Exploring the Effects of Natural Language Justifications in Food Recommender Systems

CATALDO MUSTO, University of Bari 'Aldo Moro', Italy
ALAIN D. STARKE, Wageningen University & Research, The Netherlands and University of Bergen, Norway
CHRISTOPH TRATTNER, University of Bergen, Norway
AMON RAPP, University of Torino, Italy
GIOVANNI SEMERARO, University of Bari 'Aldo Moro', Italy

Users of *food recommender systems* typically prefer popular recipes, which tend to be unhealthy. To encourage users to select healthier recommendations by making more informed food decisions, we introduce a methodology to generate and present a *natural language justification* that emphasizes the nutritional content, or health risks and benefits of recommended recipes. We designed a framework that takes a *user* and two *food recommendations* as input and produces an automatically generated natural language justification as output, which is based on the user's characteristics and the recipes' features. In doing so, we implemented and evaluated eight different *justification strategies* through two different *justification styles* (e.g., comparing each recipe's food features) in an online user study ($N = 503$). We compared user food choices for two personalized recommendation approaches, popularity-based vs our health-aware algorithm, and evaluated the impact of presenting natural language justifications. We showed that *comparative* justifications styles are effective in supporting choices for our healthy-aware recommendations, confirming the impact of our methodology on food choices.

CCS Concepts: • **Information systems → Recommender systems**; • **Computing methodologies →** *Natural language processing*.

Additional Key Words and Phrases: food recommender systems, natural language processing, explanation, decision making

ACM Reference Format:
Cataldo Musto, Alain D. Starke, Christoph Trattner, Amon Rapp, and Giovanni Semeraro. 2021. Exploring the Effects of Natural Language Justifications in Food Recommender Systems. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21), June 21–25, 2021, Utrecht, Netherlands.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3450613.3456827

1 INTRODUCTION

How do people choose what to eat? The answer is not straightforward, as research has shown that food choice motivations span from sensory appeal, convenience, and health, to ethical concerns and familiarity [23, 29, 38, 43]. Factors that affect food choices can be divided in food-related features (e.g., perceptual features and nutritional information) [35], individual differences (e.g., knowledge, skills, and anticipated consequences), and society-related features (e.g., norms and values) [9]. In this context, food recommender systems (RS) have emerged as an effective solution to *drive* and *support* people's food choices. Early technologies that generate meal recommendations to users date back to 1986 (e.g., CHEF [18]), while applications use ML techniques to automatically generate recipes that match user preferences [55].

In recent years, the idea of exploiting personalized recommendations to aid people to nourish themselves more healthily has spread [12]. This intuition is investigated by the research line regarding *health-aware* food recommender systems [41], which consider user information, such as dietary preferences and constraints (e.g., allergies) to generate a suitable meal plan. The main issue at hand is that most of the popular internet-sourced recipes used in recommendation approaches are unhealthy [53], and are, as a result, preferred by users. However, most RSs are still ill-equipped to effectively support a shift towards healthier (or more sustainable) eating habits [35, 45, 47].

In parallel recommender domains, the developments in *natural language explanations* and *justifications* strategies are promising [50]. Explanations can make the recommendation process more transparent, increasing users' trust and affecting their decision-making processes [36]. This paper fits this research theme, for we introduce a methodology to generate a *natural language justification* that supports recommendations generated by a *food recommender system*. Preliminary food recommender research shows that the interface context (i.e., *how* recommendations are presented) could affect user preferences [47, 52]. Our strategy aims to encourage people to make healthier food choices by providing them with a justification of the recommended recipe, emphasizing nutritional facts, risks, or benefits related to food consumption. Our conjecture is that justifications allow users to make better-informed and healthier food choices.

To this end, we present a framework inspired by knowledge-based Natural Language Generation [39] strategies. It takes a *user* and two *food recommendations* as input and produces an automatically generated natural language justification as output, which is based on the user's characteristics and the recipes' features. Moreover, general knowledge about health risks and benefits related to food consumption is considered to generate our justifications. Within the framework, we implement and evaluate eight different *justification strategies* through two different *justification styles*, based on the combination of different informative content and features. In particular, we generate *comparative* justifications of recommendations, which compare the main characteristics of two recipes into a single natural language sentence. For instance, such a justification could compare the fiber content content of two recipes. This taps into consumer research on the effectiveness of comparative evaluations of item attributes [3], compared to a separate representation of that information (i.e., a 'Single' justification).

The strength of the current work lies in its novelty. Generating *food* recommender systems with explanatory messages is a poorly investigated research topic, nor is there much empirical evidence on the support of healthier food choices. We evaluate our framework in a user study ($N = 503$), examining whether natural language justifications steer user food choices towards healthier recommendation. We posit the following research question:

[RQ]: *Do natural language justifications affect user choices for healthy recipe recommendations, compared to popular ones?* As we will show in the following, it emerged that users preferred healthier recipes over popularity-based recommendations, when comparative justifications are presented.

We summarize our contributions as follows: (i) We introduce a methodology to automatically generate a natural language justification to support personalized food recommendations; (ii) we design and (iii) evaluate several justification styles (i.e., None, Single, or Comparative styles) and strategies in a user study, where each justification leverages different user characteristics and recipe features. Section 2 presents an overview of related literature, while section 3 introduces our framework to automatically generate natural language justifications supporting food recommendations. Finally, we discuss the outcomes of our experimental session in Section 4, and sketch conclusions and future work in Section 5.

2 RELATED WORK

The idea of providing intelligent information systems with *explanation* facilities has been studied since the early 90s [24], and it was introduced in the area of recommender systems since 2000s [19]. It re-gained attention due to the recent

General Data Protection Regulations (GDPR), which prescribed to increase the transparency of underlying algorithms. This particularly applies to RSs, since explanation strategies have shown to positively affect both a user's acceptance of and trust in presented recommendations [10, 44]. The community's interest in the topic is shown across several studies, which each discuss the merits of explanations for recommender systems [16, 26].

We frame our current work by identifying *persuasiveness* (i.e., to promote healthier food choices) as the main goal of our justifications, which has not been investigated in other food RS research. This explanation aim is highlighted by Tintarev and Masthoff [50] and used in other domains to convince users to try or buy a recommended item. For example, [17] present a preliminary study of the persuasive power of explanations in a movie recommendation scenario.

