Nudging Towards Health? Examining the Merits of Nutrition Labels and Personalization in a Recipe Recommender System

Ayoub El Majjodi
MediaFutures, University of Bergen
Bergen, Vestland, Norway
ayoub.majjodi@uib.no

Alain Starke
Marketing and Consumer Behaviour Group, Wageningen University & Research
Wageningen, Netherlands
MediaFutures, University of Bergen
Bergen, Vestland, Norway
alain.starke@wur.nl

Christoph Trattner
MediaFutures, University of Bergen
Bergen, Vestland, Norway
christoph.trattner@uib.no

ABSTRACT

Food recommender systems show personalized recipes to users based on content liked previously. Despite their potential, often recommended (popular) recipes in previous studies have turned out to be unhealthy, negatively contributing to prevalent obesity problems worldwide. Changing how foods are presented through digital nudges might help, but these are usually examined in non-personalized contexts, such as a brick-and-mortar supermarket. This study seeks to support healthy food choices in a personalized interface by adding front-of-package nutrition labels to recipes in a food recommender system. After performing an offline evaluation, we conducted an online study (N = 600) with six different recommender interfaces, based on a 2 (non-personalized vs. personalized recipe advice) x 3 (No Label, Multiple Traffic Light, Nutri-Score) between-subjects design. We found that recipe choices made in the non-personalized scenario were healthier, while the use of nutrition labels (our digital nudge) reduced choice difficulty when the content was personalized.

CCS CONCEPTS

• **Applied computing** → **Life and medical sciences**; • **Information systems** → *Personalization.*

KEYWORDS

Personalization, Health, Food recommendations, Digital nudges, Nutrition labels

ACM Reference Format:
Ayoub El Majjodi, Alain Starke, and Christoph Trattner. 2022. 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 (UMAP '22), July 4–7, 2022, Barcelona, Spain.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3503252.3531312

UMAP '22, July 4–7, 2022, Barcelona, Spain

ACM ISBN 978-1-4503-9207-5/22/07.
https://doi.org/10.1145/3503252.3531312

1 INTRODUCTION

The popularity of recipe websites has increased in the past years, particularly due to the covid-19 pandemic [42]. At the same time, this poses a serious decision-making challenge due to the abundance of food-related content on recipe websites, such as food images, categories, and cooking videos. Recommender systems can help users to filter that information, narrowing down the options to choose from based on user preferences or needs to present the most relevant content [21].

Research on food recommender systems has shown how to facilitate people's food decision-making [24, 57]. Based on bookmarks and ratings given to recipes by users, such recommenders retrieve recipes that contain, for example, the same ingredients as recipes liked previously. Because such recommender tend to push popular content, their success backfires in the sense that the often recommended, popular content tends to be unhealthy [55], thereby negatively contributing to societal health problems, such as obesity and diabetes.

Industrial practitioners and researchers have suggested solutions to mitigate the unhealthiness of such food recommenders. Among them, health-aware recommender systems examine how health-related outcomes could be modelled [16, 28]. Moreover, food recommender approaches that do not optimize for user preferences but for nutritional needs have also emerged [2, 4, 43]. However, recommender approaches that forgo on a user's past food preferences tend to lead to lower levels of user satisfaction [36], as a health or nutrition-focused approach is at odds with the propensity to like 'common', popular foods [57].

What content is suggested (i.e., based on what user model, algorithm) is only one aspect of a recommender system. The interface, specifically *how* the content is presented, is arguably an opportunity to also steer user choices towards healthier options [35]. In this sense, nudging has shown to be an effective technique to affect user choices and to lead to behavioral change in the food domain [6, 51]. Food-related nudges have been offered to consumers in various offline contexts, such as supermarkets and cafeteria [51], in an attempt to predictably affect user choices without mitigating the freedom of choice. These nudges affect 'daily', rather unconscious decisions that rely on heuristic cognitive processes [26]. Indeed, health-related nudges could serve as a mental shortcut to users who do not wish to put effort and time into their food choices [6, 46]. A meta-review of more than 60 studies on nudging interventions for healthy food choices shows that 80% of nudging interventions (e.g., through product placement at eye sight, use of defaults, priming) leads to a 15% average increase in healthy nutritional choices [32].

The use of nudges in online contexts, often referred to as 'digital nudges', has only emerged in the past years [23]. Although digital nudges have already been applied to internet-sourced recipes [3, 49], their effectiveness has only been studied in non-personalized contexts, outside the recommender context. We argue that the effectiveness of nudges observed in one-size-fits-all contexts, such as in a brick-and-mortar supermarket, may not hold up once the content already fits a user's preferences. The effectiveness of nudges should be regarded as a means to bridge the attitude-behavior gap [27], which usually does not apply to contexts where the presented content already fits the user's attitude (i.e., as a proxy for her preferences).

Moreover, a nudging intervention is not as 'invasive' as changing the content in a recommender system [23, 25]. By adding a health-based nudge to a constant set of recommended item, it could be possible to steer user preferences towards healthier options without reducing the user's level of satisfaction, which is expected to occur when recommended on nutritional content only [57].

In a first scientific attempt to bridge the gap between healthiness and what people like (i.e., user satisfaction), we introduce digital nudges to a recipe recommender system. To emphasize the health content of recipes, we introduce a cognitively oriented, informational nudge [6, 49] in the form of nutrition labels that are used on the front of food packaging (e.g., Nutri-score). Cognitively oriented nudges mainly motivate people to make better-informed decisions based on what they know effortlessly [6], by making specific information more salient [60]. In our case, this may particularly help people who lack nutritional knowledge regarding the foods or recipes they are considering to choose [61]. The choice to do so in this paper is motivated by making healthy foods 'stand out', as many food decisions are made without much cognitive effort (cf. [26]), using nutritional labels such as "Multiple Traffic Light" and "Nutri-score." [18].

