

Boosting Health? Examining the Role of Nutrition Labels and Preference Elicitation Methods in Food Recommendation

Alain Starke^{1,2}, Ayoub El Majjodi² and Christoph Trattner²

¹Marketing and Consumer Behaviour Group, Wageningen University & Research, Wageningen, The Netherlands

²Department of Information Science and Media Studies, University of Bergen, Bergen, Norway

Abstract

How users evaluate a recommender system goes beyond the accuracy of the presented content. For food recommendation, users differ in terms of the needs they have. We investigated whether users with different levels of health consciousness evaluated food recommender interfaces differently, depending on two factors: the Preference Elicitation (PE) method and the use of a nutrition label ‘boost’, which is a nudge that is explained to the user. In an online study (2x2 between-subjects design; $N = 244$), we compared a constraint-based recipe recommender, with feature-based PE, to a collaborative filtering recipe recommender with rating-based PE. Recipes were either annotated with a multiple traffic light nutrition label (i.e. the boost), or not (i.e., baseline). We found that boosts led to healthier recipe choices across both methods of PE. Moreover, we found users to be less satisfied with the constraint-based PE, while this may depend on the user’s level of health consciousness.

Keywords

Personalization, health, food recommendations, digital nudges, nutrition labels

1. Introduction

Most recommender systems assume that people have both the capabilities and interest to make fully-informed choices. That is, interaction data such as ratings and bookmarks are assumed to be accurate reflections of one’s preferences [1, 2]. While this goes a long way in some domains, for example in movies, recommenders in other domains face users that have specific needs or wishes that they cannot always disclose to the system [3], or require users to be more experienced to make well-informed decisions [4, 5, 6].

In food recommender systems, which present personalized food or recipe content to users [7], one’s preferences may strongly depend on contextual factors. This not only includes the ‘in-person’ context, such as the time of day and allergies, but also *how* the food items are presented in the digital interface. For example, one’s preferences for a specific burger recipe may strongly depend on whether salad recipes are presented alongside it, or whether the nutritional content

IntRS’22: Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, September 22, 2022, Seattle, US (hybrid event)


✉ alain.starke@wur.nl (A. Starke); ayoub.majjodiu@uib.no (A. E. Majjodi); christoph.trattner@uib.no (C. Trattner)

🌐 <https://www.christophtrattner.info/> (C. Trattner)

🆔 0000-0002-9873-8016 (A. Starke); 0000-0002-7478-5811 (A. E. Majjodi); 0000-0002-1193-0508 (C. Trattner)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

of that burger is emphasized. Moreover, some users may find it difficult to elicit their preferences if they have specific needs [8]. For example, a user of a collaborative filtering recommender that only optimizes for ratings may find it difficult to locate recipes without specific features, such as gluten-free content.

One overarching theme in food recommender systems is to support healthier choices [9]. Most food recommenders are, however, still optimized towards popularity [10, 8], leading to unhealthy outcomes [11]. The number of food recommender studies that examine how to optimize for a user's nutritional needs, such as through knowledge-based methods [12], is rather small [9, 8]. Even so, most of the research has focused on algorithmic advancements in terms of prediction accuracy and less so on the healthiness of chosen recipes and the user's evaluation [13, 14].

In this paper, we examine recipe choices and a user's evaluation for two recommender aspects that go beyond algorithmic accuracy and the presented content. First, we investigate to what extent 'boosting' can support healthier recipe choices in a food recommender context. Like nudges [15, 16], boosts are changes to a choice architecture that lead to predictable changes in behavior [17]. Whereas nudges can also be unconscious and a user is not always aware of them [18, 16], boosts aim to empower users in their decision-making by increasing their competence [17], typically by providing more information [19]. In this sense, boosts tend to be regarded as ethically more acceptable, due to their explicitness [20]. A common approach to 'boost' healthy food choices in brick-and-mortar supermarkets is the use of front-of-package labels [21, 22]. Such labels summarize the nutritional content of a product, indicating how consuming it relates to one's allowed daily intake for different nutrients [23]. A commonly used label in the United Kingdom, is the Multiple Traffic Light label, which has shown to be effective in supporting 'offline' healthy food choices [24]. However, their effectiveness in an online context, as well as for recipes, is less clear.

Second, we investigate the role of the used preference elicitation method on a user's evaluation of a food recommender system. Whereas content-based and CF-based recommender typically ask users to interact with individual items (cf. [25, 26]), other types of recommenders seek to exploit the relation between user characteristics and recipe features, such as knowledge-based recommender systems [27, 12]. Not only does this affect what items are presented, but possibly also on how users perceive and experience the interaction. Previous research in the energy recommender domain has shown that the interplay between the used preference elicitation method and a user's domain knowledge affects the user's evaluation [4]. They show that inexperienced users tend to prefer to rate the favorability of individual items (i.e., in this case: energy-saving measures), while more experienced users could interact with the measures' features (e.g., 'effort' and 'savings') [28, 4]. For the recipe domain, this implies that preference elicitation through recipe features, as in a constraint-based recommender, would be preferred by users with a high level of experience.

We argue that the extent to which users are interested in recipe healthiness can affect how they evaluate a recommender system and its preference elicitation method. On the one hand, in a collaborative filtering context, users can only indicate their interest in healthier recipes by rating specific recipes, while this might be easier in a recommender system that inquires on specific nutrition-related features, such as knowledge-based or constraint-based recommenders. On the other hand, users who are aware of health and nutrition might be able to pick specific

recipes that fit their needs and would experience feature-based elicitation as not fully meeting their needs. To this end, we consider a user’s level of health consciousness, which measures one’s perception of one’s diet and the relation between nutrient intake and health [29]. This aspect is adapted from pre-validated scales, used in nutritional studies [30, 29].

Approaches in food recommender systems promotes popular and unhealthy content, while user preferences tend to be more complex to be extracted. We propose the following research questions:

- RQ1: To what extent does a ‘nutrition label boost’ steer users towards healthier recipe choices in a recommender system context?
- RQ2: To what extent does a user’s evaluation of a food recommender system depend on the interplay between a user’s health consciousness and the system’s preference elicitation method?