With respect to the *information content*, which is exploited to generate justifications, we frame our approach as being at the intersection between *content-based* and *knowledge-based* methods (cf. [22]). Our methodology is based on the exploitation of user characteristics and food features, along with general knowledge on food consumption that is used to justify our health-aware recommendation by emphasizing health risks and benefits. This is related to a study where health risks are highlighted in a smoking cessation application [20], but, unfortunately, no evidence concerning the effectiveness of such information is provided in the article. Conversely, our work fills this knowledge gap, by evaluating the impact of justifications, including health risks and benefits, on user food choices.

Another hallmark of the current work lies in the development of a justification framework, designed specifically for the *food* domain. As discussed in [51], studies that evaluate the impact of explanations and justification in the food domain are scarce, even though they could encourage users to stick to better eating habits. A preliminary attempt to introduce explanation mechanisms in a food RS is presented by Leipold et al. in [27], where a very simple explanation strategy based on food features is integrated with a food recommender system. However, the authors did not evaluate its impact on users' food choices. Another simple explanation interface is presented in [11], where users' food preferences are linked to the ingredients of the recommended recipe, generating explanations such as '*Because you want food containing X*'. We go beyond [11], designing and evaluating a more comprehensive set of justification strategies.

Furthermore, the novelty of this work also lies in the automatic generation of *comparative* natural language justifications that emphasize similarities and differences between two *alternative* recommendations. Consumer decision-making research has shown that *how* two alternatives are compared (e.g., separately or comparatively) affects user preferences [3]. A remotely similar approach is presented by Chen et al. [8], who introduce a user interface where different recommendations are presented together with their distinctive features, obtained automatically from user reviews. However, in contrast with [8], rather than developing a completely novel user interface, we designed a framework to automatically generate a single natural language justification that compares two alternatives.

To conclude, we frame our approach with respect to the taxonomy of explanation strategies introduced in [15], labelling it as a *black box* methodology. Hence, the explanation strategy is not aware and independent of the underlying recommendation model, generating a *post-hoc* explanation that is not linked to the recommender algorithm. Post-hoc explanations provide reliable and effective explanations that are typically preferred by final users [32, 33]. We evaluate this framework by implementing two food recommender approaches: one that identifies *popular* recipes and one that selects *healthier* recipes. More details about the algorithms will be provided in the upcoming section.

Finally, we emphasize that the term *justification* is used, instead of the 'traditional' *explanation*. Even though both concepts appear to be synonymous, we follow the definition provided by Biran [4]: an *explanation* focuses on *how* the suggestion is generated, while *justifications* describe *why* a user would be interested in an item. This supposedly provides users with a means to make a more *informed decision* about consuming an item or not, fitting seamlessly to the current study's goal, for we evaluate whether and how natural language justifications affect users' online food choices.

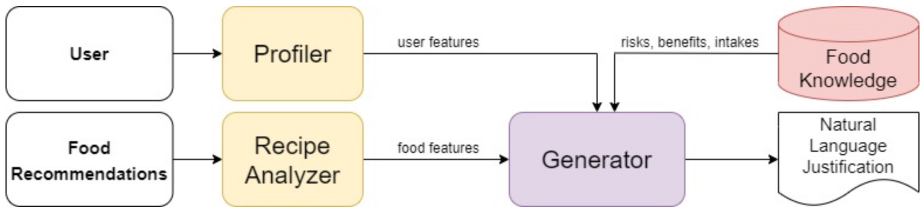

Fig. 1. Workflow to Generate Natural Language Justifications Supporting Food Recommendations.

3 METHODOLOGY

This section outlines our methodology to generate natural language justifications supporting food recommendations. We first introduce our workflow. Next, we focus on the *generation* phase, by introducing the different *strategies* we designed to justify a recommendation, along with the motivations that led to their implementation.

3.1 Description of the Workflow

The general workflow carried out by our framework is depicted in Figure 1. As shown in the figure, the methodology takes as input a *user* and two *food recommendations*, producing as output a *natural language justification*. The workflow is based on three main components: a Profiler module, whose goal is to collect information about the user, a Recipe Analyzer, which extracts the main features of the recommended recipes (e.g., nutrients, calories, ingredients), and a Generator, which builds the final justification based on user characteristics, food features, and knowledge about risks and benefits related to food consumption. As for the final output, we designed two general *justification styles* and eight different *justifications strategies*, each of which emphasizes different recipe characteristics or user features.

The Profiler module initiates the process to obtain information on a user. It implements a profiling strategy based on the *holistic user modeling paradigm* [5–7, 34], which has already been used in previous studies concerning food recommendations [35]. Table 1 outlines the seven user aspects used, which are encoded in each user profile: *demographics*, *preferences*, *goals*, *affect*, *behavioral data*, *health data*, and *domain-related information*.

Along with a user's characteristics, the workflow also acquires *food features*. These are, for example, the total amount of calories, macro-nutrient content (e.g., carbohydrates, fibers, fats, proteins), a recipe's preparation difficulty, cooking time, and its popularity on a recipe website. Some of these features are used to obtain the *healthiness* of a recipe based on the United Kingdom Food Standards Agency (FSA) Health Scores [37], which has been introduced by [47, 53] as a reference score related to food recommendations and food search. Generally speaking, all these features can be obtained by exploiting online resources, such as food communities. More details about the data collection procedure will be provided next. For our research goals, we can assume that all the above-mentioned information is available.

Table 1. User characteristics obtained by the Profiler Module in our natural language justification workflow.

User Aspect	Factors
Demographics	Gender, Age, Height, Weight
Preferences	Food Preferences and Restrictions (lactose-free, vegan, etc.)
Goals	Losing Weight (binary)
Affects	Mood (positive, negative, neutral)
Behavioral Data	Level of Physical Activity
Health Data	Lifestyle, BMI, Amount of Sleep, Stress
Domain Knowledge	Cooking Experience, Available Time, Cost Constraints

3.2 Generating Natural Language Justifications

After obtaining user characteristics and food features, the Generator module comes into play. This component's goal is to generate a natural language justification that supports the recommendation by emphasizing a recipe's nutritional facts, risks, or benefits, in order to encourage people to make a more informed and, possibly, healthier decision.

