia, C. Méjean, E. Kesse-Guyot, M. Touvier, L. K. Fezeu, S. Hercberg, and S. Péneau. Impact of different front-of-pack nutrition labels on consumer purchasing intentions: a randomized controlled trial. *American journal of preventive medicine*, 50(5):627–636, 2016. 1.1, 2.3

[18] P. Ducrot, C. Méjean, C. Julia, E. Kesse-Guyot, M. Touvier, L. Fezeu, S. Hercberg, and S. Péneau. Effectiveness of front-of-pack nutrition labels in french adults: results from the nutrinet-sante cohort study. *PLoS One*, 10(10):e0140898, 2015. 2.3.1

[19] ebizma. Top 15 Best Recipe Websites | March 2021 . http://www.ebizmba.com/articles/recipe-websites/, 2020. Online last accessed 31 May 2021. 3.1

[20] M. Egnell, P. Ducrot, M. Touvier, B. Allès, S. Hercberg, E. Kesse-Guyot, and C. Julia. Objective understanding of nutri-score front-of-package nutrition label according to individual characteristics of subjects: Comparisons with other format labels. *PLoS One*, 13(8):e0202095, 2018. 2.3

[21] D. Elsweiler, C. Trattner, and M. Harvey. Exploiting food choice biases for healthier recipe recommendation. In *Proceedings of the 40th international acm sigir*

conference on research and development in information retrieval, pages 575–584, 2017. 1.1, 2.1.1, 2.1.2, 3.1, 5.2

[22] G. I. Feunekes, I. A. Gortemaker, A. A. Willems, R. Lion, and M. Van Den Kommer. Front-of-pack nutrition labelling: testing effectiveness of different nutrition labelling formats front-of-pack in four european countries. *Appetite*, 50(1):57–70, 2008. 2.3.1, 5.1

[23] E. A. Finkelstein, F. J. L. Ang, B. Doble, W. H. M. Wong, and R. M. van Dam. A randomized controlled trial evaluating the relative effectiveness of the multiple traffic light and nutri-score front of package nutrition labels. *Nutrients*, 11(9):2236, 2019. 1.1, 2.3, 2.3.1, 2.3.2, 5.1

[24] J. Freyne and S. Berkovsky. Intelligent food planning: personalized recipe recommendation. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 321–324, 2010. 2.1.1

[25] K. M. Gallagher and J. A. Updegraff. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. *Annals of behavioral medicine*, 43(1):101–116, 2012. 2.2.1

[26] N. Garg, R. Govind, and A. Nagpal. Message framing effects on food consumption: A social marketing perspective. *Australian Journal of Management*, page 0312896221989398, 2021. 2.2.1

[27] M. Ge, M. Elahi, I. Fernaández-Tobías, F. Ricci, and D. Massimo. Using tags and latent factors in a food recommender system. In *Proceedings of the 5th International Conference on Digital Health 2015*, pages 105–112, 2015. 2.1.1

[28] M. Ge, F. Ricci, and D. Massimo. Health-aware food recommender system. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 333–334, 2015. 2.1.1

[29] D. J. Graham and M. N. Laska. Nutrition label use partially mediates the relationship between attitude toward healthy eating and overall dietary quality among college students. *Journal of the Academy of Nutrition and Dietetics*, 112(3):414–418, 2012. 2.4

[30] K. G. Grunert, J. M. Wills, and L. Fernández-Celemín. Nutrition knowledge, and use and understanding of nutrition information on food labels among consumers in the uk. *Appetite*, 55(2):177–189, 2010. 2.3

[31] D. Hagmann and M. Siegrist. Nutri-score, multiple traffic light and incomplete nutrition labelling on food packages: Effects on consumers' accuracy in identifying healthier snack options. *Food Quality and Preference*, 83:103894, 2020. 5.1

[32] M. Hof. Questionnaire evaluation with factor analysis and cronbach ' s alpha an example -. 2012. 4.2, 4.3

[33] S. Howard, J. Adams, and M. White. Nutritional content of supermarket ready meals and recipes by television chefs in the united kingdom: cross sectional study. *Bmj*, 345, 2012. 3.1, 3.4.7

[34] Q. Jin, A. Animesh, and A. Pinsonneault. When popularity meets position. 2019. 2.2.3

[35] E. J. Johnson, S. B. Shu, B. G. Dellaert, C. Fox, D. G. Goldstein, G. Häubl, R. P. Larrick, J. W. Payne, E. Peters, D. Schkade, et al. Beyond nudges: Tools of a choice architecture. *Marketing Letters*, 23(2):487–504, 2012. 3.2.2

[36] C. Julia, S. Péneau, C. Buscail, R. Gonzalez, M. Touvier, S. Hercberg, and E. Kesse-Guyot. Perception of different formats of front-of-pack nutrition labels according to sociodemographic, lifestyle and dietary factors in a french population: cross-sectional study among the nutrinet-sante cohort participants. *BMJ open*, 7(6):e016108, 2017. 2.3.1

[37] M. W. Kreuter, L. K. Brennan, D. P. Scharff, and S. N. Lukwago. Do nutrition label readers eat healthier diets? behavioral correlates of adults' use of food labels. *American journal of preventive medicine*, 13(4):277–283, 1997. 2.4

[38] F. M. Kroese, D. R. Marchiori, and D. T. de Ridder. Nudging healthy food choices: a field experiment at the train station. *Journal of Public Health*, 38(2):e133–e137, 2016. 2.2

[39] M. Li and G. B. Chapman. Nudge to health: Harnessing decision research to promote health behavior. *Social and Personality Psychology Compass*, 7(3):187–198, 2013. 2.2

[40] G. L. Lohse. Consumer eye movement patterns on yellow pages advertising. *Journal of Advertising*, 26(1):61–73, 1997. 2.3

[41] L. Machín, J. Aschemann-Witzel, M. R. Curutchet, A. Giménez, and G. Ares. Does front-of-pack nutrition information improve consumer ability to make

healthful choices? performance of warnings and the traffic light system in a simulated shopping experiment. *Appetite*, 121:55–62, 2018. 2.3

[42] T. M. Marteau, D. Ogilvie, M. Roland, M. Suhrcke, and M. P. Kelly. Judging nudging: can nudging improve population health? *Bmj*, 342, 2011. 2.2