We present an online recommender study in which users can disclose their preferences for recipes, after which they are presented a personalized recommendation list. By comparing recipe lists with and without nutrition labels and by using different preference elicitation methods, we show that:

- Healthier recipe choices can be supported by boosts, without changing the recommended content.
- A user’s perception (i.e., effort) and evaluation (i.e., choice difficulty and satisfaction) are more favorable among users of a constraint-based recommender with a low level of health consciousness, and vice versa for a collaborative filtering recommender.

2. Methodology

2.1. Dataset

We consulted a database that comprised recipes of Allrecipes.com, used in previous food recommender systems studies (e.g., [11, 14]). From the total of 58000+ recipes, we extracted a sample of 991 recipes. Our dataset included the basic metadata for each recipe, such as image URLs, serving sizes, the number of ingredients, preparation times, calories, sugar, salt, (saturated) fat, and protein. Table 1 presents the number of recipes per food category, which were selected because they contained metadata on features required to generate recommendations.

Table 1

Allrecipes.com dataset used for algorithm training and the user study.

Recipe Category	Number of Recipes
Meat and Poultry	444
Fruit and Vegetables	339
Barbecue	123
Pasta, Noodles and Seafood	85

2.2. Recommender Approaches

To address our research questions, and specifically RQ2, we compared two recommender approaches that were distinct in terms of their preference elicitation (PE) methods¹. Collaborative Filtering (CF) relies on rating-based PE, asking users to indicate preferences for individual items (i.e., recipes). Such approaches tend to outperform other item-based PE methods, such as content-based recommendation [31]. In contrast, Constraint-based (CB) recommendation exploits user preferences for recipe features, retrieving content based on the relation between user characteristics and recipe features. Both of the selected approaches involve explicit preference elicitation, as this was found to be the best representation of user preferences in food domain [8].

2.2.1. Collaborative Filtering (CF)

Before implementing the CF-based recommender, we evaluated several rating-based prediction algorithms in an offline setting using our dataset. The results of this analysis were also reported in [32]. Singular Value Decomposition (SVD) [33] was found to outperform algorithms (e.g., SVD++, KNNBaseline, NMF) by 10% in terms of the Root Mean Squared Error and Mean Absolute Error, and was deployed for our online evaluation.

As part of the study, users were presented 10 recipes to rate on a 5-point scale. These recipes were all part of a preferred cuisine by the user (cf. subsection 2.3). Subsequently, a list of ten recipes was retrieved that was closest to the inferred user profile, based on the SVD recommender. Five recipes were retrieved from a healthy set, and the other five were retrieved from a less healthy set (cf. subsection 2.5.1).

2.2.2. Constraint-based (CB)

Our CB recommender inquired on preferred user constraints for the recipe recommender. Rather than relying on the relation between user characteristics and recipe features, such as was done in the knowledge-based recommender of Musto et al. [12], we focused on ‘pure’ feature-based PE, which was consistent with Knijnenburg et al. [4]. The recommendation process was initiated by asking users what type of recipes they preferred, based on the food category and different features. Features addressed different aspects, such as practicalities (i.e., number of servings) and health (i.e., preferred amount of calories). An overview of features is depicted in Figure 1. After obtaining feature-based preferences, a similarity function was used to score recipes (based on [27]), eventually retrieving recipes that were deemed most relevant.

2.3. Research Design and System Procedure

Users were subject to 2 (Preference Elicitation (PE): Collaborative Filtering (CF), Constraint-Based (CB)) X 2 (Labelling conditions: No label, Boost) between-subject design. For one arm, users either interacted with a CF-based or a CB system, which differed in terms of PE and the recommender algorithm. For the other arm, users either interacted with educational pages about the use of Multiple Traffic Light (MTL) nutrition labels, before being presented personalized

¹Materials used for this study: https://github.com/ayoubGL/Boosting_TowardsHealth

What are your recipe preferences ?

Please select the food category that like the most, then answer carefully the following questions. You will receive personalized recommendations according to your preferences.

Food category

I want recipes at least with 3 stars 4 stars No preferences

The preferred number of servings in my recipes are

Preferred amount of calories in my recipes

The time I have available for cooking (in min)

The preferred number of ingredients in my recipes

Next

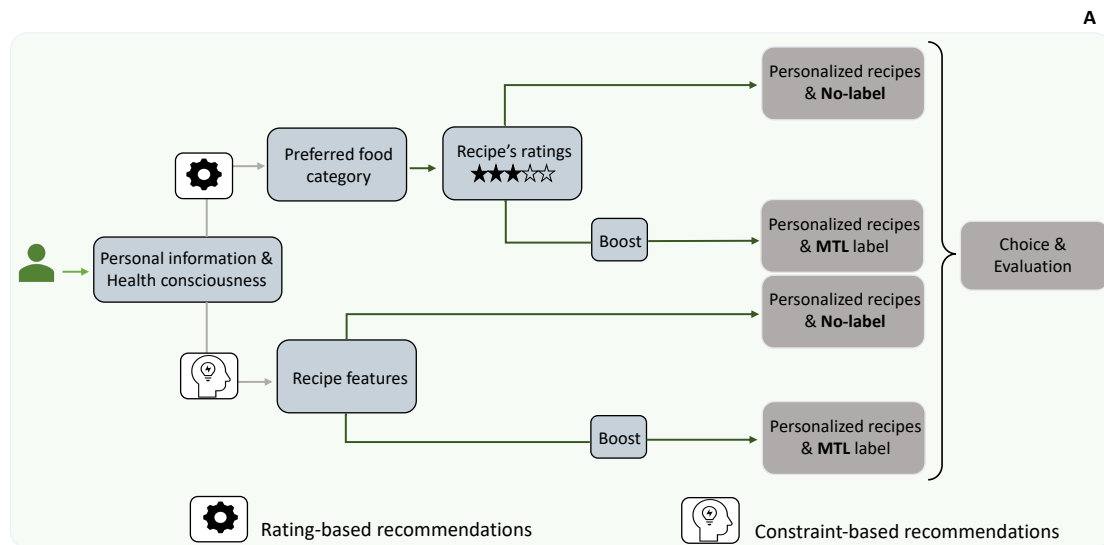
Figure 1: Features extracted for the Constraint-based preference elicitation and recommender approach.

recipes annotated with MTL labels (i.e., Boost condition), or were not exposed to any education or label. Figure 2B depicts an example of an MTL label, Figure 3 depicts the educational prompt of the boost.