First, it is important to emphasize that the generation process follows the principles of Natural Language Generation systems [39], thus it is completely *automated* and *unsupervised*, and does not require any human intervention. Given this general setting, our framework can generate its output by following two different justification *styles*: *single* and *comparative*. As a reminder, our framework takes as input two different recipes: by following the first justification style, both of them are processed separately and each recipe is provided with a different justification. In contrast, a comparative justification compares the characteristics of the recipes, which is generated automatically by the algorithm.

To generate justifications, the algorithm also relies on general *food knowledge*. Our approach uses a food knowledge base that comprises *facts* related to the daily intake of macro-nutrients, as well as food consumption risks and benefits. This knowledge is based on general guidelines concerning food consumption, such as government publications, academic studies, and commonsense knowledge. In particular, for each of the main nutrients (i.e., carbohydrates, sugar, proteins, fats, fibers), around 10 facts are encoded. For example, "Consuming too much sugar increases the risk of diabetes", "High sodium intake increases health pressure", and "High protein intake improves muscle development". In total, we have encoded around 150 facts in our knowledge base, which are used in several justification strategies.

3.2.1 Overview of the Justification Styles and Strategies. Based on the above-mentioned setting, *eight* different justification strategies are implemented in the framework, across two justification styles. These strategies exploit different information sources and focus on different aspects. Regardless of the specific strategies, justifications are generated by exploiting a *template-based structure*. Each output follows a *fixed* structure and is *dynamically* filled in, based on: (i) characteristics of the user; (ii) features of the recipe; (iii) facts extracted from the food knowledge base. These aspects of the justification are generated separately and are concatenated to each other by using adverbs and conjunctions. In the following, we provide an overview of the *eight* justification strategies. Table 2 summarizes the output produced by the different strategies, along with details of the features they rely on, using *'Spaghetti Cacio and Pepper'*[1] and *'Vegetable Soup'* as running examples, providing an overview of the behavior of the framework.

Description. This justification strategy is based on a *textual description* of the recipe, which is gathered from online sources and is stored in our dataset as *food feature*. In this case, single and comparative justifications do not differ. The goal of this strategy is to provide the user with very general information about the recipe.

Popularity. This justification strategy is based on the *popularity score* of the recipe. This information is obtained as a food feature. For single justifications, we map the popularity score to a *categorical* popularity feature. In particular, we rank all the recipes based on their popularity scores and split them into four bins of equal size. When the explanation is generated, labels of the bins are used to provide information about the popularity of the recipe. Conversely, when a comparative explanation based on popularity is generated, popularity scores of the recipes are compared, and the one with the higher popularity score is emphasized. In this case, the justification provides information about how much popular is the recipe in the community, based on the the fact that people often use this criterion in food choices [13].

User Skills. This justification is grounded in the construct of self-efficacy, which is defined as *"beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments."* [2]. As hypothesized by Bandura [1, 13], people having high levels of self-efficacy belief tend to undertake more difficult and challenging tasks

[1] https://www.gimmesomeoven.com/cacio-e-pepe/

Table 2. Recap of the available Justification Strategies. We present examples for 'Single' and 'Comparative' justification styles.

Just. Strategy	Information Source		Example	
	User Features	*Food Features*	*Single Justification*	*Comparative Justification*
Description	none	Recipe Description	Spaghetti Cacio and Pepper are one of the dishes of the Roman Tradition: grated pecorino and peppercorns, a quick and tasty recipe. Vegetable Soup is a genuine and healthy dish, a perfect winter comfort food	n.a. (same as single)
Popularity	none	Recipe Popularity	Spaghetti Cacio and Pepper is very popular in the community. Vegetables Soup is poorly popular in the community	Spaghetti cacio and pepper is more popular than Vegetable Soup in the community
User Skills	Cooking Experience	Level of Difficulty	Spaghetti Cacio and Paper has a medium level of difficulty. It might not be adequate to your cooking skills, which are low. Vegetable Soup is very easy to prepare. The recipe seems adequate to your cooking skills, which are low.	Vegetable Soup is easier to prepare than Spaghetti Cacio and Pepper. They could be more adequate to your cooking skills, which are low.
Food Goals	Diet Goals	Calories	Spaghetti Cacio and Pepper has 491 calories. Please consider it, since your goal is to lose weight. Vegetable Soup has 462 calories. Please consider it, since your goal is to lose weight.	Spaghetti cacio and pepper has more calories than Vegetable Soup (491 vs. 462). Past of vegetables can better help to reach your goal of losing weight.
User Lifestyle	Personal Lifestyle	FSA Healthy Score	Spaghetti Cacio and Pepper is an unhealthy recipe according to FSA Score. Please consider this, since you aim to have a healthy lifestyle. Vegetable Soup is a healthy recipe, according to FSA Score. Please consider this, since you aim to have a healthy lifestyle.	Accordingly to FSA Score, Vegetable Soup is healthier than Spaghetti Cacio and Pepper. Please consider this, given the importance you give to a healthy lifestyle
Food Features	Preferences and Restrictions	Nutritional Information, Ingredients	Spaghetti Cacio and Pepper has 8.7gr of saturated fats and 2.3gr of fibers. Vegetable Soup has 4.55gr of saturated fats and 7.3gr of fibers.	Spaghetti Cacio and Pepper has a higher amount of saturated fats (8.7gr vs. 4.55gr) and a lower amount of fibers (4.55gr vs. 7.3gr) than Vegetable Soup.
Health Risks	BMI, Mood Sleep, Stress, Physical Activity,	Nutritional Information, Ingredients	Spaghetti Cacio and Pepper has 8.7 gr of saturated fats and 2.3gr of fibers. To intake many saturated fats increases the risk of heart diseases. Given your high BMI, you should take into account this fact. Vegetable Soup has 4.55gr of saturated fats and 7.3gr of fibers. To intake many fibers increases the risk of constipation.	Spaghetti Cacio and Pepper has a higher amount of saturated fats (8.7gr vs. 4.55gr) and a lower amount of fibers (4.55gr vs. 7.3gr) than Vegetable soup. To intake many saturated fats increases the risk of heart diseases. Given your high BMI, you should take into account this fact. On the other side, to intake many fibers increases the risk of constipation.
Health Benefits	BMI, Mood, Sleep, stress, Physical Activity	Nutritional Information, Ingredients	Spaghetti Cacio and Pepper has 8.7gr of saturated fats and 2.3gr of fibers. To intake many saturated fats improves your energy supply. Given your current stress level, this can be helpful. Vegetable Soup has 4.55gr of saturated fats and 7.3gr of fibers. To intake many fibers reduces the risk of cancer.	Spaghetti Cacio and Pepper has a higher amount of saturated fats (8.7gr vs. 4.55gr) and a lower amount of fibers (4.55gr 7.3gr) than Vegetable Soup. To intake many saturated fats improves your energy supply. Given your current stress, this can be helpful. On the other side, to intake many fibers reduces the risk of cancer.