[43] M. Milosavljevic, V. Navalpakkam, C. Koch, and A. Rangel. Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology*, 22(1):67–74, 2012. 2.1.1

[44] B. B. Murdock Jr. The serial position effect of free recall. *Journal of experimental psychology*, 64(5):482, 1962. 2.2.2

[45] C. Musto, C. Trattner, A. Starke, and G. Semeraro. Towards a knowledge-aware food recommender system exploiting holistic user models. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, pages 333–337, 2020. 2.1.1

[46] W. H. Organization et al. Global strategy on diet, physical activity and health: a framework to monitor and evaluate implementation. 2006. 2.3

[47] F. Pecune, L. Callebert, and S. Marsella. A recommender system for healthy and personalized recipe recommendations. 2.1.2, 5.1

[48] P. Pu, L. Chen, and R. Hu. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 157–164, 2011. 3.4.5, 3.4.7

[49] R. Raghunathan, R. W. Naylor, and W. D. Hoyer. The unhealthy= tasty intuition and its effects on taste inferences, enjoyment, and choice of food products. *Journal of Marketing*, 70(4):170–184, 2006. 5.1

[50] D. Rawson, I. Janes, and K. Jordan. Pilot study to investigate the potential of eye tracking as a technique for fsa food labelling behaviour research. *Report for FSA*, 2008. 2.3

[51] F. Ricci, L. Rokach, and B. Shapira. Introduction to recommender systems handbook. pages 1–35, 2011. 2.1.1

[52] A. Robertson, C. Tirado, T. Lobstein, C. Knai, J. Jensen, A. Ferro-Luzzi, and W. James. Food and health in europe: A new basis for action (european series no 96), 2004. 1.1

[53] M. Rokicki, C. Trattner, and E. Herder. The impact of recipe features, social cues and demographics on estimating the healthiness of online recipes. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12, 2018. 2.1

[54] A. J. Rothman and P. Salovey. Shaping perceptions to motivate healthy behavior: the role of message framing. *Psychological bulletin*, 121(1):3, 1997. 2.2.1

[55] S. S. Schiffman, B. G. Graham, E. A. Sattely-Miller, and Z. S. Warwick. Orosensory perception of dietary fat. *Current Directions in Psychological Science*, 7(5):137–143, 1998. 2.4

[56] C. Schneider, M. Weinmann, and J. Vom Brocke. Digital nudging: guiding online user choices through interface design. *Communications of the ACM*, 61(7):67–73, 2018. 2.2

[57] L. R. Skov, S. Lourenco, G. L. Hansen, B. E. Mikkelsen, and C. Schofield. Choice architecture as a means to change eating behaviour in self-service settings: a systematic review. *Obesity Reviews*, 14(3):187–196, 2013. 2.2

[58] A. D. Starke, M. C. Willemsen, and C. Trattner. Nudging healthy choices in food search through visual attractiveness. *Frontiers in Artificial Intelligence*, 4, 2021. 3.1, 3.1, 5.1, 5.2

[59] N. J. Temple. Front-of-package food labels: A narrative review. *Appetite*, 144:104485, 2020. 2.3, 2.3.2

[60] R. H. Thaler and C. R. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, 2009. 2.2

[61] T. N. T. Tran, M. Atas, A. Felfernig, and M. Stettinger. An overview of recommender systems in the healthy food domain. *Journal of Intelligent Information Systems*, 50(3):501–526, 2018. 2.1.2

[62] C. Trattner and D. Elsweiler. Food recommender systems: important contributions, challenges and future research directions. *arXiv preprint arXiv:1711.02760*, 2017. 2.1.1, 2.1.2, 2.3.1, 3.1

[63] C. Trattner and D. Elsweiler. Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proceedings of the 26th international conference on world wide web*, pages 489–498, 2017. 3.4.7, 5.1

[64] C. Trattner, D. Elsweiler, and S. Howard. Estimating the healthiness of internet recipes: a cross-sectional study. *Frontiers in public health*, 5:16, 2017. 1.1, 2.1.1, 3.1

[65] C. Trattner, D. Moesslang, and D. Elsweiler. On the predictability of the popularity of online recipes. *EPJ Data Science*, 7:1–39, 2018. 2.2.3, 5.1

[66] E. Van Herpen and H. C. Van Trijp. Front-of-pack nutrition labels. their effect on attention and choices when consumers have varying goals and time constraints. *Appetite*, 57(1):148–160, 2011. 2.3

[67] Y. Van Pinxteren, G. Geleijnse, and P. Kamsteeg. Deriving a recipe similarity measure for recommending healthful meals. In *Proceedings of the 16th international conference on Intelligent user interfaces*, pages 105–114, 2011. 2.1.2

[68] V. H. Visschers, R. Hess, and M. Siegrist. Health motivation and product design determine consumers' visual attention to nutrition information on food products. *Public health nutrition*, 13(7):1099–1106, 2010. 2.4

[69] B. Wansink and A. S. Hanks. Slim by design: serving healthy foods first in buffet lines improves overall meal selection. *PloS one*, 8(10):e77055, 2013. 1.1, 2.2, 2.2.2

[70] J. Wardle, K. Parmenter, and J. Waller. Nutrition knowledge and food intake. *Appetite*, 34(3):269–275, 2000. 2.4

[71] W. L. Watson, B. Kelly, D. Hector, C. Hughes, L. King, J. Crawford, J. Sergeant, and K. Chapman. Can front-of-pack labelling schemes guide healthier food choices? australian shoppers' responses to seven labelling formats. *Appetite*, 72:90–97, 2014. 2.3

[72] M. Weinmann, C. Schneider, and J. Vom Brocke. Digital nudging. *Business & Information Systems Engineering*, 58(6):433–436, 2016. 2.2

[73] WHO. Obesity and overweight. `https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight`. Accessed: 2020-03-03. 1.1

[74] M. C. Willemsen, M. P. Graus, and B. P. Knijnenburg. Understanding the role of latent feature diversification on choice difficulty and satisfaction. *User Modeling and User-Adapted Interaction*, 26(4):347–389, 2016. 5.2