For the online evaluation, users were asked to provide their consent for participation. They were informed that our food recommender system would help them to find recipes they would like to cook and eat. Figure 2A depicts the user flow of the proposed system. First, users were asked to disclose basic demographic characteristics (e.g., age, gender, level of education) and to respond to questionnaire items about their level of health consciousness, as well as to choose a preferred food category. In the CF scenario, users were asked to rate ten recipes from the preferred category using a 5-star rating scale. In the CB scenario, the user filled out a form expressing her needs in terms of desired recipe features (see Figure 1). Users in both conditions were presented a list of ten personalized recipe recommendations, among which five were relatively healthy (i.e., having an FSA score of 8 or lower) and five relatively unhealthy (having an FSA score of 9 or higher). Recipes were either annotated with the Multiple Traffic Light (MTL) nutrition labels or not, according to the intervention conditions. Afterwards, users were asked to evaluate their perception of the system and their experience with the chosen recipes.

2.4. Participants

A total of 244 participants (75% female) completed our 5-minute study, for which they were rewarded with GBP 0.75. They were recruited on the crowdsourcing platform Prolific. They had at least an approval rate of 95% and previously completed at least 30 submissions. Among them, 99% had attained at least a high school diploma. Participants all lived in Great Britain, as



B

Fiery Fish Tacos with Crunchy Corn Salsa					
Servings			Serving Size		
8			300 (g)		
Cal	Sugar	Fat	Sat Fat	Salt	
351	0.30g	15.30g	4.60g	0.53g	
Kcal	Low	Medium	Medium	Low	
<input type="button" value="Select Recipe"/>					

Figure 2: A (on the top): User flow of the online evaluation. B (on the bottom): An example of a recipe that was annotated with a Multiple Traffic Light label.

they were more likely to have experience with the Multiple Traffic Light Label. Note that the research setup conformed to the ethical standards of the Norwegian Centre for Research Data (NSD).

2.5. Measures

2.5.1. Recipe healthiness

We assessed the healthiness of chosen recipes. In other studies, this was typically based on nutritional guidelines proposed by various health organizations [34, 35]. We used the well-validated FSA score [36], which was issued by the British Food Standards Agency [37]. The FSA score was computed by assigning points to four nutrients in a given recipe: sugar, fat, saturated fat, and salt. For each nutrient, we discerned between low, medium, or high content, assigning one point for each level (low, medium, high). This led to a scored scale from 4 (healthiest) to 12 (least healthy). We used a score of eight as a threshold to discern 50% healthy and 50% unhealthy recipes in our recommendation sets.

Understanding the use of nutrition labels

On the next page, you will be presented personalized recipes recommendations, along with Multiple Traffic Light nutrition labels. Please carefully read the text below to understand what they mean before proceeding to the next page.

Nutritional labels give you information that can help you make healthier and more informed choices when deciding which food products to buy: "By checking the label each time you purchase something, you will take more control of your eating habits."

The traffic light labelling system will tell you whether a recipe has high, medium or low amount of fat, saturated fat, sugars and salt. It will also tell you how much of each nutrient a recipe contains per serving.

- **Red:** means the product is high in a nutrient and you should try to cut down, eat less often or eat smaller amounts.
- **Amber:** means medium, if a food contains mostly amber, you can eat it most of the time.
- **Green:** means low, the more green lights a label shows, the healthier the food choice is.

Next

Figure 3: Prompt to Boost a user's understanding of the Multiple Traffic Light nutrition labels.

2.5.2. Nutrition Labels

For our boosting intervention, we relied on a Front-of-Package Nutrition Label to inform users about the health content of recipes. For years, food products displayed nutritional details on the back of packaging. Although this was found to be associated with low-fat food intake and healthier food choices overall [38, 39], many people found the information too complex to use [40, 41]. This spurred the development of Front-of-Package labels that summarize a product's nutritional content [42].

In this study, we used the Multiple Traffic Light (MTL) Front-of-Package nutrition label as a healthy eating boost (see Figure 2B). The MTL label was based on the FSA score, representing different levels of nutrient content by displaying red, amber, or green colors for high, medium, or low levels of nutrients, respectively [37].

2.5.3. User Characteristics, Perception, and Experience

To examine the user's evaluation of our food recommender approaches, we inquired on user characteristics and evaluation aspects. In line with the recommender system user experience framework [43, 44], we examined perception and experience aspects and user characteristics. Item responses were submitted to 5-point Likert Scales. Items for choice satisfaction [43, 45, 14], choice difficulty [46, 14], and perceived effort [6] were adapted from previous recommender studies. Item for health consciousness was adapted from a pre-validated scale in the food domain [30, 29], in line with a procedure followed by [47].

Whether the items formed the expected aspects was examined using a principal component factor analysis. Table 2 outlines the results, describing the factor loadings for the used aspects and items. Items with low loadings or too many cross-loadings were removed from analysis. Whereas choice difficulty, choice satisfaction and perceived effort could be inferred reliably,

there were doubts about the reliability of health consciousness. We observed a low value of Cronbach’s Alpha (0.37), even after dropping unreliable items. Since the used items were part of a pre-validated scale and the factor loadings with the retained items were good, we decided to proceed with our analyses including health consciousness.

For our analyses, all aspects were standardized and predicted using regression scoring. Health consciousness was considered as both a continuous and dichotomous variable. For the latter, we differentiated between low and high levels of health consciousness, performing a mean split on the standardized variable.

Table 2

Results of the principal component factor analysis across different user characteristics and experience aspects. Items were measured on 5-point Likert scales. Cronbach’s Alpha is denoted by α , items in gray were omitted from analysis.