We are among the first to use digital nudges in the context of food [24, 48, 60], as well as among the first to apply such front-of-package labels to recipes. In an online user study, we test the effectiveness of two different nutrition label across both non-personalized and personalized recommender interfaces. In addition, we examine whether this also depends on whether a user is interested in cooking healthy recipes, by inquiring on self-reported dietary goals. We address the following research questions:

- **RQ1:** To what extent do nutrition labels steer users to healthy recipe choices across personalized and non-personalized food recommender systems?
- **RQ2:** To what extent do personalization and nutrition labels affect user choice satisfaction and difficulty?
- **RQ3:** To what extent do user-based and evaluative factors predict the healthiness of a chosen recipe?

This paper is structured as follows. Section 2 presents the related work to our research, while our methodology is presented in Section 3, where we report the results of the offline recommender evaluation of our recipe dataset, as well as the research design of our online evaluation. Section 4 presents the results of our statistical analyses, of which the implications are further discussed in Section 5.

2 RELATED WORK

Digital technology can play an important role in supporting healthier food choices. However, the current approaches may be biased towards short-term user preferences [12, 29], rather longer-term goals [44]. We discuss how digital nudges, specifically nutrition labels, can support healthier food choices across personalized and non-personalized recommender interfaces.

2.1 Recommender Systems

Recommender systems retrieve and present content to users based on what they liked in current or previous sessions [22]. Whereas much work has been conducted in leisure and e-commerce domains, such as movies and books [40], the number of studies performed on food recommender systems have only increased in the past decade [57]. To make sense of contemporary food recommender systems, it is argued that there are three dominant types of approaches, in terms of what type of data or goals are used for personalization [36, 46].

The most traditional approach for food recommenders is to optimize their algorithms based on a user's eating preferences only [14]. This could come in the form of ratings and bookmarks on recipe websites [14], or through past purchases in an online supermarket environment [64]. For the recipe domain, most models assume that users like to receive recipe suggestions containing ingredients that they liked in the past [14, 20], or recipes from the same category [45, 58], typically exploiting Collaborative Filtering and Content-Based methods [57].

The two other types of food recommenders either only focus on the nutritional needs of the user [36], or aim to balance user preferences and nutritional needs [4, 46, 52, 57]. This can be incorporated in the form of constraints for specific nutrients in recipe retrieval [37, 53], or by suggesting foods to eat or buy based on missing nutrient or a user's health status [33, 59]. Agapito et al. [2] present a knowledge-based nutritional recommender system based on the user's health condition, using a profiler to process user information and matching that to a database of nutritional advice. Whereas nutrition-based recommenders can lead to comparatively lower levels of user satisfaction [57], other approaches apply a hybrid recommender approach. To balance both health and user preferences, a few approaches have adopted a hybrid approach in which similar recipes are retrieved and re-ranked based on a specific health-related feature [45, 55]. Beyond the food domain, health has also been the focal point of investigation [43], such as to promote physical activity or to suggest medical adherence behaviors.

2.2 Digital Nudges

Most food recommender studies do not investigate beyond changes in the recommended content [23, 44, 49]. Nudges can support users to make healthier food choices [51], for example by making them more aware of a recipe's nutritional content [3], without changing the presented content. Although food-related nudges have been successfully applied to offline contexts [6], such as by re-arranging a supermarket shelf to display healthy products at eye level, much less is known about the effectiveness in digital contexts [49].

An important difference between, say, a brick-and-mortar supermarket and a recipe website with a food recommender system is the level of personalization. Although nudges are effective in a physical supermarket [6], it remains an open question whether they are effective if the context is already personalized towards what a user likes? And, as a problem that is specific to this paper, would they still support healthier recipe choices amidst a personalized list of recipe recommendations?

Different types of nudges could be used to address these questions. Cadario et al. [6] discern between three types of healthy eating nudges: cognitively oriented nudges (e.g., through informational visibility or cues), affectively oriented nudges (e.g., through attractiveness food images [49]), and behaviorally oriented nudges (e.g., re-ranking lists of recommendations on health [3, 49]). The focal point of this paper is the use of informational nudges, as such could also be easily applied beyond the food domain. For example, emphasizing specific information on an e-commerce website might also 'nudge' users towards different purchases (cf. [23]). Moreover, behaviorally oriented digital nudges are less interesting to examine in this context, as some are commonly applied in recommender systems: the most relevant items are typically presented first [40].

2.3 Nutrition Labels

The health-based cognitively oriented or informational nudge examined in this paper is the addition of a nutrition label. Our work is based on Front-of-Package (FoP) labels [11, 61], found on individual products in supermarkets. MRI studies have revealed that the addition of a food label that either emphasizes the healthiness (e.g., high in calories, low in fat) or taste (e.g., sweet and juicy) of a food item, leads to varying brain activity [17], which could thus facilitate a shift towards healthier food choices.

Recently, more research has emphasized the importance of nutritional food labels to support people in meeting dietary intake levels [50]. Several guidelines have been found in the literature for designing food labels, such as capturing consumers' attention, as well the ease with which consumers can process, evaluate, and influence the decision-making [19, 38]. Accordingly, in several studies, Multiple Traffic Light (MTL) and Nutri-score nutrition labels have been found to lead to an increase in healthy food choices, compared to other types of food labels [10, 35, 41].

2.4 Contribution

Our work examines to what extent we support users in making food decisions *online*, while not mitigating their experience with using a personalized recommender system. The reviewed related work shows that we are among the first to combine personalization and nudging (cf. [47]), particularly in the food domain. In doing so, we propose a novel application of behavioral economics strategies within a recommender system, with the following contributions:

- Applying nutrition labels to recipes to examine whether they can support healthy food choices.
- Comparing the effectiveness of nudges across personalized and randomized advice interfaces in the recipe retrieval domain, combining content based on user preferences with context based on health needs.