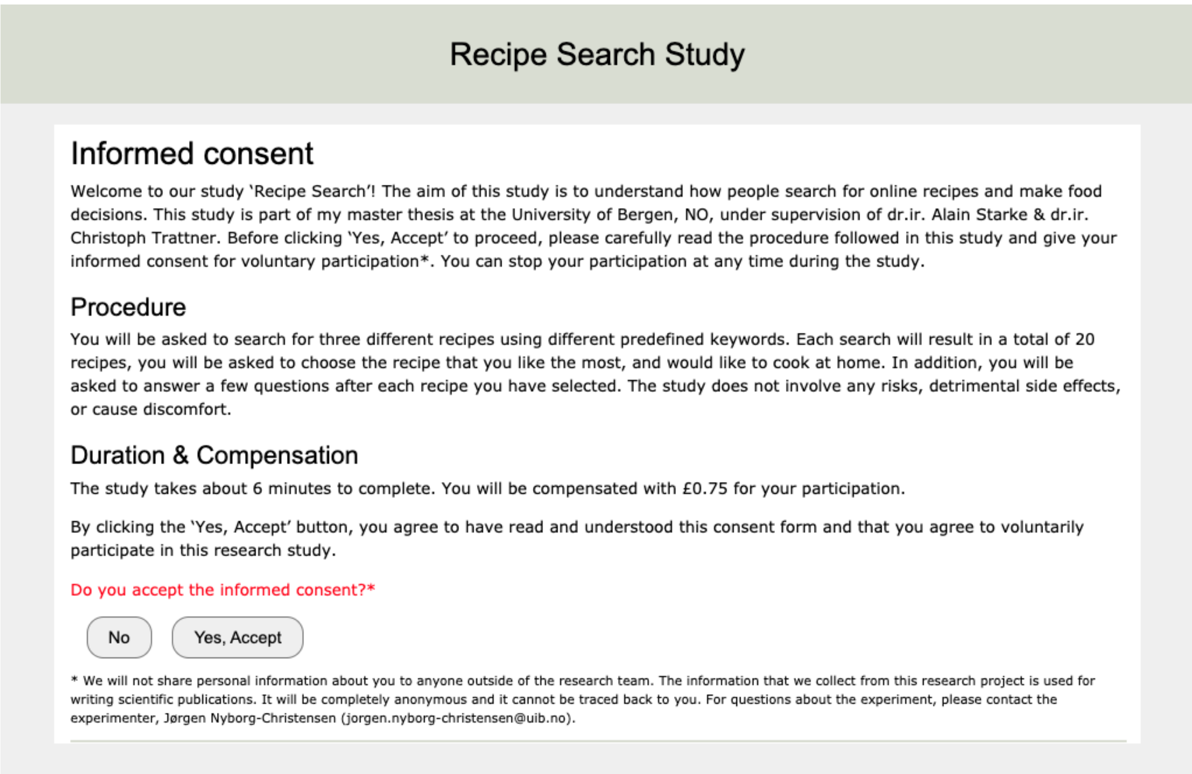
[75] Q. Zhang, D. Elsweiler, and C. Trattner. Visual cultural biases in food classification. *Foods*, 9(6):823, 2020. 2.1.1

Appendix

5.4 Screenshots

This appendix contains screenshots of each step of the Prolific and Mturk user study carried out in this thesis.

Step 1: Instructions



Recipe Search Study

Informed consent

Welcome to our study 'Recipe Search'! The aim of this study is to understand how people search for online recipes and make food decisions. This study is part of my master thesis at the University of Bergen, NO, under supervision of dr.ir. Alain Starke & dr.ir. Christoph Trattner. Before clicking 'Yes, Accept' to proceed, please carefully read the procedure followed in this study and give your informed consent for voluntary participation*. You can stop your participation at any time during the study.

Procedure

You will be asked to search for three different recipes using different predefined keywords. Each search will result in a total of 20 recipes, you will be asked to choose the recipe that you like the most, and would like to cook at home. In addition, you will be asked to answer a few questions after each recipe you have selected. The study does not involve any risks, detrimental side effects, or cause discomfort.

Duration & Compensation

The study takes about 6 minutes to complete. You will be compensated with £0.75 for your participation.

By clicking the 'Yes, Accept' button, you agree to have read and understood this consent form and that you agree to voluntarily participate in this research study.

Do you accept the informed consent?*

No | Yes, Accept

* We will not share personal information about you to anyone outside of the research team. The information that we collect from this research project is used for writing scientific publications. It will be completely anonymous and it cannot be traced back to you. For questions about the experiment, please contact the experimenter, Jørgen Nyborg-Christensen (jorgen.nyborg-christensen@uib.no).

Figure 5.1: The instructions given to the participants recruited on Prolific.