Aspect	Item	Loading
Choice Satisfaction $\alpha = .86$	I like the recipe I have chosen.	.841
	I think I will prepare the recipe I have chosen.	.874
	The chosen recipe fits my preference.	.789
	I know many recipes that I like more than the one I have chosen.	.817
Choice Difficulty $\alpha = .75$	I would recommend the chosen recipe to others.	.817
	I changed my mind several times before making a decision.	.920
	Making a choice was overwhelming.	-.794
Health Consciousness $\alpha = .37$	It was easy to make this choice.	.651
	My diet is well-balanced and healthy.	.809
	The amount of sugar I get in my food is important.	.752
	I have the impression that I sacrifice a lot for my health.	.752
Perceived Effort $\alpha = .61$	My health does not depend on the food I consume.	.752
	I am concerned about the quantity of salt that I get in my food.	.752
	The system takes up a lot of time.	.782
	I quickly understood the functionalities of the system.	.843
	Many actions were required to use the system.	.843

3. Results

We present the analyses for our two research questions. First, we investigated to what extent annotating recipes with nutrition labels, with (i.e., boosting) or without (i.e., nudging) explanation, led to healthier recipe choices (RQ1). Second, we examined the interplay between the used Preference Elicitation (PE) method and user’s evaluation method, specifically examining how the user’s health consciousness and the PE method led to differences in user perception (i.e., perceived effort) and evaluation (i.e., choice difficulty and choice satisfaction; RQ2).

3.1. RQ1: Boosting Towards Healthier Choices

We first examined to what extent nutrition label boosted affected the healthiness of recipes chosen. We performed a two-way ANCOVA to predict whether the FSA score of chosen recipes

differed significantly across conditions, while adjusting for a user’s level of health consciousness. With regard to the labeling conditions, the results in Table 3 show that the FSA score was significantly lower in the boost condition ($M = 7.98$, $SD = 1.63$) than in the no-label condition ($M = 8.65$, $SD = 1.50$): $F(1, 239) = 10.41$, $p = 0.0014$. This showed that in the context of personalized recipe recommendations, boosting and annotating recipes with a multiple traffic light nutrition label leads to an increase in the healthiness of recipe choices.

Table 3

Results of a two-way ANCOVA, predicting the healthiness of chosen recipes across different recommendation approaches and interventions conditions. A user’s health consciousness was included as a covariate. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Factor (Predicting FSA score)	df	F
Model	4	4.45 **
Labelling Condition (No Label-Boost)	1	10.41**
Preference Elicitation (CF-CB)	1	0.79
Preference Elicitation * Labelling Cond.	1	1.25
Health Consciousness	1	1.76

The two-way ANCOVA reported in Table 3 further revealed that the healthiness of recipe choices did not depend on the Preference Elicitation (PE) method. Although chosen recipes were slightly less healthier after a constraint-based (CB) PE and recommendation method ($M = 8.39$, $SD = 1.39$) than after a collaborative filtering (CF) PE ($M = 8.23$, $SD = 1.79$), this difference was not significant ($p = 0.42$). This suggested that the used PE and recommendation method did not directly affect the recommended content.

In addition, we neither observed an interaction effect between the PE method and the use of a boosted nutrition label. To better understand all effects, please inspect Figure 4, which shows that users across both PE methods chose healthier recipes when being presented nutrition labels. Finally, Table 3 did not reveal that a user’s level of health consciousness significantly affected the healthiness of chosen recipes ($p = 0.057$); we further checked for interaction effects with the labelling conditions, but did not observe any.

3.1.1. Conclusion

Overall, we found that annotating recipes with multiple traffic light nutrition labels, in conjunction with an explanation, can support users in making healthier choices. On the other hand, the recommendation approach did not affect on the healthiness of chosen recipes (see Table 3), nor was it affected by a user’s level of health consciousness. Recipe choices were further related to how users evaluated our recommender system in the next subsection.

3.2. RQ2: User Evaluation of Preference Elicitation Methods

We examined a user’s evaluation of our recipe recommender system, based on the used Preference Elicitation (PE) method. In doing so, we first examined whether the user’s perception was affected by the interplay between a user’s health consciousness and the preference elicitation

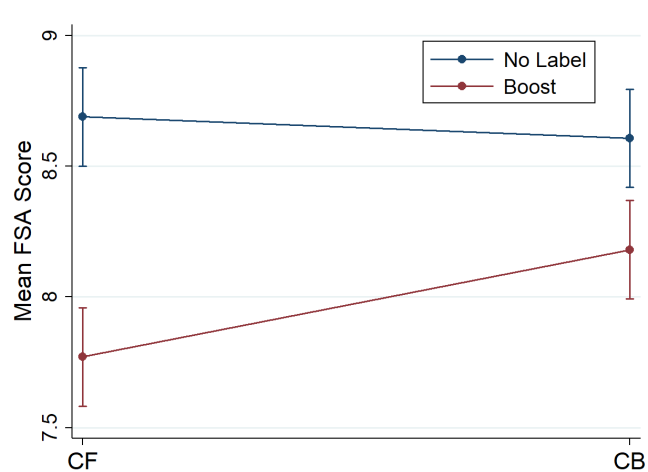


Figure 4: Mean FSA score of chosen recipes across different conditions for preference elicitation (CF vs CB) and labelling (No Label vs Boost). Lower scores indicate healthier recipe choices.

method (RQ2). Then, we examined whether this led to further differences in a user’s experienced choice difficulty and choice satisfaction.

3.2.1. Perceived Effort

We performed a one-way ANCOVA on a user’s perceived effort of using our recommender system, including an interaction effect between health consciousness and the used PE method². The results are outlined in Table 4. We found no main effects for the used elicitation method, as the perceived effort of using the CB recommender ($M = -.014, SD = 1.03$), which relied on disclosing preferences for recipe features, was only somewhat lower than that of the CF-based recommender ($M = .014, SD = .97$), which relied on rating-based PE. In addition, we neither observed a main effect of a user’s health consciousness on effort: $F(1, 240) = 1.16, p = 0.28$.