than people with low levels of self-efficacy. In this perspective, this justification strategy aims to match users' beliefs in their own cooking skills with the difficulty of the recipe execution. People who rate themselves as being low-skilled are matched to easy recipes, while users with high self-perceived skills are shown challenging recipes, an intuition that is rarely applied in (food) RS research [42, 46, 48]. Specifically, we compare the *'cooking experience'* feature encoded in the profile to a recipe's *'level of difficulty'*. Single justifications first present the recipe's level of difficulty of (e.g., high, medium, low), which is compared to the user's self-reported cooking skills afterwards: if a user's skills are lower than or equal to the recipe's difficulty, a string indicating that the recipe is adequate for the user is concatenated to the justification. Vice versa, the opposite information (e.g., Recipe X is not adequate) is shown. As for comparative justifications, the levels of difficulty of the recipes are compared. If the self-reported cooking skills of the user are low, the easiest recipe is emphasized. Otherwise, the justification first shows the most difficult one.

User Goals. Food choices are often driven by specific goals of users, such as losing weight. Goal-setting theory [28] shows that people make decisions and take action in line with their set goal, particularly if that goal is important to the individual (e.g., self-set rather than assigned) [31]. Accordingly, this explanation strategy links a user's self-set goals to the total amount of calories of the recipe. For single justifications, we generate a sentence presenting the calories of the recipe, along with the suggestion to consider this information in a user's food decision. For single justifications, sentences are filled in by using the *amount of calories* of each recipe. As for comparative justifications, they contrast the calories of the two recipes in the format (*'X has more calories than Y'*). Furthermore, if it is the user's goal to lose weight, we generate a second sentence that indicates the recipe with fewer calories.

User Lifestyle. Users' personal values, such as the importance of maintaining a healthy lifestyle, can strongly influence food choices. The value-attitude-behavior model explains that both *values* and *attitudes* impacts on *behavior* [54]. In this perspective, research has shown that people's health values have a positive effect on both their attitudes towards low-fat or low-calories menu items and their behavioral intentions to choose healthy menus [25]. Accordingly, this justification strategy links users' health values to a recipe's healthiness, which is computed based on the popular FSA Health Score [53]. In single justifications, our template fills in a recipe's name and a categorical label (*unhealthy, quite healthy, healthy*), based on the FSA Health Score of the recipe. Moreover, in line with the previous justification strategy, we generate a new sentence that suggest the user to interpret this information in line with their lifestyle self-assessment. Comparative justifications work in a similar way, by first comparing the recipes' FSA scores and. subsequently, generating a sentence that emphasizes the healthier recipe if that is in line with the user's self-health assessment. If not, a simple comparison between the recipes is presented (e.g., 'X is healthier than Y').

Food Features. The goal of this strategy is to inform the user about the ingredients of each recipe. Research highlights that nutritional knowledge contributes to better food choices and a more adequate nutrient intake [21]. Moreover, people with higher food knowledge are more likely to meet the current recommendations for fruit, vegetable and fat intake than individuals with lower knowledge levels [56]. For single justifications, our framework compares the amount of macronutrients (e.g., protein) and salt in a recipe to the recommended daily intake [30]. Next, two randomly selected nutrients that in the top-3 in terms of % of the recommended daily intake are used to fill in the template. Comparative justifications compare the nutrients by automatically generating a lexicalization of the characteristics, such as *'X contains more protein and fats than Y, but fewer carbohydrates'*.

Health Risks. This justification strategy can be seen as an extension to the *'food features'* strategy, linking information about macronutrients to *health risks*. Justifications are based on the *health belief model*, which posits that health behavior is affected by the *perceived susceptibility* to illness or health problems and the *perceived severity* of the consequences associated with the state or condition, the sum thereof is called *perceived threat* [40, 49]. Based on this theoretical foundation, we generate *risk-aware* (single) justifications that are split into three parts. First, we follow the *food features* strategy by presenting the main macronutrients of each recipe. Second, this is linked to the previously mentioned *food knowledge base*. In this case, we retrieve facts that match the main characteristics of the recipe. For example, if *'saturated fats'* has been previously selected by the algorithm, a fact describing health risks related to overconsuming saturated fats is randomly retrieved among those available in the knowledge base (e.g., 'the intake of too many saturated fats increases the risk of heart disease'). Finally, the third part of the justification links some characteristics of the *user* to further health risks. For instance, if the user's self-reported features indicate that she is overweight or do not do engage in sufficient physical activity, our framework could highlight a risk related to heart diseases. For comparative justifications, the algorithm first compares the different levels of macronutrients, presenting two different sentences that each link food characteristics to health risks. 'Health Risks' is our most comprehensive strategy, for it links information about food, risks, and user characteristics into a single natural language justification.

Health Benefits. This justification strategy is analogous to the previous one and is also based on the health belief model. The key difference is that the current strategy focuses on *health benefits* rather than health risks, which is also highlighted by the health belief model: the *perceived benefits* of a health behavior are also an important determinant [40, 49]. Apart from this aspect, the justification follows the same structure as the previous one. It presents nutritional information and food characteristics first, after which it selects a number of recipe aspects that are linked to food facts encoded in our knowledge base. Similar to the Health Risks strategy, a randomly selected fact that matches the selected characteristic of the recipe is chosen. Finally, the fact is linked to a user's characteristics. If a further match emerges, a new sentence is introduced in the justification. This setup applies to both single and comparative justifications.