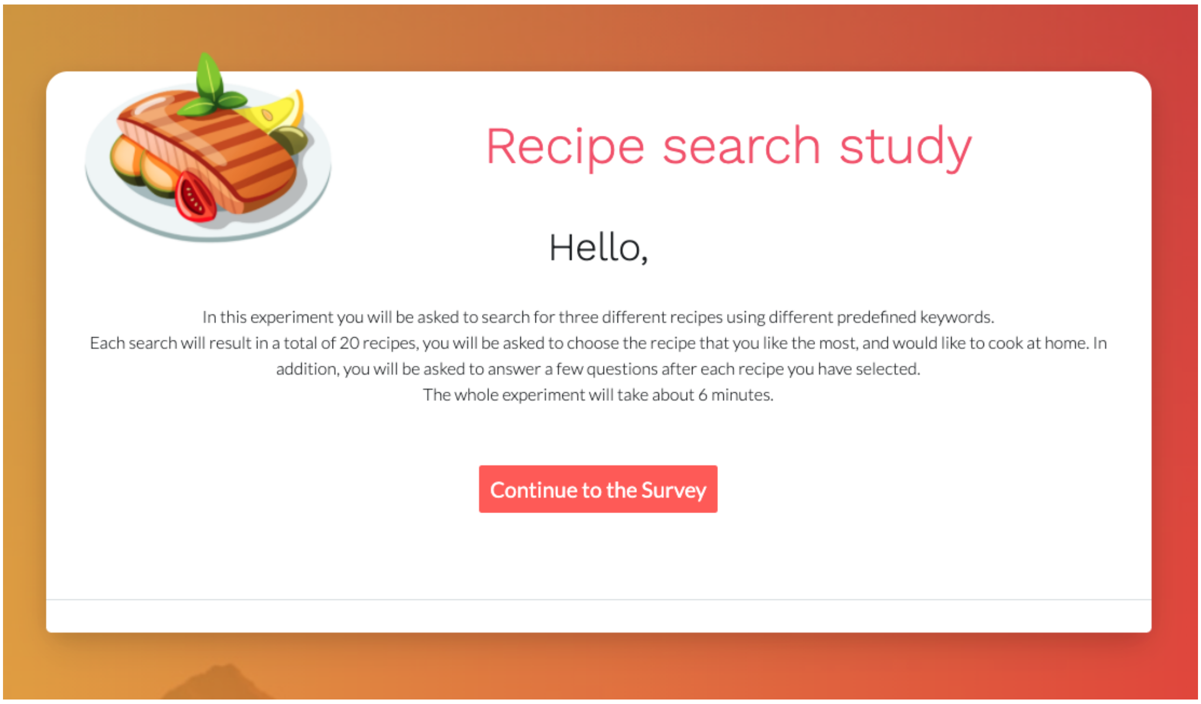
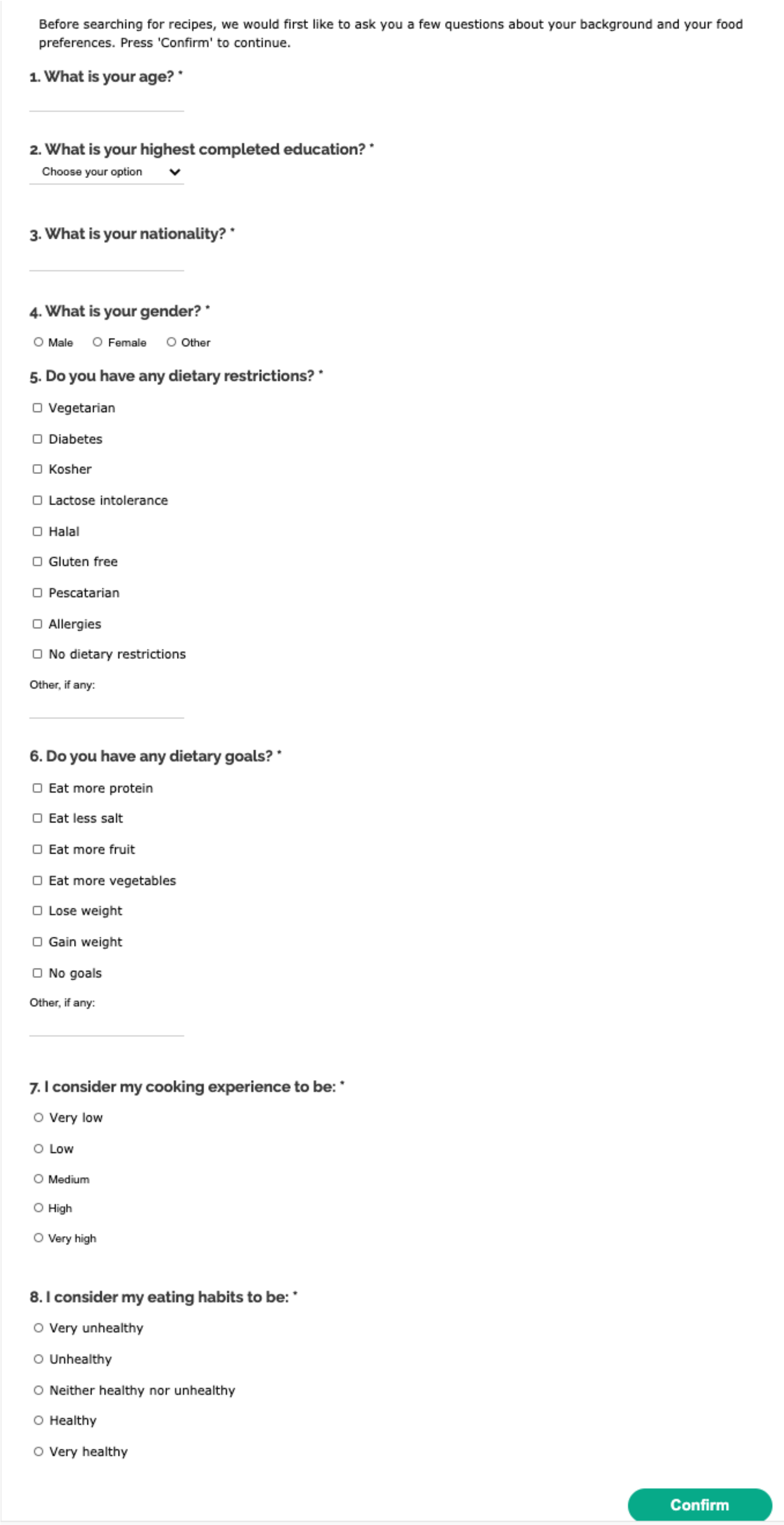


Figure 5.2: The instructions given to the participants recruited on Amazon Mechanical Turk.

Step 2: Demographic Questionnaire


Before searching for recipes, we would first like to ask you a few questions about your background and your food preferences. Press 'Confirm' to continue.

1. What is your age? *

2. What is your highest completed education? *
Choose your option

3. What is your nationality? *

4. What is your gender? *
○ Male ○ Female ○ Other

5. Do you have any dietary restrictions? *
- ☐ Vegetarian
- ☐ Diabetes
- ☐ Kosher
- ☐ Lactose intolerance
- ☐ Halal
- ☐ Gluten free
- ☐ Pescatarian
- ☐ Allergies
- ☐ No dietary restrictions

Other, if any:

6. Do you have any dietary goals? *
- ☐ Eat more protein
- ☐ Eat less salt
- ☐ Eat more fruit
- ☐ Eat more vegetables
- ☐ Lose weight
- ☐ Gain weight
- ☐ No goals

Other, if any:

7. I consider my cooking experience to be: *
- ○ Very low
- ○ Low
- ○ Medium
- ○ High
- ○ Very high

8. I consider my eating habits to be: *
- ○ Very unhealthy
- ○ Unhealthy
- ○ Neither healthy nor unhealthy
- ○ Healthy
- ○ Very healthy

Confirm

Figure 5.3: The demographics questionnaire all participants had to complete before doing the search task.