Table 4

Results of a one-way ANCOVA with interaction effect on the user’s level of perceived effort. It was examined across different PE conditions (CF, CB), while controlling for a user’s level of health consciousness. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Factor (Predicting Perceived Effort)	df	F
Model	3	3.02*
Preference Elicitation (CF, CB)	1	0.08
Health Consciousness	1	1.16
PE method (CF, CB) * Health Consciousness	1	8.42**

What stands out from Table 4 is an interaction effect between the PE method and a user’s health: $F(1, 240) = 8.42, p = 0.0041$. This suggested that a user’s perceived effort depended on

²We also examined whether the user’s perceived effort differed across labelling conditions, such as by performing a two-way ANOVA across different labelling and PE conditions. However, we observed no differences.

the interplay between the PE method and the user's level of health consciousness. The direction of this effect can be understood best by inspecting Figure 5, in which we differentiated between low and high levels of health consciousness based on a mean split. While users with low levels of health consciousness perceived the CB method as less effortful, this increased significantly for users with a high level of health consciousness. In contrast, Figure 5 depicts much smaller differences in perceived effort for a CF-based PE. This suggested that users who were likely to seek out healthier recipes found our constraint-based recommender, with feature-based elicitation, more effortful to use.

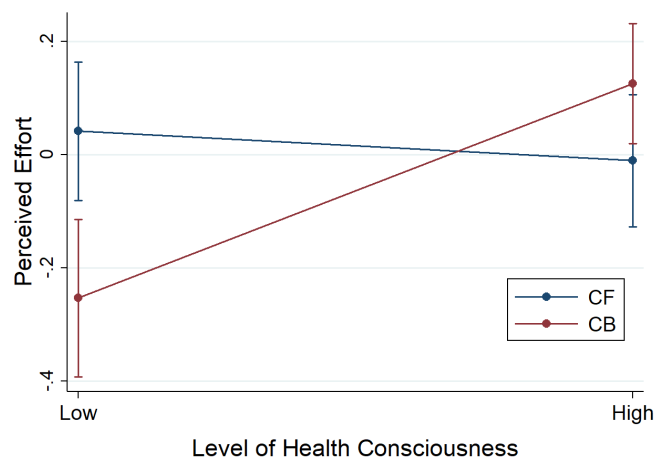


Figure 5: Users' perceived effort of using the different Preference Elicitation (PE) methods, based on a user's level of health consciousness. The Collaborative Filtering (CF) PE was rating-based, while the Constraint-Based (CB) PE was feature-based. To interpret the interaction effect in Table 4, health consciousness was divided into low and high based on a mean split. Error bars are 1 S.E.

3.2.2. Choice Difficulty and Choice Satisfaction

We further examined to what extent different elicitation methods and user characteristics affected the user experience aspects of choice difficulty and choice satisfaction. For each aspect, we performed a linear regression analysis, in which we also checked for differences across labelling conditions, as well as whether perception aspects (i.e., effort) and choice metrics (i.e., FSA score) played a role.

Table 5 reports both analyses. For choice difficulty, we found that users of the constraint-based recommender found it more difficult to use ($\beta = .31, p = 0.01$), compared to users of our CF-based recommender. In contrast, choice difficulty was not affected by the use of nutrition labels (i.e., our boost), neither by the user's level of health consciousness, nor by the interaction between the PE method and the user's health consciousness (all $p > 0.05$). This suggested that it was more difficult to choose between the recipes generated by the constraint-based recommender (compared to CF), while the use of labels did not support easier decision-making.

With regard to other aspects, we found that users who perceived a recommender as effortful to use, also reported higher levels of choice difficulty: $\beta = .26, p < 0.001$. This suggested a

possible indirect effect of the interplay between a user's level of health consciousness and the PE method on choice difficulty, via perceived effort. Hence, an earlier analysis revealed that users in the CB condition reported higher levels of perceived effort if they had a higher level of health consciousness, and vice versa. In contrast, the healthiness of the chosen recipe (i.e., FSA) was not related to choice difficulty.

Table 5

Linear regression analyses, with models to predict the user's experienced choice difficulty and choice satisfaction, based on the experimental conditions, user characteristics, perception aspects and choices. 'Boost' and 'CB' were coded as 0.5, 'No label' and 'CF' as -0.5. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Factor	Choice Difficulty		Choice Satisfaction	
	β	S.E.	β	S.E.
Labelling Condition (Boost vs No label)	.11	.13	.13	.13
Preference Elicitation (CB vs CF)	.31*	.12	-.43**	.12
Health Consciousness	.050	.062	.051	.062
Preference Elicitation (CB vs CF) * Health Consciousness	.0075	.13	-.021	.13
Choice Difficulty			-.21**	.065
Perceived Effort	.26***	.063	.033	.065
FSA	-.045	.040	.066	.040
Intercept	.37	.33	-.55	.34
R^2		.11***		.11***

With regard to choice satisfaction, we observed similar effects as for choice difficulty. Again, we observed no significant effects for the use of nutritional labels, health consciousness and the chosen recipe's FSA score. In a similar vein, users reported lower levels of choice satisfaction for our constraint-based recommender with feature-based PE ($M = -.24$, $SD = 1.02$), than for our rating-based CF recommender ($M = .24$, $SD = .92$): $\beta = -.43$, $p = 0.001$. In addition, we also observed a negative, significant relation between the experienced choice difficulty and choice satisfaction: $\beta = -.21$, $p = 0.002$. This suggested two possible mediated paths towards choice satisfaction. First, the constraint-based PE method increased the experienced choice difficulty and, in turn, lowered the user's level of experienced choice satisfaction. Second, the interaction between a user's health consciousness and the PE method affected effort, which affected choice difficulty and satisfaction subsequently. All the effects, regarding the experimental conditions, can be understood further by inspecting Figures 6 and 7.