3 METHODOLOGY

The following sections describe the proposed methodology for our offline and online evaluations. We first describe the dataset used, after which we determine which algorithm attains the highest accuracy level. The setup of our recommender interface, the followed procedure and research design, and the used measures are explained thereafter.

3.1 Dataset

We consulted a recipe database from the website Allrecipes.com, which was used in previous recommender studies [45, 58]. It initially contained over 58263 recipes, which were arranged into several food categories. For our studies, we narrowed down the dataset to four food categories (cf. Table 1), from which we randomly sampled a dataset of 991 recipes. The dataset included basic and nutritional recipe metadata: *URL image, number of calories, servings and serving size, and saturated fat, sodium, protein, fat, and salt.* The mean rating given to the recipes was rather high: 4.45 on a 5-point scale (*SD*=0.04).

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

3.2 Offline Evaluation

To determine which recommender algorithm could best predict user preferences, we performed offline evaluation on our dataset. In doing so, we focused on the highest level of accuracy based on the prediction error, through the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE). The data was evaluated using five-fold cross-validation, as was common in other recommender studies [14, 56].

Table 2 reports the results. It was apparent that Singular Value Decomposition (SVD) [15] outperformed other the other algorithms that were included (e.g., SVD++, NMF, KNNWithMeans), in terms of the prediction error measured. Hence, SVD was integrated into our recommender interface for our online evaluation, to match the presented recipes to elicited user preferences. This would be compared against a random recommendation scenario, based on a random generator function described in [34].

3.3 System Design and Procedure

We developed an interactive recommender system that generated recipe recommendations. All users were asked to fill out a

Table 2: Offline evaluation: comparison of different recommender algorithms based on the prediction errors.

Algorithm	RMSE	MAE
BaseLine predictor	0.72	0.55
Co-clustering	0.71	0.53
KNNBaseline	0.33	0.20
KNNWithMeans	0.33	0.21
NMF	0.62	0.49
SlopeOne	0.52	0.38
SVD	0.18	0.12
SVD++	0.38	0.28

questionnaire on their basic demographics (i.e., age, gender, level of education), as well as their self-reported food-related behaviors, such as their *level of cooking experience*, the healthiness of their *eating habits*, any specific *eating goals* (e.g., eating less sugar), and *dietary restrictions*. Subsequently, they were asked to select one out of four preferred food categories, from which they would receive recommendations.

To elicit user preferences, all users were asked to provide preference ratings to a list of ten recipes from the preferred food category. Half of the presented recipes were designated as healthy, based on an FSA health metric (cf. Subsection 3.6.1), while the other half were designated as unhealthy. Afterwards, all users were presented a list of ten recipes, which was either personalized on the given ratings or not (cf. Subsection 3.4), and again discerned between five healthy and five unhealthy recipes. Each recipe depicted its calories, the number of servings, the serving size in grams, the title of the recipe, and a photo; see Figure 1. Depending on the research design, a nutrition label (i.e., Nutri-Score or Multiple Traffic Light) was shown or not.

Each user was asked to choose one recipe they would like to cook at home. This was followed by a short questionnaire to evaluate the user experience regarding choice satisfaction and choice difficulty.

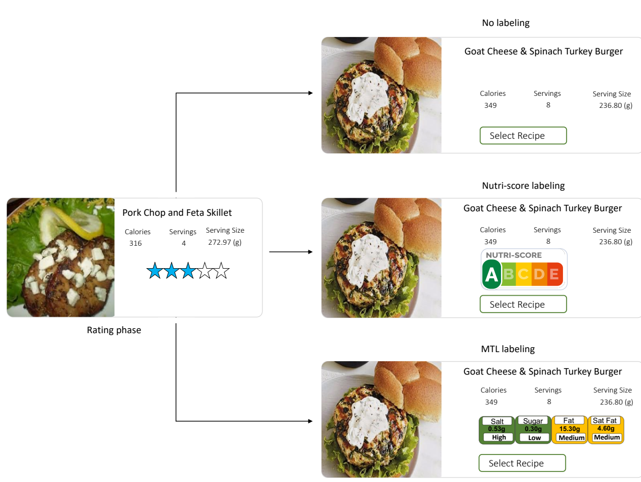

Figure 1: Examples of how individual recipes were presented to the active user across different labeling conditions.

3.4 Research Design

The recommender interface and the presented recipes were subject to a 2 (recommendations: personalized vs. non-personalized) x 3 (Labeling systems: no label vs. MTL vs. Nutri-score) between-subject design. In the personalized scenario, we used the given ratings to generate a list of ten personalized recipes using an SVD recommendation method. In contrast, the non-personalized scenario generated a list of random recipes from the preferred food category. Within each recommendation scenario, the baseline group was presented ten recipes without any labeling annotation, while the other two treatment groups interacted with recipes that were annotated with either MTL or Nutri-score labels. This variation in label annotation is also depicted in Figure 1. Accordingly, the participant was randomly assigned to any of the six conditions.

3.5 Study Participants

We recruited 600 Amazon MTurk workers to participate in our online study. The recruitment was based on a high level of hits (> 500 hits) and each participant was compensated with 1 USD for the task that approximately required 10min. Overall, participants (42% female) in this experiment were on average 39.53 years old, and had almost all attained their high school diploma.

3.6 Measures

3.6.1 Recipe healthiness. The healthiness of recipes could be assessed using various metrics (e.g., WHO, HCTS [7, 39]). In our study, we adopted the most commonly validated measure for food healthiness, the FSA score, which was issued by the British Food Standards Agency [9].