Step 3: Search Task

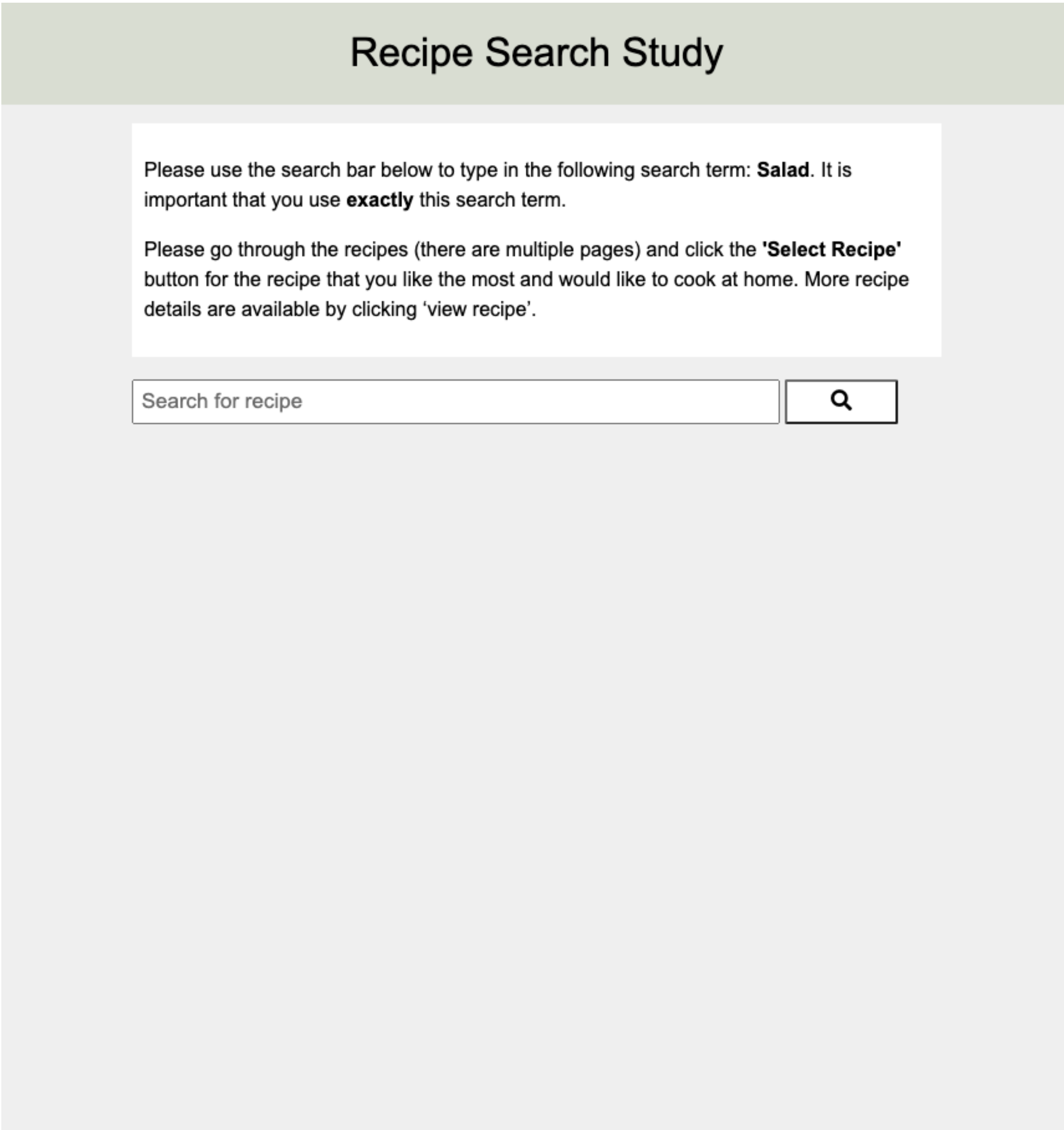


Figure 5.4: Search bar with instructions. Participants were told to use a predefined keyword to search for and select a recipe they would like to cook at home.