3.2.3. Conclusion

We observed that the user's evaluation of our food recommender system depended on the interplay between the PE method and user characteristics. Specifically, we observed that if a user's level of health consciousness was high, users perceived higher levels of effort of using the constraint-based recommender, compared to lower levels of effort for users with a low level of health consciousness. Effort was further found to positively affect choice difficulty, which had, in turn, a negative relation with choice satisfaction. On top of that, users had on average more

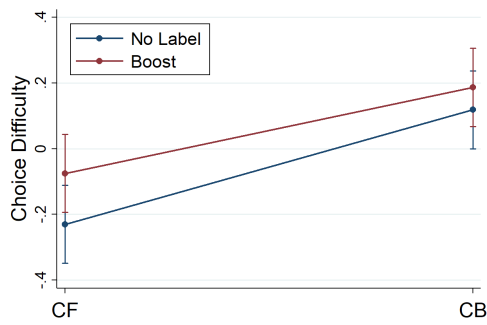


Figure 6: Standardized scores for the choice difficulty experience aspect across conditions. Errors bars represent 1 S.E.

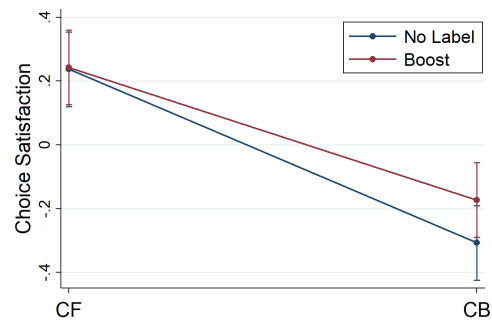


Figure 7: Standardized scores for the choice satisfaction experience aspect across conditions. Errors bars represent 1 S.E.

difficulties in choosing a recipe in the constraint-based condition, while we observed no effects on the used labeling condition. This showed that how a food recommender is evaluated depends on the healthy food interests of the user, which is one of the possible user characteristics that could have been inquired on.

4. Discussion

In the context of personalized recipe recommendations, this work has examined two ways how a recommender can cater to users who are interested in healthy eating. Food recommender systems have faced difficulties in optimizing for nutritional content [9, 10, 11], particularly while maintaining a user’s level of satisfaction [36]. In an attempt to ‘boost’ healthier recipe choices, we have gone beyond optimizing algorithmic accuracy by not necessarily changing what recipes are presented but how they are presented and how user preferences are elicited.

First, we have found that annotating recipes with nutrition labels leads to healthier recipe choices (RQ1). Our work is among the first to examine such a digital nudge in a personalized context [48], particularly in the domain food [49], and one of the first to use the concept of ‘boosting’ in a recommender system context. The idea that our (interface) interventions, not only the algorithm, should be explainable to a person or user is gaining ground in behavioral economics [17, 20]. The purpose of highlighting a specific interface this way is to make it more salient to a user, which can increase one’s knowledge level or awareness [50]. Although it seems sensible that a nutrition label can support healthier choices [40], evidence for the effectiveness of nudges in personalized recommender contexts is scarce [45]. It seems that they need to be well-designed to be effective, as the content already fits the user’s preferences.

Second, we have found that the user’s evaluation of a food recommender system is significantly associated with the used preference elicitation method, based on that user’s level of health consciousness (RQ2). We have compared two distinct recommender approaches, collaborative filtering and constraint-based, that also involve different methods of preference. Across all types of users, we find that constraint-based recipe recommendation lists are more difficult to choose

from and, in turn, lead to lower levels of satisfaction. With regard to choice satisfaction, it could be argued that this outcome is unsurprising, since a constraint-based recommender system tends to be less effective than a CF-based recommender as it makes much simpler assumptions about the user's preferences [51].

When combining health consciousness and the PE method, we would have expected to observe an additional interaction effect. Hence, our user-specific analyses are more striking. Similar to the interaction between knowledge and the PE method in [28], we have observed an interaction effect between health consciousness and the PE method. However, although we have observed *lower* perceived effort for lower levels of health consciousness and feature-based PE, previous findings from the energy domain show the opposite with *higher* system satisfaction levels for high domain knowledge and feature-based PE [28, 4]. We argue that the examined interplay of user characteristics and PE is different. Whereas the work of Knijnenburg et al. [4] focuses on domain knowledge and the resulting understandability of different energy-related features, the current paper focuses on whether a specific aspect in which users can be interested (i.e., food healthiness) can be catered to effectively using different PE methods. Whether similar findings would be observed when examining the relation between PE and food knowledge (e.g., subjective food knowledge [52, 53]) will be examined in an upcoming study. Regardless, our findings stress the complexity and multifacetedness of the food domain [7], the domain-specificity of findings in recommender studies.

The main limitation to our findings is that health consciousness faced construct validity issues. Although the factor loadings of the eventually used items are good, the observed Cronbach's Alpha was found to be too low and multiple items were dropped. This can raise some doubts about whether we have measured health consciousness reliably. Although we wholeheartedly recommended to replicate our findings, we have proceeded with our analysis in the current paper, because the used items are part of a pre-validated scale [30], implying that some studies have already used the scale without validating it further (e.g., [47]).

Another aspect that can be improved is the method of analysis. Whereas Knijnenburg et al. [4] has examined mediation effects using structural equation modelling, this was not possible in the current study due to fit issues. We have not been even to converge to a model that would meet all assumptions. Instead, we have examined a simplified version of a structural equation model, by examining mediation through multiple separate analyses (i.e., ANCOVAs, regression). This approach is in line with the earlier 2009 RecSys work of Knijnenburg and Willemsen [28], inferring mediation in a stepwise manner. This approach is also prescribed by Baron and Kenny [54]. Nonetheless, it has been argued that this approach faces more limitation than a structural equation model would [55]. Hence, in a follow-up study, we intend to mitigate the issues regarding construct validity by examining our findings in a structural equation model by using different questionnaire items. Even though health consciousness has been adapted from previous studies [30, 29], we opt for pre-validated aspects in a follow-up study. In doing so, we will also consider subjective food knowledge scales, which would be in line with the earlier work on PE from Knijnenburg et al. [4]. All in all, this would allow us to paint a full picture of the interplay between a food recommender's PE method and multiple relevant user characteristics that cannot simply be captured by a recommender algorithm.