The FSA score was composed of four different nutrients: fat, saturates, sugar, and sodium. For each nutrient, it discerned between low, medium, or high content within a recipe. One point is assigned for each level (low, medium, high) per nutrient, leading to a scored scale that ran from 4 (healthiest) to 12 (least healthy). For example, fat content was designated as low if it fell below 3g per 100g served, while a medium range for saturated fat fell between 3g/100g and 17g/100g served. High recipe content not only considered the per 100g content, but also the total weight in g per serving. All computational details about the FSA score were reported in Starke et al. [49]. Table 3 presents the FSA score distribution of recipes found in our dataset.

We discerned between healthy and unhealthy recipes based on the FSA score. Recipes were considered healthy if their FSA score fell between 4 to 8, while 'less healthy' recipes had an FSA score between 9 and 12.

Table 3: The FSA scores for recipes used in our study.

FSA score	4	5	6	7	8	9	10	11	12
Number of recipes	4	43	102	150	158	199	295	24	16

3.6.2 Nutrition Labels. The FSA score formed the foundation of the MTL labeling system [9]. Accordingly, each recipe nutrient was represented by a color that indicated whether the amount found in the recipe was considered low (green), medium (amber), or high (red). On the other hand, the Nutri-score labeling system [8] signalled recipe healthiness through a color-coded summary evaluation, ranging from dark green -A- (healthiest) to dark red -E- (least healthy), which was based on energy, nutrients, and ingredients. Table 4 presents the Nutri-score of the recipes found in our study.

Table 4: The Nutri-scores for recipes used in our study. (A): highest nutritional quality, (E): lowest nutritional quality.

Nutri score	A	B	C	D	E
Number of recipes	234	278	898	171	10

3.6.3 User Evaluation. For a user's evaluation, we assessed their experienced choice difficulty and choice satisfaction. Each metric was measured using pre-validated questionnaire items [49, 62], which are outlined in Table 5. All responses to our propositions were recorded on 5-point Likert scales. Two questionnaire items had to removed, for they negatively affected the respective values of Cronbach's Alpha, making it uncertain whether they measured the same construct. The remaining four items resulted in acceptable to good levels of reliability.

3.6.4 User Characteristics. As mentioned in Subsection 3.3, we also inquired on a number of user characteristics and goals, which were used to address RQ3. Besides basic demographics that were added to the model as continuous variables (i.e., gender, level of education, age), we also inquired on a user's self-reported level of cooking experience and healthiness of eating habits (both on 5-point scales). Moreover, we asked users to disclose any eating goals they would have, such as eating less sugar or more protein. For our analysis, we included the number of self-reported healthy eating goals a continuous variable in our model (M=1.79, SD=1.53).

4 RESULTS

We analyzed the healthiness of chosen recipes across different recommendation approaches and label annotations (RQ1), as well as the choice satisfaction and choice difficulty reported by our system users using two-way ANOVAs (RQ2). Finally, we predicted the healthiness of chosen recipes using different types of factors in a regression model (RQ3).

4.1 RQ1: Healthiness of Chosen Recipes

We examined the FSA score of chosen recipes across all conditions. Figure 2 depicts the choice distribution in terms of whether recipes were designated as healthy (FSA < 9) or unhealthy (FSA > 8). We found that 65% of chosen recipes were healthy in the non-personalized scenario (random recommender algorithm), while 60% of recipes chosen in the personalized scenario were unhealthy.

Whether any of the observed differences were statistically significant was examined using a two way between-subjects ANOVA. We predicted the FSA score of the chosen recipe using the employed recommendation approach (control: *Random algorithm*, treatment:*SVD*) and the employed front-of-pack nutrition labels (control: no-label, treatments: Nutri-score, MTL). Table 6 indicates that whether recommendations were personalized significantly affected the healthiness of chosen recipes: $F(1{,}594) = 12.91$, $p < 0.01$.

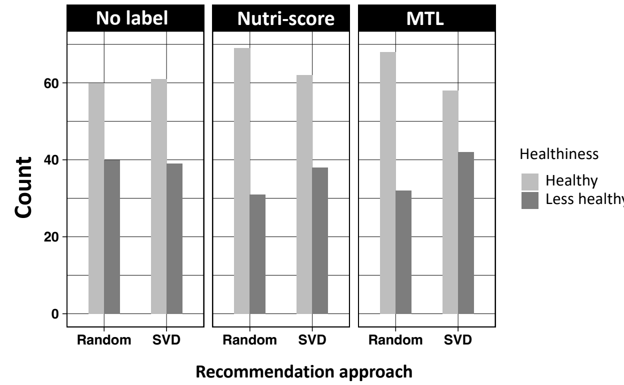

Figure 2: Distribution of healthy and less healthy recipe choices across our different recommender and labelling conditions.

Concerning the nutrition labels, Table 6 indicates there were no significant differences between any of the two labelled conditions and the no-label baseline (both p-values > 0.05). Moreover, we neither observed any interaction effects between the recommendation approach and the used labels (both p-values > 0.05). Figure 2 suggests that the use of nutrition labels (both Nutri-score and MTL) led to slightly more healthy recipe choices in the non-personalized condition (compared to No Label), while the number of healthy choices made in the personalized condition was actually lower for both of the labelled conditions (again compared to No Label). However, as indicated by our ANOVA results, this interaction effect was not significant.

We checked for additional differences using a post-hoc Tukey HSD test. This confirmed that the mean FSA score of recipes chosen in the no-personalization approach *(M=7.87, SD=1.86)* was significantly lower than those chosen in the personalized condition (M=8.35, SD=1.54). The Tukey test did not reveal any additional differences. Taken together, these results suggested that a high level of personalization of recipes led users to make unhealthier recipe choices, while nutrition labels did not seem to mitigate this effect and had only a small, non-significant effect in the non-personalized condition.

4.2 RQ2: User Evaluation

We examined the user experience across all recommender system conditions. We used two different two-way ANOVAs to predict differences in choice satisfaction and choice difficulty levels.