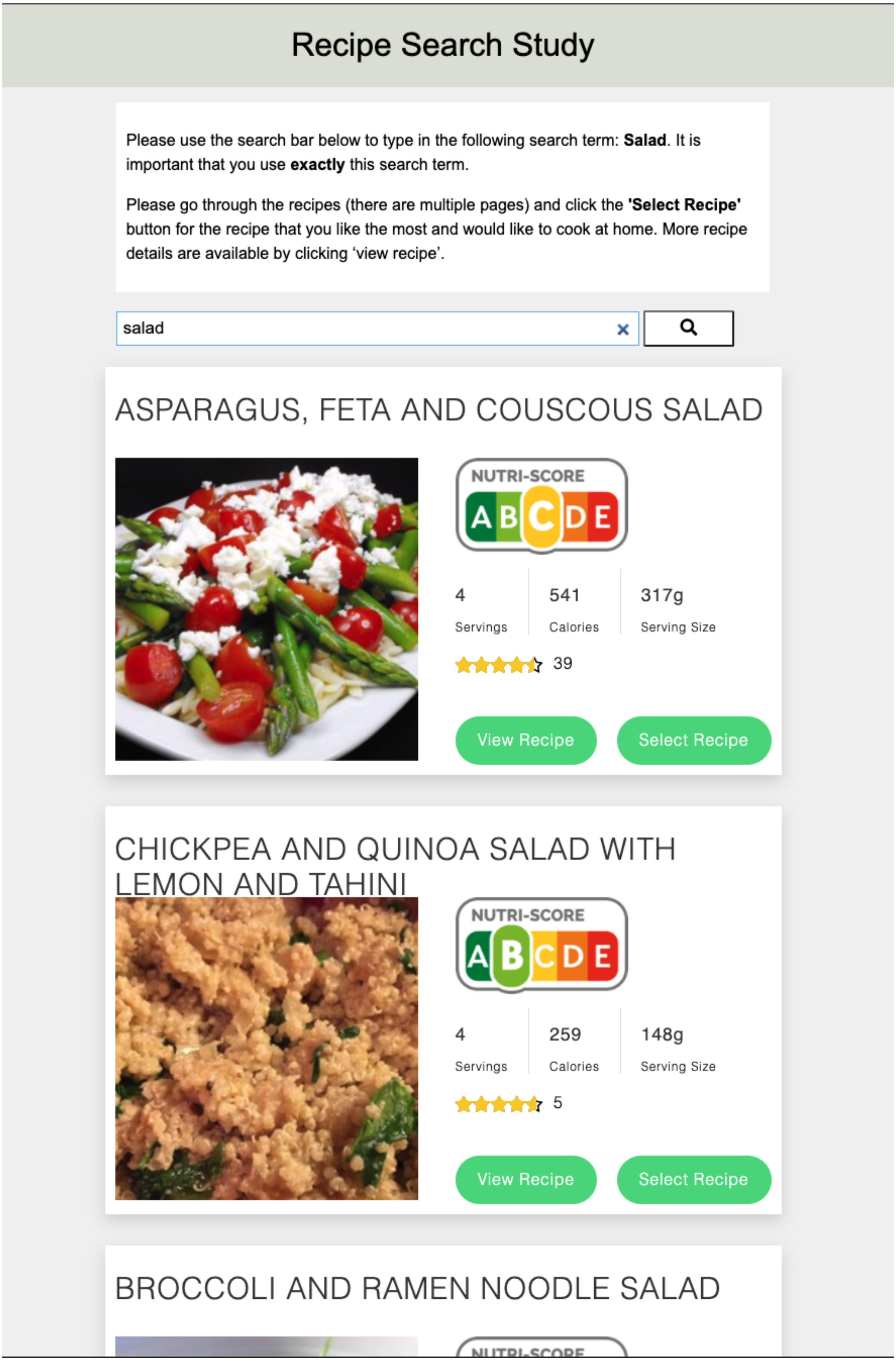
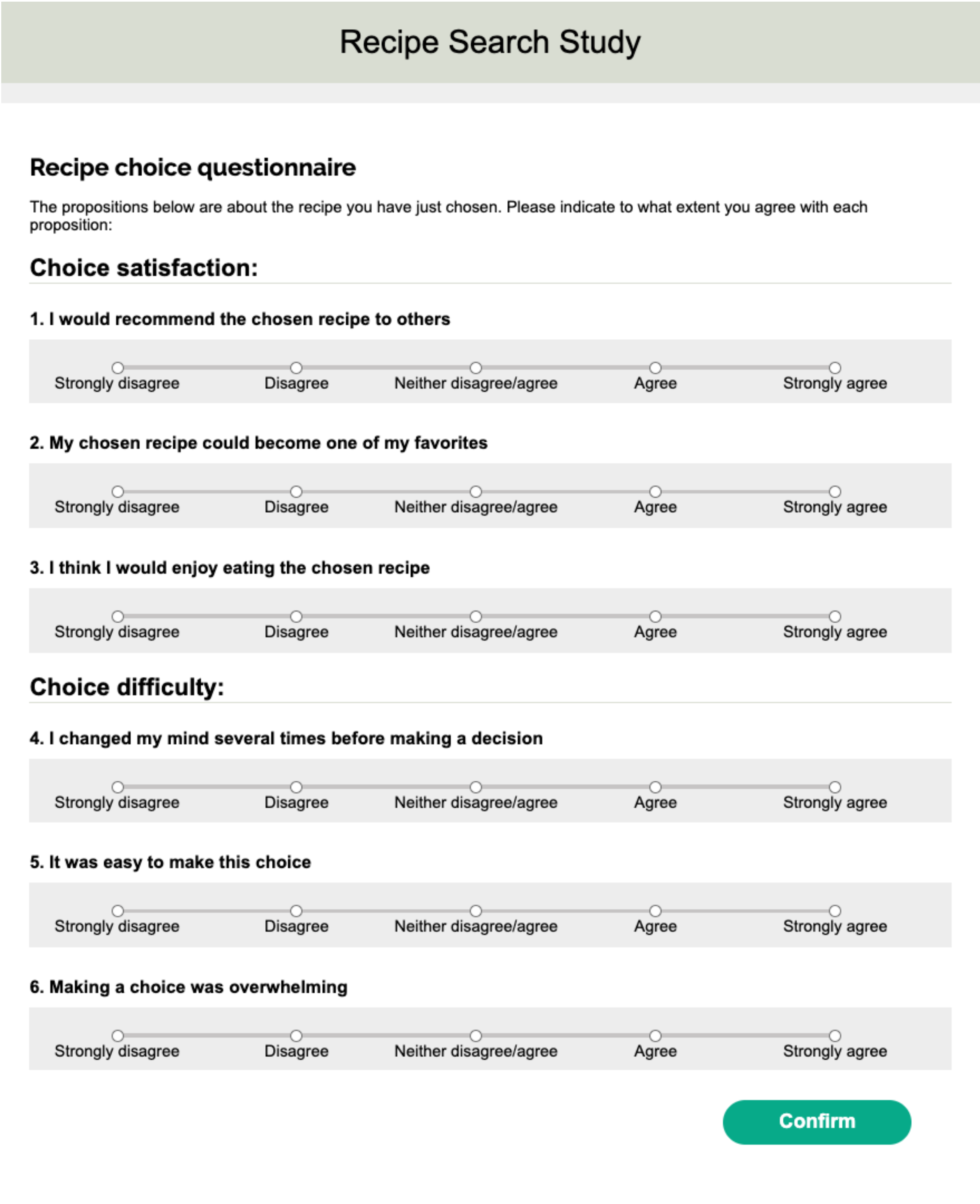


Figure 5.5: Search results for the keyword "Salad" with a Nutri-Score label on the recipe card.

Step 4: Survey About Chosen Recipe


Recipe Search Study

Recipe choice questionnaire

The propositions below are about the recipe you have just chosen. Please indicate to what extent you agree with each proposition:

Choice satisfaction:

1. I would recommend the chosen recipe to others

Strongly disagree | Disagree | Neither disagree/agree | Agree | Strongly agree

2. My chosen recipe could become one of my favorites

Strongly disagree | Disagree | Neither disagree/agree | Agree | Strongly agree

3. I think I would enjoy eating the chosen recipe

Strongly disagree | Disagree | Neither disagree/agree | Agree | Strongly agree

Choice difficulty:

4. I changed my mind several times before making a decision

Strongly disagree | Disagree | Neither disagree/agree | Agree | Strongly agree

5. It was easy to make this choice

Strongly disagree | Disagree | Neither disagree/agree | Agree | Strongly agree

6. Making a choice was overwhelming

Strongly disagree | Disagree | Neither disagree/agree | Agree | Strongly agree

Confirm

Figure 5.6: Participants had to complete this survey after each of the three recipe choices in the study.