4 EXPERIMENTAL EVALUATION

We examined whether natural language justification affected user preferences for healthy recipe recommendations, compared to popular ones. In the following, we described the setup of our online user study, introducing the dataset, participants, and the research design. Subsequently, we compared our main conditions (i.e., a *single* justification and a *comparative* justification) to our no explanation baseline. Moreover, we explored which specific justifications strategies (e.g., health benefits) affected food choices the most and the examined the underlying choice motivations.

4.1 Method

4.1.1 Dataset. Recipes were sampled from a database of 4,671 recipes, which is available online:[2]. Recipes were obtained from a popular food community platform[3], and translated to English. The recipes contained information about their name, category, preparation difficulty, as well as their ingredients, (macro-)nutrients, calories, rating count, and average website rating. Moreover, they also included several binary tags, such as *vegetarian*, *vegan*, *lactose-free*, and *low-nickel*.

4.1.2 Food Recommendation Algorithms. Recipes were retrieved using two different personalized food algorithms. In the following, we refer to our recommendation algorithms as *health-aware* or healthy, and *popular* or popularity-based. For the former, we obtained *healthy* recipes based on user characteristics, goals, and constraints, retrieved through our healthy-aware food recommendation algorithm [35]. In the second case, a *popular* recipe was identified by the algorithm. We wish to reiterate that the algorithms were entities separate from our natural language justification framework, and were thus considered as independent parameters in our analyses.

[2] https://tinyurl.com/recipes-data-umap
[3] https://www.giallozafferano.it/

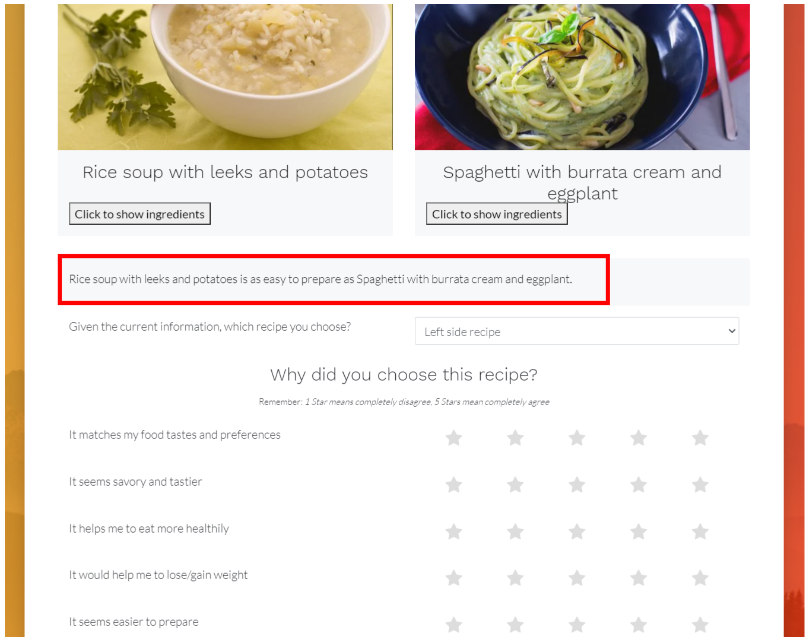

Fig. 2. The study's interface for two first course meals. The recipe depicted on the left is our healthy-algorithm recommendation, the one on the right is generated by a popular algorithm. Depicted within the red box is a justification in a specific style, in this case a 'Comparative' User Skills justification; the box is missing in the 'No Justification' condition. Users were asked to choose one recipe or neither of them, and to provide reasons why they had chosen a recipe.

4.1.3 Participants. Participants were recruited at Amazon MTurk to complete a study in which they would receive three recipe pairs of food they could enjoy. Participants were required to be US-based and to have a hit rate of 98%, with a minimum of 500 approved hits. [4] The typical completion time fell between 3 and 10 minutes, for which participants were reimbursed with 0.5 USD. In total, 503 participants (54.7% Male) completed our user study, among which 61.0% was between 20 and 39 years old. The majority of users was employed (73.6%; 14.9% was student) and had a weight loss goal (51.1%), while only 70 users (13.9%) had a weight gain goal.

4.1.4 Procedure. Users were first asked various questions that were used to model their profile. The features used are outlined in Table 1, and included questions about demographics, self-reported health and well-being, experience with home cooking, and dietary restrictions and preferences. After users submitted their responses, the profiler (cf. Figure 1) would generate three pairs of recommendations. One example of a such recommendation pair is presented in Figure 2, of which the left one based on the *healthy* food RS, and the one on the right was generated using a *popularity-based* algorithm. Three different pairs were presented sequentially to a user, presenting two first course meals, two second courses and, finally, two desserts. For each pair, users were asked to either choose the left-hand side or right-hand side recipe, or neither. Note that we did not inform the users about which recipe was the *healthy* recommendation, or if there was any for that matter. Users who had chosen one of the two recipes were asked, in turn, to indicate to what factors were underlying their decision, such as a recipe's healthiness, taste, or ease of preparation.

[4] Such participants are more likely to generate high-quality data and to meet attention checks.

4.1.5 Research Design. Whether justifications were presented underneath each recipe pair or not was subject to three between-subject conditions. Users were either presented no justification for the presented recipes (i.e., baseline), a justification style that focused on each recipe separately (i.e., 'Single Justification'), or a justification style that compared the two recipes (i.e., 'Comparative Justification'). The strategy in which a single or comparative justification was presented, was subject to eight within-subject conditions, which are outlined in Table 2. This way, one user could be presented three different single justifications (e.g., Popularity, Food Goals, and Health Risks), while another user would be presented three different comparative justifications (e.g., User Lifestyle, Food Features, Health Benefits), or no explanation for each recipe. Figure 2 shows an example of a 'User Skills' justification, depicted within the red box.