The results for choice difficulty are described in Table 7. We found a main effect of personalized on choice difficulty, indicating that personalized interfaces led to a lower perceived choice difficulty ($M = 3.38$, $SD = 0.15$) compared to our non-personalized recommenders ($M = 3.41$, $SD = 0.05$): $F(1{,}594) = 10.04$, $p = 0.002$. Although the addition of nutrition labels did not significantly affect choice difficulty as a main effect (both MTL and Nutri-score: p>0.05), we did observe two interaction effects with whether the content was personalized or not. The combined presence of both

Table 5: Questionnaire items for choice satisfaction and choice difficulty. Items in gray were omitted from analysis.

Measure	Item	Mean	Alpha
Choice Satisfaction	I would recommend the chosen recipe to others.	4.02	0.60
	I think I would enjoy the chosen recipe.	4.27	
	My chosen recipe could become one of my favorites.	4.06	
Choice Difficulty	I changed my mind several times before making a decision.	2.97	0.78
	Making a choice was overwhelming.	2.96	
	It was easy to make this choice.	3.83	

Table 6: Results of a Two-Way ANOVA, predicting the healthiness of chosen recipes across different recommendation and labeling conditions. Note that label predictors were added separately, as there was no clear hierarchy between the Nutri-Score and MTL in terms of the expected effectiveness. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Factor (FSA score)	**df**	***F***
Model	5	3.19**
Nutri-score	1	1.39
MTL	1	0.55
Recommendation approach	1	12.91**
Recommendation approach * Nutri-score	1	0.74
Recommendation approach * MTL	1	0.86

SVD and Nutri-score (p<0.001), as well as SVD and MTL (p<0.001) significantly affected choice difficulty.

To understand this interaction effect, please refer to Figure 3. It depicts that the effect of personalization on choice difficulty depended on the presence of nutrition labels. For the No Label condition, it seemed that personalization *increased* the perceived choice difficulty. In contrast, for both the Nutri-score and MTL conditions, personalization *decreased* the perceived choice difficulty. It seemed that the merits of adding nutrition labels depended on whether the content was personalized.

Table 7: Results of a two-way ANOVA that predicted choice difficulty across recommendation and labelling conditions. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Factor (Choice difficulty)	**df**	***F***
Model	5	4.74**
MTL	1	0.00
Nutri-score	1	1.64
Recommendation approach	1	10.04**
Recommendation approach * MTL	1	19.01***
Recommendation approach * Nutri-score	1	11.82***

For choice satisfaction, however, we could not reliably infer a model. We found that the Two-Way ANOVA model with personalization and label factors was not significantly different from an empty baseline model: $F(5,595) = 1.59$, $p > 0.05$. This indicated that

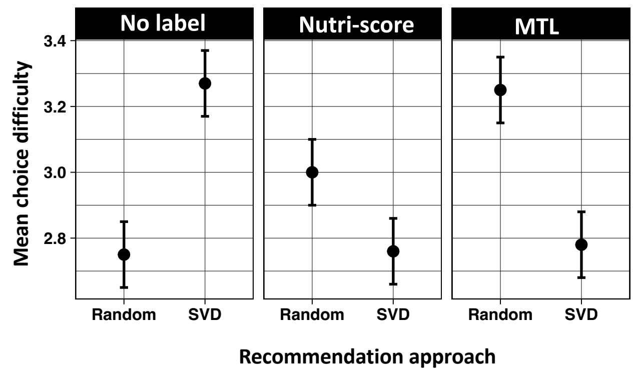

Figure 3: Means and standard errors of choice difficulty levels, reported by users across conditions.

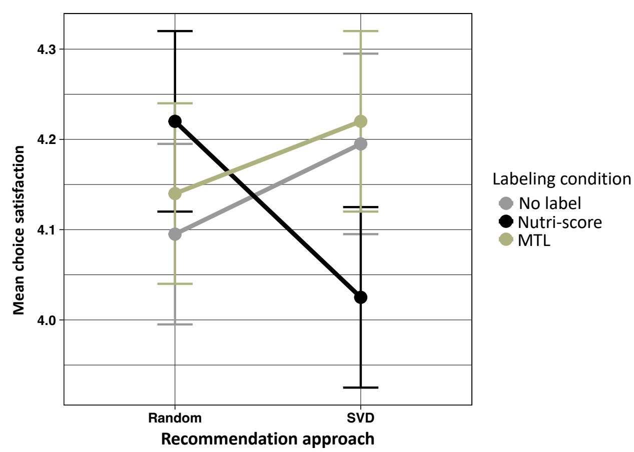

Figure 4: Mean levels of choice satisfaction across all personalization (random vs SVD) and labeling conditions.

we could *not* reliably interpret the model's parameters, and suggested that there were likely no relevant differences in the model. The lack of differences is also suggested by Figure 4, depicting only small differences between personalization approaches across each label condition.

Table 8: Results of the linear regression model that predicted the FSA score (i.e., inverse healthiness) of recipes chosen by users, based on factors from the research design, user characteristics, and user perception. $^{***}p$ < 0.001, $^{**}p$ < 0.01, $^{*}p$ < 0.05.

Factor (FSA score)	β	$S.E$
Personalization & Labels		
Nutri-score	-.23	.16
MTL	-.14	.16
Personalization	.47***	.13
User Characteristics		
Age	-.054	.063
Cooking experience	-.012	.085
Eating goals	-.29**	.12
Gender (Male)	-.018	.13
Healthy eating habits	-.080	.091
Level of education	-.049	.11
User Perception & Interactions		
Choice difficulty	-.28**	.09
Choice difficulty * Goals	.079*	.036
Intercept	9.54***	.49
R^2	0.053	

4.3 RQ3: Predicting Recipe Healthiness

Finally, we investigated the predictability of the chosen FSA score (i.e., inverse healthiness), based on different types of factors using a linear regression. In our model, we differentiated between factors from the research design (i.e., personalization, labelling systems), user characteristics (i.e., demographics, self-reported eating goals and habits), and system perception (i.e., choice difficulty). Table 8 outlines the results[1], which again confirms that personalization led to unhealthier recipe choices (p<0.001), while this did not apply to the different labelling systems.