Step 5: End Survey

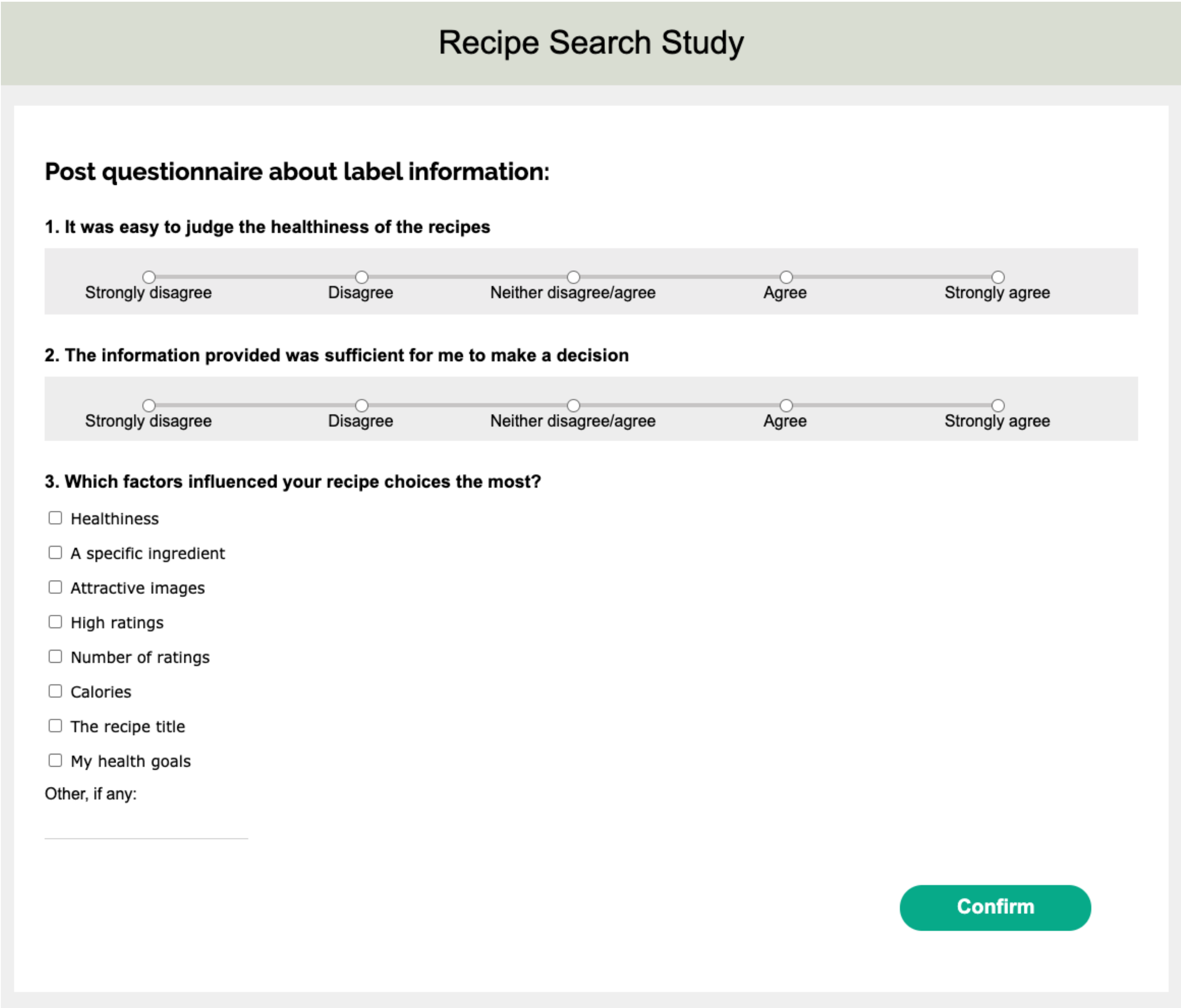


Figure 5.7: Participants had to complete this survey at the end of the study.

5.5 Exploratory data analysis

This section of the appendix contains graphs showing the top 8 answers to the question in the end survey.

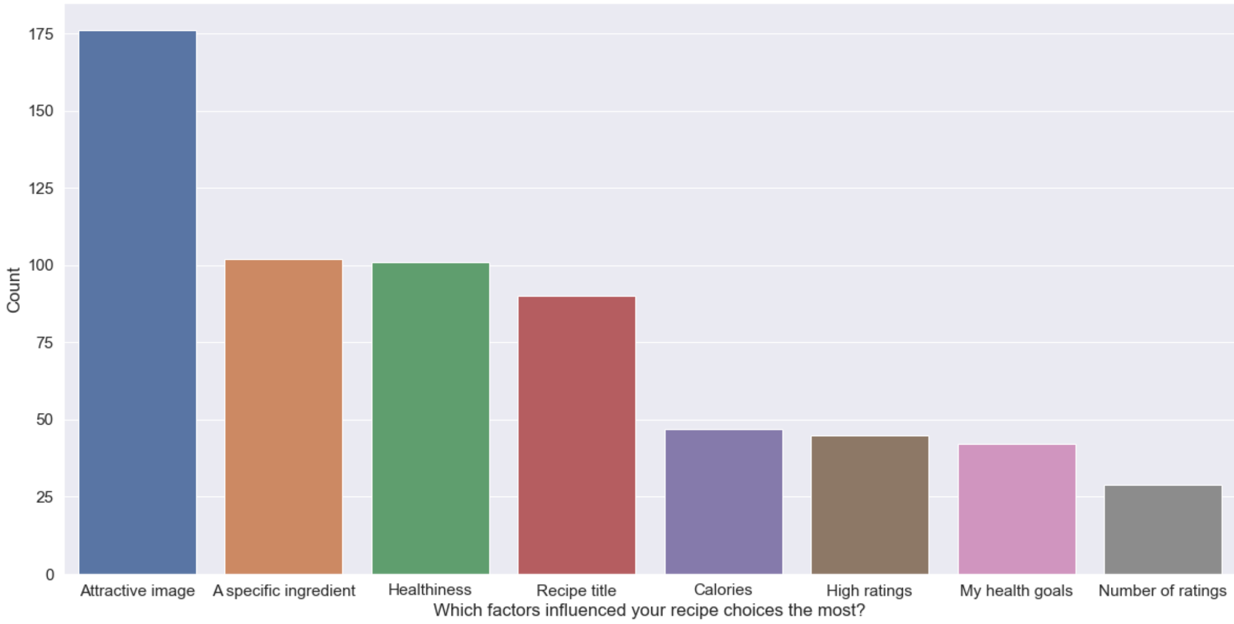


Figure 5.8: Count plot of the most selected answers to the question "Which factors influenced your recipe choices the most?" in the end survey for condition 1: Nutri-Score label.