4.1.6 Measures. For our analyses, we considered the effect of different justification styles on the percentage of healthy recommendations chosen. We did so by comparing the 'No Justification' baseline either with any justification style, with 'Single' and 'Comparative' justifications separately, or across all different strategies outlined in Table 2. The effectiveness of different justification styles were contrasted against the no explanation baseline, across all dish types for all choices made (i.e., choosing the popular recommendation or choosing neither of the recipes). Different justification strategies were compared between the no explanation baseline and the comparative style, because we found that 'Comparative' was the most effective justification style.

Furthermore, we examined a user's motivation for choosing any of the two presented recipes. Users were asked to indicate on 5-point scales to what extent a reason was applicable as to why they had chosen either recipe. Choice motivation items were related to 'a match with the user's preferences', the recipe's taste, healthy eating goals, weight-loss or gain goals, and ease to prepare.

We also inquired on a set of user characteristics, which was also used by the profiler to generate healthy recommendations, as described in Table 1. Besides obtaining information on demographics (i.e., gender, age, BMI) and food preferences, we asked users whether they had any eating goals (i.e., either weight-loss, weight-gain, or no goals), to rate the healthiness or their lifestyle and the importance of having that (5-point scales). Users were also asked to rate their frequency (5-point scale) of either making healthy food choices, looking at food's nutritional values, using recipe websites, and engaging in home cooking. Moreover, with regard to well-being, we inquired on their current mood, level of sleep and level of physical activity (3-point scales), whether they reported to be stressed or depressed ('yes' or 'no'). Finally, we inquired on users' domain knowledge, asking them to indicate their self-reported cooking experience (5-point scale), as well as their time and cost constraints for cooking.

4.2 Results

We examined user choice behavior through three different analyses.[5] First, we investigated whether presenting *any* explanation affected user preferences for healthy recommendations. Second, we examined preferences for different justification strategies. Third, we examined why users had either chosen healthy or popular recipes.

4.2.1 Single and Comparative Justifications. We investigated whether users were more likely to choose healthier recipes if any justification was presented. We used a one-way ANOVA to examine choices made across all meal types, which showed that the healthy recommendation was chosen more often if any justification was presented alongside it (47.4% of choices, $S.E. = 1.6\%$), compared to the 'No Justification' baseline ($M = 38.1\%, S.E. = 2\%$): $F(1, 1507) = 11.80, p < 0.001$. This suggested that justifications helped to steer user preferences towards the health-aware recommendation. To

[5] The collected data, as well as the analysis scripts can be obtained via https://osf.io/vytdx/.

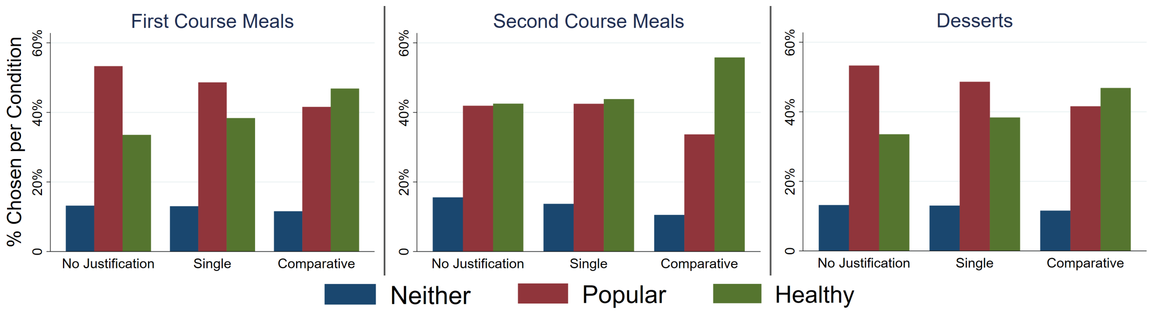

Fig. 3. Percentages of choices per condition, per meal type. Depicted are choices for neither recipe (in blue), the Popular recipe (in red), and the Healthy recommendation across three different meal types. Conditions are the three different justification styles: No justification, single justifications, and comparative justifications. Meal types are First Course, Second Course, and Dessert.

inspect the 'Single' and 'Comparative' styles separately, we performed a second one-way ANOVA across all meal types. Although users were not more likely to choose the healthy recommendation when presented a 'Single Justification' (42.7% of choices, $S.E. = 2.4\%$, $p = 0.16$), compared the baseline (38.1%), they were more likely to do so if a 'Comparative Justification' ($M = 51.1\%$, $S.E. = 2.1\%$) was presented: $F(1, 1506) = 18.25$, $p < 0.001$. This suggested that comparative justifications were particularly effective in promoting the healthy recommendation.

Since previous research showed that preferences differed across meal types [35], we also examined choices for the healthy recommendation per dish type. Using multiple one-way ANOVA's, we found that depicting a justification increased the number of choices for healthy recommendations for first courses ($F(1, 502) = 4.70$, $p < 0.05$) and desserts ($F(1, 502) = 4.32$, $p < 0.05$), but found no such effect for second course meals ($F(1, 502) = 2.92$, $p = 0.09$).[6] Inspecting the effects more closely by differentiating between 'Single' and 'Comparative' justifications per meal type in multiple one-way ANOVAs, we found an important distinction. While 'Single' justifications did not significantly boost healthy recommendation choices in any dish type (all p-values > 0.1), 'Comparative' justifications did: for first courses ($F(1, 500) = 5.37$,$p < 0.05$), second courses ($F(1, 500) = 6.33$, $p < 0.05$), and desserts ($F(1, 500) = 6.61$, $p < 0.05$). This provided further evidence that justifications that compared healthy and popular recommendations were more effective in steering user preferences towards the healthy recommendation, compared to separate justifications per recipe.

The reported tests can be understood better by inspecting Figure 3. Depicted are choices per meal type (from left to right: first course, second course, dessert), for which we examined per justification style what percentage of the preferred options were chosen: neither recipe, the popularity-based recommendation, or the health-aware recommendation. For first course meals and desserts, it was clear that the 'Single' justification only increased the number of choices for healthy recommendation a little, while Comparative justifications increased that effect much further. For second course meals, there was little difference between 'No Justification' and 'Single' in terms of choices made, while 'Comparative' boosted the choices for the healthy recommendation.