With regard to user characteristics and perception, Table 8 points out two additional significant predictors and an interaction effect. First, whereas cooking experience, habits, and demographics were not significant predictors, it did indicate that users with more healthy eating goals chose recipes with on average lower FSA scores (p<0.05). This suggested that the system could support users with such goals to find appropriate recipes. Second, users who perceived the decision-making process as difficult (i.e., a higher choice difficulty) also made healthier choices: $\beta = -.27$, $p < 0.01$. At the same time, we also observed a positive interaction effect between the number of healthy eating goals and the reported choice difficulty (p<0.05). This could be understood by considering both predictors as being either high (positive) or low (negative): users with many healthy eating goals and who perceived the decision to be difficult made unhealthier choices, as well as users with no healthy eating goals and little choice difficulty. In contrast, users with many healthy goals seemed to particularly choose healthy recipes if they perceived little choice difficulty.

[1]We also explored other interaction effects, but found no relevant ones. Note that we excluded choice satisfaction from this model, as this was an aspect that measured the user's experienced satisfaction *after* the recipe was chosen. Therefore, from a causal point of view, it would not make sense to use it to predict the FSA score of the chosen recipe.

5 DISCUSSION

The effectiveness of digital nudges within recommender systems has received attention to date [23]. In the food domain, however, several studies stressed the potential of nudging strategies to advance public health in offline contexts [6]. In an attempt to bridge brick-and-mortar supermarkets and recipe websites, we have filled this research gap by applying an informational, cognitively oriented nudge in a recommender system through nutrition labels.

Our results particularly contribute to the overall understanding of the effectiveness of personalization and nudges in the food domain. In line with the literature that describes how popular recipes in food recommender systems lead to unhealthy outcomes [55, 56], we have found that personalized rather than random recipe recommendations lead to a *decrease* in the healthiness of chosen recipes. This confirms that the commonly used, preference-based and popularity-driven approach in recommender research [13], leads to unhealthy outcomes in recipe recommendation.

Arguably surprisingly, we have found that this effect is not moderated by the use of our informational nudges, the two front-of-package nutrition labels. The latter can be contextualized in terms of previous food recommender system research. Recommender approaches are typically assessed in terms of their focus on either user preferences, nutritional needs, or a trade-off between both [36, 46]. In the current study, it seems that unhealthy user preferences have prevailed over any health-related needs. This is arguably exacerbated by the limitations of the dataset sample used for our offline and online evaluations, which contained rather popular recipes (i.e., a mean preference rating of 4.5 out of 5), even more than so than related datasets used in previous studies [55, 56]. Therefore, in future studies, we opt to use a more diverse dataset in terms of popular and non-popular recipes, to examine this problem using a more representative sample of food-related products. Alternatively, even though content-based recommender approaches may not yield higher accuracy levels than collaborative approaches [14, 57], such as the SVD employed in the current study, content-based recommendation might be able to mitigate the popularity bias in recipe recommenders (cf. [1]).

This study has applied a single nudging technique to a personalized recommendation scenario. Although front-of-package labels, such as the Nutri-score and the Multiple Traffic Light (MTL) label, have increased healthy food purchases in brick-and-mortar supermarkets [6], their effectiveness for online recipes is less profound. Recent research has suggested that digital nudges might need to be combined to increase their effectiveness [23], although this applied mostly to a non-personalized food retrieval system.

The findings on choice difficulty are *also* important for studies beyond the food domain. Although choice difficulty seems to increase due to the use of personalization (i.e., using SVD over random recommendation), this effect is reversed by the introduction of nutrition labels. This suggests that whereas nutrition labels are not helpful in a random recommendation context, their merit is higher when the content is more alike, which is likely in a personalized scenario [5]. Our study has, thus, shown that although they may not overcome the popularity bias in recipe recommendation, they may facilitate better decision-making.

Contrary to recommender studies in other domains [30, 63], we have not observed a significant increase in choice satisfaction for the SVD recommender compared to the random approach. The descriptive results, as indicated by Figures 2 and 4, suggest that choice satisfaction may be related to the unhealthiness of chosen recipes. Or, in other words, that there may be a positive relation between the FSA score and choice satisfaction, which would further confirm that users appreciate popular recommender content. At the same time, this poses additional challenges for food recommender studies, which may need to sacrifice accuracy to facilitate healthy outcomes [46].

The results from RQ1-2 open up new research directions. For example, it is interesting to examine what are other nudging techniques, beyond nutrition labels and other informational nudges, are more effective in supporting healthy choices, when integrated with a personalization recommender. In line with the rationale of this study and previous studies [24, 49], such an approach should not come at the cost of evaluative outcomes, like choice satisfaction. Moreover, it might also be interesting to examine to what extent such nudging techniques can support changes in a user's longer-term eating habits and diet, for previous studies have suggested that both recommender systems and nudges could affect long term habits [31, 54].

For our third research question (RQ3), we predicted the healthiness of recipes (FSA score). The results suggest that the number of healthy eating goals that a user has, affects their recipe choices at the health level. We have found the effect is moderated by the perceived choice difficulty. This suggests that a non-confusing, unambiguous decision environment is more effective at facilitating the healthy eating goals of users. In terms of other user characteristics, such as level of education and cooking experience, we have observed no other significant predictors. We further find that personalization positively correlates with FSA score. This suggests that a high level of preference matching with the recommended recipes can lead to unhealthy choices.