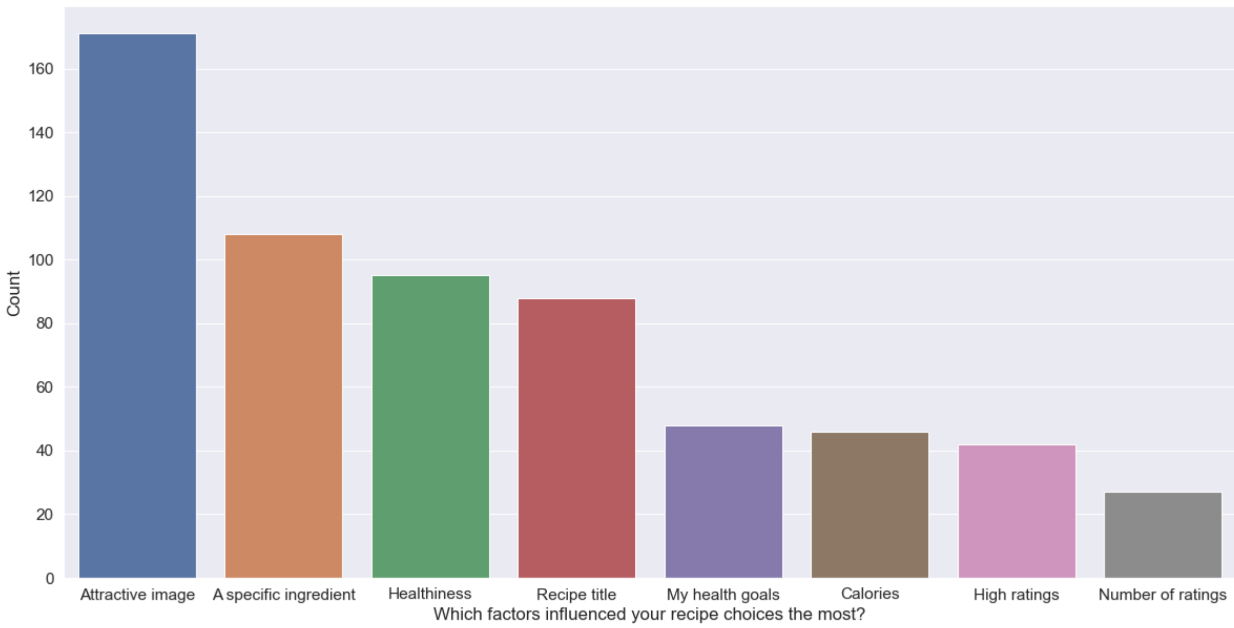


Figure 5.9: Count plot of the most selected answers to the question "Which factors influenced your recipe choices the most?" in the end survey for condition 2: Multiple Traffic Light label.

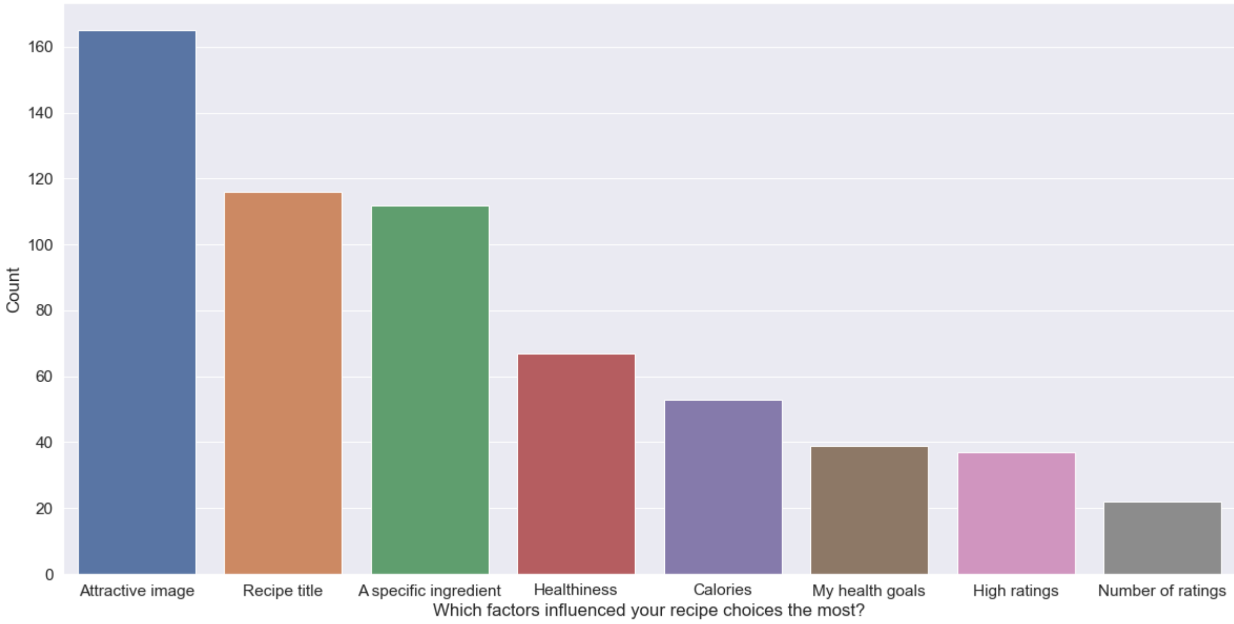


Figure 5.10: Count plot of the most selected answers to the question "Which factors influenced your recipe choices the most?" in the end survey for condition 3: No-label.