4.2.2 Justification Strategies. The previous subsection showed that pairwise justifications were the most effective in steering user preferences towards healthy recommendation. Here, we examined the effectiveness of the specific justification strategies (cf. Table 2) to promote our healthy recommendations.[7]

[6] Performing a Repeated Measures ANOVA that included 'meal type' as a categorical variable did not affect the main effects of the explanation styles.
[7] We also examined choices for different justification strategies across all conditions (both 'Single' and 'Comparative'), as well as for 'Single' Justifications only. Although nearly all effects pointed into a similar direction, fewer differences were significant; mostly for 'Single' justifications. Since 'Comparative' justifications were shown to be the most effective in the previous subsection, we only reported the results for that style.

We examined the effectiveness across all meal types, as well as per type. Table 3 describes four different logistic regression analyses, which each predicted whether our health-aware recommendation was chosen (compared to a popularity-based choice or no recipe chosen). We found effects to be mixed across the different meal types, while the second course and dessert models had the highest pseudo R^2-values. However, all significant effects across all models were positive, indicating that the different justification strategies in the comparative condition increased the likelihood that the healthier recommendation was chosen, not the popularity-based option.

We first examined significant differences. The model across all meal types in Table 3 shows that three justification strategies effectively supported health-aware choices. A comparison of the food features of the two recipes (e.g., Recipe A contains less fat than Recipe B) was related to a higher likelihood of choosing the healthy recommendation compared to the no justification baseline: $\beta = .86$, $p < 0.001$ (also in the first course model), as did justification that compared the health risks of both recipes: $\beta = .98$, $p < 0.001$ (also in the second course and dessert models).

In a similar vein, comparing recipes in terms of their health benefits led users to choose the healthier dessert more often: $\beta = .84$, $p < 0.05$, but not for other meal types. Table 3 also shows that comparing recipes in terms of food goals increased the likelihood of choosing the healthy option for first courses: $\beta = .78$, $p < 0.05$, but not for second courses and dessert. In contrast, a somewhat counterintuitive effect was that a popularity justification strategy, which typically showed that the healthy recipe was less popular than the popularity-based recommendation, increased the likelihood of choosing the healthy recommendation: $\beta = .59$, $p < 0.05$ (also in the dessert model).

Table 3 also highlights which strategies did not affect preferences between the 'Comparative' and 'No Explanation' conditions. Providing comparative descriptions of the recipe contents (e.g., the ingredients) did not affect user preferences, nor did comparing whether the recipes match with the user's lifestyle – for each meal type. Moreover, notable was that comparative justifications of food goals did not affect dessert choices, while emphasizing health risks and benefits did not influence choices for first course meals.

4.2.3 Choice Motivation. Beyond justification styles and strategies, we finally examined *why* users had chosen either recipe. To do so, we performed four logistic regression analyses to compare why users had either chosen the healthy or popular recommendation, ignoring cases where neither recipe was chosen. Table 4 outlines a model that includes a user's choice motivation across all meal types, as well as three meal-specific models, where in each model positive effects indicated reasons why the healthy recommendation was chosen, while negative effects indicated why the popular

Table 3. Four logistic regression models, predicting choices for healthy-aware recommendations (against no choice or popularity-based choices) in the 'Comparative' justification condition, compared to the no explanation baseline. The first model examines choices across all meal types, the other models are meal type-specific. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Justification Style	**All Meal Types** β *(S.E.)*	**First Course** β *(S.E.)*	**Second Course** β *(S.E.)*	**Dessert** β *(S.E.)*
Description	.37 (.24)	.56 (.43)	.094 (.40)	.49 (.40)
Popularity	.59 (.28)*	-.30 (.52)	.92 (.49)	1.14 (.51)*
User Skills	.58 (.32)	.48 (.60)	.62 (.49)	.50 (.63)
Food Goals	.44 (.23)	.78 (.39)*	.64 (.44)	-.058 (.42)
User Lifestyle	.047 (.25)	.35 (.39)	-.17 (.43)	-.23 (.51)
Food Features	.86 (.24)***	1.11 (.44)*	.74 (.41)	.76 (.42)
Health Risks	.98 (.26)***	.39 (.45)	1.51 (.49)**	1.09 (.44)*
Health Benefits	.42 (.27)	.28 (.48)	.079 (.16)	.84 (.43)*
Intercept	-.48 (.092)***	-.48 (.16)**	-.30 (.16)	-.68 (.16)***
Pseudo R^2	.0199	.0238	.0361	.0337

Table 4. Four logistic regression models, each predicting user choices for the *Healthy Recommendation*. Models either included choices across all meal types ($N = 1,339$), or only meal-specific choices: First Course ($N = 462$), Second Course ($N = 437$), and Desserts ($N = 440$). We only considered recipe pairs why users had either chosen the healthy recommendation (positive effects) or the popular recommendation (negative effects). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Choice Motivation	**All Meal Types** β *(S.E.)*	**First Course** β *(S.E.)*	**Second Course** β *(S.E.)*	**Dessert** β *(S.E.)*
Matched User Preferences	.13 (.070)*	-.052 (.13)	.52 (.13)***	-.20 (.12)
Tastiness	-.47 (.072)***	-.54 (.14)***	-.58 (.13)***	-.22 (.11)
Healthiness	.41 (.063)***	.78 (.12)***	.13 (.11)	.47 (.12)***
Matched Food Goals	.061 (.062)	-.0031 (.11)	.21 (.11)*	-.13 (.12)
Easiness	-.080 (.054)	-.26 (.099)**	-.030 (.098)	-.047 (.096)
Intercept	.040 (.32)	.52 (.62)	.-52 (.55)	.62 (.56)
Pseudo R^2	.0628	.134	.0671	.0545

recommendation was chosen. The best model fit was observed for the first course meals model, for which the pseudo R^2 was around two times higher than for the other models.