Finally, we conclude our discussion section by proposing novel research questions, as a guide for future work. Some of these questions can also be applied to recommender domains beyond food:

- How can other types of nudging strategies (e.g., defaults, social norms) be integrated with (food) recommender approaches to facilitate better (i.e., for food: healthier) decision-making?
- How can user eating goals be integrated with both the suggested content and the decision context of a recommender system?
- To what extent do short-term food choices contribute to behavioral change in the long term?

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] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2017. Controlling popularity bias in learning-to-rank recommendation. In *Proceedings of the eleventh ACM conference on recommender systems*. ACM, New York, NY, USA, 42–46.

[2] Giuseppe Agapito, Barbara Calabrese, Pietro Hiram Guzzi, Mario Cannataro, Mariadelina Simeoni, Ilaria Caré, Theodora Lamprinoudi, Giorgio Fuiano, and Arturo Pujia. 2016. DIETOS: A recommender system for adaptive diet monitoring and personalized food suggestion. In *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 1–8.

[3] Arngeir Berge, Vegard Velle Sjøen, Alain D. Starke, and Christoph Trattner. 2021. Changing Salty Food Preferences with Visual and Textual Explanations in a Search Interface. In *HEALTHI'21: Joint Proceedings of ACM IUI 2021 Workshops*. CEUR.

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

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

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

[7] Health Canada. 2014. *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*. http://publications.gc.ca/collections/collection_2014/sc-hc/H164-158-2-2014-eng.pdf

[8] Julia Chantal, Serge Hercberg, World Health Organization, et al. 2017. Development of a new front-of-pack nutrition label in France: the five-colour Nutri-Score. *Public Health Panorama* 3, 04 (2017), 712–725.

[9] Department of Health and Social Care UK. 2016. *Front of Pack nutrition labelling guidance*. https://www.gov.uk/government/publications/front-of-pack-nutrition-labelling-guidance

[10] Pauline Ducrot, Chantal Julia, Caroline Méjean, Emmanuelle Kesse-Guyot, Mathilde Touvier, Léopold K Fezeu, Serge Hercberg, and Sandrine Péneau. 2016. Impact of different front-of-pack nutrition labels on consumer purchasing intentions: a randomized controlled trial. *American journal of preventive medicine* 50, 5 (2016), 627–636.

[11] Manon Egnell, Zenobia Talati, Serge Hercberg, Simone Pettigrew, and Chantal Julia. 2018. Objective understanding of front-of-package nutrition labels: an international comparative experimental study across 12 countries. *Nutrients* 10, 10 (2018), 1542.

[12] Michael D Ekstrand and Martijn C Willemsen. 2016. 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, 221–224.

[13] Mehdi Elahi, Danial Khosh Kholgh, Mohammad Sina Kiarostami, Sorush Saghari, Shiva Parsa Rad, and Marko Tkalčič. 2021. Investigating the impact of recommender systems on user-based and item-based popularity bias. *Information Processing & Management* 58, 5 (2021), 102655.

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

[15] Simon Funk. 1999. *Netflix update: Try this at home (December 2006)*. http://sifter.org/simon/journal/20061211.html

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

[17] Fabian Grabenhorst, Frank P Schulte, Stefan Maderwald, and Matthias Brand. 2013. Food labels promote healthy choices by a decision bias in the amygdala. *Neuroimage* 74 (2013), 152–163.

[18] Klaus G Grunert, Laura Fernández-Celemín, Josephine M Wills, Stefan Storcksdieck genannt Bonsmann, and Liliya Nureeva. 2010. Use and understanding of nutrition information on food labels in six European countries. *Journal of public health* 18, 3 (2010), 261–277.

[19] Klaus G Grunert and Josephine M Wills. 2007. A review of European research on consumer response to nutrition information on food labels. *Journal of public health* 15, 5 (2007), 385–399.

[20] Morgan Harvey, Bernd Ludwig, and David Elsweiler. 2013. You are what you eat: Learning user tastes for rating prediction. In *International symposium on string processing and information retrieval*. Springer, 153–164.

[21] Dietmar Jannach and Michael Jugovac. 2019. Measuring the business value of recommender systems. *ACM Transactions on Management Information Systems*

(TMIS) 10, 4 (2019), 1–23.

[22] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. 2011. An introduction to recommender systems. *New York: Cambridge* 10 (2011), 1941904.

[23] Mathias Jesse and Dietmar Jannach. 2021. Digital nudging with recommender systems: Survey and future directions. *Computers in Human Behavior Reports* 3 (2021), 100052.

[24] Mathias Jesse, Dietmar Jannach, and Bartosz Gula. 2021. Digital Nudging for Online Food Choices. *Frontiers in Psychology* 12 (2021).

[25] Eric J Johnson, Suzanne B Shu, Benedict GC Dellaert, Craig Fox, Daniel G Goldstein, Gerald Häubl, Richard P Larrick, John W Payne, Ellen Peters, David Schkade, et al. 2012. Beyond nudges: Tools of a choice architecture. *Marketing Letters* 23, 2 (2012), 487–504.

[26] Daniel Kahneman. 2011. *Thinking, fast and slow.* Macmillan.

[27] Florian G Kaiser, Oliver Arnold, and Siegmar Otto. 2014. Attitudes and defaults save lives and protect the environment jointly and compensatorily: Understanding the behavioral efficacy of nudges and other structural interventions. *Behavioral sciences* 4, 3 (2014), 202–212.

[28] Mansura A Khan, Ellen Rushe, Barry Smyth, and David Coyle. 2019. Personalized, health-aware recipe recommendation: an ensemble topic modeling based approach. In *HEALTHI'21: Joint Proceedings of ACM IUI 2021 Workshops*. CEUR, 4–10.

[29] Bart P Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. 2016. Recommender systems for self-actualization. In *Proceedings of the 10th acm conference on recommender systems*. ACM, New York, NY, USA, 11–14.

[30] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User modeling and user-adapted interaction* 22, 4 (2012), 441–504.