A limitation regarding the recommended items is the arguably smaller dataset of recipes. While some other studies with internet-sourced recipes have been able to leverage datasets

with more recipes [11], our dataset is smaller due to the focus on high-quality ratings for our collaborative filtering recommender. We have only used ratings from users that have rated at least 20 recipes, to make sure that the collected ratings are of rather active and experienced users. In addition,

Follow-up studies should also address the limitations of one-off user choices as a measure of recommender ‘success’. While food choices may go a long way in digital interfaces to predict subsequent user behavior (e.g., [56]), an optimal study design would also check whether chosen recipes are actually prepared and consumed. In a sense, repeated interactions with a recommender systems through some kind of application would be representative to assess whether user preferences would shift. Therefore, we opt for a recommender systems with a longitudinal design, as has been demonstrated in other domains [57].

Acknowledgments

This work was supported by industry partners and the Research Council of Norway with funding to MediaFutures: Research Centre for Responsible Media Technology and Innovation, through the centers for Research-based Innovation scheme, project number 309339.

References

- [1] W. Hill, L. Stead, M. Rosenstein, G. Furnas, Recommending and evaluating choices in a virtual community of use, in: Proceedings of the SIGCHI conference on Human factors in computing systems, 1995, pp. 194–201.
- [2] J. Konstan, J. Riedl, Recommended for you, *IEEE Spectrum* 49 (2012) 54–61. doi:10.1109/MSPEC.2012.6309257.
- [3] M. D. Ekstrand, M. C. Willemsen, Behaviorism is not enough: better recommendations through listening to users, in: Proceedings of the 10th ACM conference on recommender systems, ACM, New York, NY, USA, 2016, pp. 221–224.
- [4] B. Knijnenburg, M. Willemsen, R. Broeders, Smart sustainability through system satisfaction: tailored preference elicitation for energy-saving recommenders, in: 20th Americas Conference on Information Systems (AMCIS 2014), August 7-9, 2014, Savannah, Georgia, United States, AIS/ICIS, 2014, pp. 1–15.
- [5] H. Schäfer, M. C. Willemsen, Rasch-based tailored goals for nutrition assistance systems, in: Proceedings of the 24th International Conference on Intelligent User Interfaces, 2019, pp. 18–29.
- [6] A. Starke, M. Willemsen, C. Snijders, Effective user interface designs to increase energy-efficient behavior in a rasch-based energy recommender system, in: Proceedings of the eleventh ACM conference on recommender systems, 2017, pp. 65–73.
- [7] C. Trattner, D. Elsweiler, Food recommendations, in: Collaborative recommendations: Algorithms, practical challenges and applications, World Scientific, 2019, pp. 653–685.
- [8] T. N. T. Tran, M. Atas, A. Felfernig, M. Stettinger, An overview of recommender systems in the healthy food domain, *Journal of Intelligent Information Systems* 50 (2018) 501–526.

- [9] D. Elswailer, H. Hauptmann, C. Trattner, Food recommender systems, in: F. Ricci, L. Rokach, B. Shapira (Eds.), *Recommender Systems Handbook*, Springer US, New York, NY, 2022, pp. 871–925. URL: https://doi.org/10.1007/978-1-0716-2197-4_23. doi:10.1007/978-1-0716-2197-4_23.
- [10] S. Mika, Challenges for nutrition recommender systems, in: *Proceedings of the 2nd Workshop on Context Aware Intel. Assistance*, Berlin, Germany, CEUR, 2011, pp. 25–33.
- [11] C. Trattner, D. Elswailer, 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*, ACM, New York, NY, USA, 2017, pp. 489–498.
- [12] C. Musto, C. Trattner, A. Starke, 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*, ACM, New York, NY, USA, 2020, pp. 333–337.
- [13] C. Musto, A. D. Starke, C. Trattner, A. Rapp, G. Semeraro, Exploring the effects of natural language justifications in food recommender systems, in: *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, 2021, pp. 147–157.
- [14] A. Starke, E. Asotic, C. Trattner, “serving each user”: Supporting different eating goals through a multi-list recommender interface, in: *Fifteenth ACM Conference on Recommender Systems*, ACM, New York, NY, USA, 2021, pp. 124–132.
- [15] 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 (2012) 487–504.
- [16] R. H. Thaler, C. R. Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*, Penguin, 2009.
- [17] R. Hertwig, T. Grüne-Yanoff, Nudging and boosting: Steering or empowering good decisions, *Perspectives on Psychological Science* 12 (2017) 973–986.
- [18] G. Loewenstein, C. Bryce, D. Hagmann, S. Rajpal, Warning: You are about to be nudged, *Behavioral Science & Policy* 1 (2015) 35–42.
- [19] L. M. König, B. Renner, Boosting healthy food choices by meal colour variety: results from two experiments and a just-in-time ecological momentary intervention, *BMC Public Health* 19 (2019) 1–15.
- [20] T. Rouyard, B. Engelen, A. Papanikitas, R. Nakamura, Boosting healthier choices, *bmj* 376 (2022).
- [21] M. Egnell, Z. Talati, S. Hercberg, S. Pettigrew, C. Julia, Objective understanding of front-of-package nutrition labels: an international comparative experimental study across 12 countries, *Nutrients* 10 (2018) 1542.
- [22] E. J. Van Loo, V. Caputo, R. M. Nayga Jr, W. Verbeke, Consumers’ valuation of sustainability labels on meat, *Food Policy* 49 (2014) 137–150.
- [23] N. J. Temple, J. Fraser, Food labels: a critical assessment, *Nutrition* 30 (2014) 257–260.
- [24] P. Ducrot, C. Julia, C. Méjean, E. Kesse-Guyot, M. Touvier, L. K. Fezeu, S. Hercberg, S. Péneau, Impact of different front-of-pack nutrition labels on consumer purchasing intentions: a randomized controlled trial, *American journal of preventive medicine* 50 (2016) 627–636.
- [25] D. Jannach, M. Zanker, A. Felfernig, G. Friedrich, An introduction to recommender systems,

New York: Cambridge 10 (2011) 1941904.

- [26] F. Ricci, L. Rokach, B. Shapira, Recommender systems: introduction and challenges, in: Recommender systems handbook, Springer, 2015, pp. 1–34.
- [27] C. C. Aggarwal, Knowledge-based recommender systems, in: Recommender systems, Springer, 2016, pp. 167–197.
- [28] B. P. Knijnenburg, M. C. Willemsen, Understanding the effect of adaptive preference elicitation methods on user satisfaction of a recommender system, in: Proceedings of the third ACM conference on Recommender systems, 2009, pp. 381–384.
- [29] S. J. Gould, Health consciousness and health behavior: the application of a new health consciousness scale, *American Journal of Preventive Medicine* 6 (1990) 228–237.
- [30] A. Gámbaro, A. C. Ellis, V. Prieto, Influence of subjective knowledge, objective knowledge and health consciousness on olive oil consumption—a case study, *Food and Nutrition* 4 (2013) 445–453.
- [31] C. Trattner, D. Elswiler, An evaluation of recommendation algorithms for online recipe portals, in: Proceedings of the 4th International Workshop on Health Recommender Systems, CEUR, Aachen, DE, 2019.
- [32] A. El Majjodi, A. D. Starke, C. Trattner, Nudging towards health? examining the merits of nutrition labels and personalization in a recipe recommender system, in: Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, 2022, pp. 48–56.
- [33] S. Funk, Netflix update: Try this at home (december 2006), 1999. URL: <http://sifter.org/simon/journal/20061211.html>.
- [34] W. H. Organization, Diet, nutrition, and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation, volume 916, World Health Organization, 2003.
- [35] H. Canada., The development and use of a surveillance tool: The classification of foods in the canadian nutrient file according to eating well with canada’s food guide, 2014. URL: http://publications.gc.ca/collections/collection_2014/sc-hc/H164-158-2-2014-eng.pdf.
- [36] A. D. Starke, M. C. Willemsen, C. Trattner, Nudging healthy choices in food search through visual attractiveness, *Frontiers in Artificial Intelligence* 4 (2021) 20.
- [37] Department of Health and Social Care UK, Front of Pack nutrition labelling guidance, 2016. URL: <https://www.gov.uk/government/publications/front-of-pack-nutrition-labelling-guidance>.
- [38] K. Anastasiou, M. Miller, K. Dickinson, The relationship between food label use and dietary intake in adults: A systematic review, *Appetite* 138 (2019) 280–291.
- [39] M. L. Neuhouser, A. R. Kristal, R. E. Patterson, Use of food nutrition labels is associated with lower fat intake, *Journal of the American dietetic Association* 99 (1999) 45–53.
- [40] M. Cecchini, L. Warin, Impact of food labelling systems on food choices and eating behaviours: a systematic review and meta-analysis of randomized studies, *Obesity reviews* 17 (2016) 201–210.
- [41] K. G. Grunert, J. M. Wills, L. Fernández-Celemín, Nutrition knowledge, and use and understanding of nutrition information on food labels among consumers in the uk, *Appetite* 55 (2010) 177–189.
- [42] K. L. Hawley, C. A. Roberto, M. A. Bragg, P. J. Liu, M. B. Schwartz, K. D. Brownell, The science on front-of-package food labels, *Public health nutrition* 16 (2013) 430–439.

- [43] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, C. Newell, Explaining the user experience of recommender systems, *User modeling and user-adapted interaction* 22 (2012) 441–504.
- [44] B. P. Knijnenburg, M. C. Willemsen, Evaluating recommender systems with user experiments, in: *Recommender systems handbook*, Springer, 2015, pp. 309–352.
- [45] A. D. Starke, C. Trattner, Promoting healthy food choices online: A case for multi-list recommender systems, in: *Proceedings of the ACM IUI 2021 Workshops*, CEUR, Aachen, DE, 2021.
- [46] M. C. Willemsen, M. P. Graus, B. P. Knijnenburg, Understanding the role of latent feature diversification on choice difficulty and satisfaction, *User Modeling and User-Adapted Interaction* 26 (2016) 347–389.
- [47] R. Mai, S. Hoffmann, How to combat the unhealthy= tasty intuition: The influencing role of health consciousness, *Journal of Public Policy & Marketing* 34 (2015) 63–83.
- [48] M. Jesse, D. Jannach, Digital nudging with recommender systems: Survey and future directions, *Computers in Human Behavior Reports* 3 (2021) 100052.
- [49] M. Jesse, D. Jannach, B. Gula, Digital nudging for online food choices, *Frontiers in Psychology* 12 (2021).
- [50] L. N. van der Laan, O. Orcholska, Effects of digital just-in-time nudges on healthy food choice—a field experiment, *Food Quality and Preference* 98 (2022).
- [51] A. Felfernig, R. Burke, Constraint-based recommender systems: technologies and research issues, in: *Proceedings of the 10th international conference on Electronic commerce*, 2008, pp. 1–10.
- [52] L. R. Flynn, R. E. Goldsmith, A short, reliable measure of subjective knowledge, *Journal of business research* 46 (1999) 57–66.
- [53] Z. Pieniak, J. Aertsens, W. Verbeke, Subjective and objective knowledge as determinants of organic vegetables consumption, *Food quality and preference* 21 (2010) 581–588.
- [54] R. M. Baron, D. A. Kenny, The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations., *Journal of personality and social psychology* 51 (1986) 1173.
- [55] A. Pardo, M. Román, Reflections on the baron and kenny model of statistical mediation, *Anales de psicología* 29 (2013) 614–623.
- [56] W.-Y. Chao, 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*, Springer, 2020, pp. 221–234.
- [57] Y. Liang, M. C. Willemsen, A longitudinal study – exploring the effect of nudging on users’ genre exploration behavior and listening preference, in: *Sixteenth ACM Conference on Recommender Systems*, ACM, New York, NY, USA, 2022.