[31] Thomas C Leonard. 2008. Richard H. Thaler, Cass R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness.

[32] Christophe Leys, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology* 49, 4 (2013), 764–766.

[33] Jennifer Mankoff, Gary Hsieh, Ho Chak Hung, Sharon Lee, and Elizabeth Nitao. 2002. Using low-cost sensing to support nutritional awareness. In *International conference on ubiquitous computing*. Springer, 371–378.

[34] Makoto Matsumoto and Takuji Nishimura. 1998. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 8, 1 (1998), 3–30.

[35] Cliona Ni Mhurchu, Helen Eyles, Yannan Jiang, and Tony Blakely. 2018. Do nutrition labels influence healthier food choices? Analysis of label viewing behaviour and subsequent food purchases in a labelling intervention trial. *Appetite* 121 (2018), 360–365.

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

[37] Cataldo Musto, Christoph Trattner, Alain Starke, and Giovanni Semeraro. 2020. 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, 333–337.

[38] Pan American Health Organization. 2020/09/25. *Launching of Front-of-Package Labeling as a Policy Tool for the Prevention of Noncommunicable Diseases in the Americas*. PAHO. https://www.paho.org/en/documents/front-package-labeling-policy-tool-prevention-noncommunicable-diseases-americas

[39] World Health Organization. 2003. *Diet, nutrition, and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation*. Vol. 916. World Health Organization.

[40] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender systems: introduction and challenges. In *Recommender systems handbook*. Springer, 1–34.

[41] Osvaldo Santos, Violeta Alarcão, Rodrigo Feteira-Santos, João Fernandes, Ana Virgolino, Catarina Sena, Carlota Pacheco Vieira, Maria João Gregório, Paulo Nogueira, Pedro Graça, et al. 2020. Impact of different front-of-pack nutrition labels on online food choices. *Appetite* 154 (2020), 104795.

[42] Barthélemy Sarda, Corinne Delamaire, Anne-Juliette Serry, and Pauline Ducrot. 2022. Changes in home cooking and culinary practices among the French population during the COVID-19 lockdown. *Appetite* 168 (2022), 105743.

[43] Hanna Schäfer, Santiago Hors-Fraile, Raghav Pavan Karumur, André Calero Valdez, Alan Said, Helma Torkamaan, Tom Ulmer, and Christoph Trattner. 2017. Towards health (aware) recommender systems. In *Proceedings of the 2017 international conference on digital health*. ACM, New York, NY, USA, 157–161.

[44] Alain Starke. 2019. RecSys Challenges in achieving sustainable eating habits. In *4th International Workshop on Health Recommender Systems, HealthRecSys 2019*. CEUR, 29–30.

[45] Alain Starke, Edis Asotic, and Christoph Trattner. 2021. "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, 124–132.

[46] Alain Starke, Christoph Trattner, Hedda Bakken, Martin Johannessen, and Vegard Solberg. 2021. The cholesterol factor: Balancing accuracy and health in recipe recommendation through a nutrient-specific metric. In *Proceedings of the 1st Workshop on Multi-Objective Recommender Systems (MORS 2021)*. CEUR, 11 pages.

[47] Alain Starke, Martijn Willemsen, and Chris Snijders. 2021. Promoting Energy-Efficient Behavior by Depicting Social Norms in a Recommender Interface. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 11, 3-4 (2021), 1–32.

[48] Alain D Starke and Christoph Trattner. 2021. Promoting Healthy Food Choices Online: A Case for Multi-List Recommender Systems. In *Proceedings of the ACM IUI 2021 Workshops*. CEUR, Aachen, DE, 3 pages.

[49] Alain D Starke, Martijn C Willemsen, and Christoph Trattner. 2021. Nudging healthy choices in food search through visual attractiveness. *Frontiers in Artificial Intelligence* 4 (2021), 20.

[50] Norman J Temple and Joy Fraser. 2014. Food labels: a critical assessment. *Nutrition* 30, 3 (2014), 257–260.

[51] Richard H Thaler and Cass R Sunstein. 2009. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

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

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

[54] Thi Ngoc Trang Tran, Alexander Felfernig, Christoph Trattner, and Andreas Holzinger. 2021. Recommender systems in the healthcare domain: state-of-the-art and research issues. *Journal of Intelligent Information Systems* 57, 1 (2021), 171–201.

[55] Christoph Trattner and David Elsweiler. 2017. 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, 489–498.

[56] Christoph Trattner and David Elsweiler. 2019. An evaluation of recommendation algorithms for online recipe portals. In *Proceedings of the 4th International Workshop on Health Recommender Systems*. CEUR, Aachen, DE.

[57] Christoph Trattner and David Elsweiler. 2019. Food Recommendations. In *Collaborative recommendations: Algorithms, practical challenges and applications*. World Scientific, 653–685.

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

[59] Tsuguya Ueta, Masashi Iwakami, and Takayuki Ito. 2011. A recipe recommendation system based on automatic nutrition information extraction. In *International Conference on Knowledge Science, Engineering and Management*. Springer, 79–90.

[60] L Nynke van der Laan and Oliwia Orcholska. 2022. Effects of Digital Just-In-Time Nudges on Healthy Food Choice–a Field Experiment. *Food Quality and Preference* 98 (2022), 104535 pages.

[61] Ellen J Van Loo, Vincenzina Caputo, Rodolfo M Nayga Jr, and Wim Verbeke. 2014. Consumers' valuation of sustainability labels on meat. *Food Policy* 49 (2014), 137–150.

[62] Martijn C Willemsen, Mark P Graus, and Bart P Knijnenburg. 2016. Understanding the role of latent feature diversification on choice difficulty and satisfaction. *User Modeling and User-Adapted Interaction* 26, 4 (2016), 347–389.

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

[64] Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. A dynamic recurrent model for next basket recommendation. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 729–732.