We observed mixed evidence for why healthy recommendations were chosen across different meal types. Our health-aware recommendations were chosen more often because of health-related reasons. A positive effect was found across all meal types (β=.41, $p < 0.001$), as well as for first course meals (β=.78, $p < 0.001$) and desserts (β=.47, $p < 0.001$). In contrast, tastiness was related to popular meal choices: averaged across meal types (β=-.47, $p < 0.001$), as well as for first course (β=-.54, $p < 0.001$) and second course meals (β=-.58, $p < 0.001$). Furthermore, users who indicated to choose recipes because they matched their preferences, were more likely to choose our health-aware recommendations across all meal types (β=.13, $p < 0.05$), in particular for second course meals (β=.52, $p < 0.001$). Second course healthy recipes were also chosen more often because a match in food goals: β=0.21, $p < 0.05$. In contrast, easiness was negatively related to choosing healthy first course recommendations (β=-.26, $p < 0.01$), suggesting that users had chosen first course popular recommendations because they were easier to prepare, while no such effects were observed for second course meals and desserts.

5 CONCLUSIONS AND FUTURE WORK

The contribution of this paper is twofold. First, we present a recommendation approach that captures a user's eating preferences. In contrast with most earlier work [14, 52], we do not focus on recipes that users liked in the past, but we consider a user's general eating preferences, affect, self-reported skills, and domain knowledge. This has resulted in a recommendation pipeline that presents personalized, yet healthier recommendations. Second, we have presented an approach to generate natural language justifications food recommendations. While the NLP pipeline is a contribution in its own respect, particularly in a food recommender system, we have also validated its effectiveness by showing what types of justifications are most effective to promote our health-aware recommendations, through a user study. Whereas popular recipes are preferred by most users if no explanation is presented (our 'baseline'), we have shown that most users prefer our health-aware recommendations over a challenging popularity-based recommendation baseline, when presenting both recommendations along with a comparative justification.

With regard to specific justification styles, we find that comparative approaches are more effective in promoting choices for health-aware recommendations than single justifications. This taps into research that people are much at making comparative judgments than combining two 'singular' observations [3], which is reflected by the effectiveness of our 'Comparative' justification style over the 'Single' style. The obtained evidence is convincing, since we have observed this effect across different meal types – even desserts, for which food choices tend to be more related to taste instead of health [35]. Moreover, we have also examined the effectiveness of specific justification strategies, suggesting that presenting a comparison of each recipe's features and health risks seems to cater towards a user's healthy food preferences. The sophistication of these strategies may have contributed to their effectiveness, for they link and compare different aspects, namely user characteristics, recipe features, and food goals. Although the large number of comparisons for specific justification styles may have been prone to a higher false positive rate, the overall results point out that all explanation strategies either promote healthy food choices – even the popularity-based strategy – or have no net effect.

We have also examined what drives users to choose healthier recommendations, and whether this differs per meal type. For most meal types, we have found evidence that popularity-based choices are related to taste motivations, while choices for our health-aware recommendation are linked to health-related reasons. This confirms that our health-aware recommendation pipeline caters to users with healthy eating goals, which is promising for future applications that seek to support such users. Moreover, 'because it fits my preferences' is also found to be a reason to choose the healthy recommendation across all meal types, suggesting that our approach could generate both satisfactory and healthy food recommendations, which is rarely found in food RSs to date [52].

An interesting avenue of future research is to test whether the insights can be generalized in a practical application if more than two recipes in a recommendation list. Moreover, we will introduce justifications combining several user-focused aspects, such as food taste and goals, to assess whether these can persuade a user to choose the healthier recommendation. Moreover, we will investigate whether such natural language justifications can be personalized further, and whether this would increase their effectiveness. For example, presenting justification styles that address healthy eating goals make more sense if a user has indicated to have such a goal. While the current user study has done so by inquiring on the user's preferences in the first screen, such questions would only need to be asked when a user's profile is created, for instance on a recipe website.

Finally, we wish to emphasize that the study can serve as a blueprint for future studies on healthy food recommendation. We have shown that our algorithm successfully generates healthy recommendations, as users who chose them indicated to have health-related choice reasons. Moreover, we have also shown *how* such recommendations should be presented to support healthy food choices. Such a combination of a knowledge-aware algorithm and UI design should pave the way for even more sophisticated applications in food recommendation, as well as for applications in other behavioral recommendation domains. Moreover, future work should extend the number of inputs in the recommender framework, by taking into account a larger and more comprehensive set of algorithms and to evaluate them.

6 ACKNOWLEDGMENTS

This work is in part funded by MediaFutures partners and the Research Council of Norway (grant number 309339). Moreover, we would thank Andrea Montemurro and Ivan Paparella for the implementation effort.

REFERENCES

[1] Albert Bandura. 1986. *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall.

[2] Albert Bandura. 1997. *Self-Efficacy: The Exercise of Control.* W. H. Freeman.

[3] James R Bettman, Mary Frances Luce, and John W Payne. 1998. Constructive consumer choice processes. *Journal of consumer research* 25, 3 (1998), 187–217.

[4] Or Biran and Courtenay Cotton. 2017. Explanation and justification in machine learning: A survey. In *IJCAI-17 Workshop on Explainable AI (XAI).* 8.

[5] Federica Cena, Silvia Likavec, and Amon Rapp. 2018. Real World User Model: Evolution of User Modeling Triggered by Advances in Wearable and Ubiquitous Computing: State of the Art and Future Directions. *Information Systems Frontiers* (2018), 1–26. https://doi.org/10.1007/s10796-017-9818-3 cited By 0; Article in Press.

[6] Federica Cena, Amon Rapp, Cataldo Musto, and Pasquale Lops. 2018. Towards a Conceptual Model for Holistic Recommendations. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization* (Singapore, Singapore) *(UMAP '18)*. Association for Computing Machinery, New York, NY, USA, 207–210. https://doi.org/10.1145/3213586.3225248

[7] Federica Cena, Amon Rapp, Cataldo Musto, and Giovanni Semeraro. 2020. Generating Recommendations From Multiple Data Sources: A Methodological Framework for System Design and Its Application. *IEEE Access* 8 (2020), 183430–183447. https://doi.org/10.1109/ACCESS.2020.3028777

[8] Li Chen and Feng Wang. 2017. Explaining Recommendations based on Feature Sentiments in Product Reviews. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. ACM, 17–28.

[9] Pin-Jane Chen and Marta Antonelli. 2020. Conceptual Models of Food Choice: Influential Factors Related to Foods, Individual Differences, and Society. *Foods* 9, 12 (2020), 1898.

[10] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. The Effects of Transparency on Trust and Acceptance of a Content-based Art Recommender. *User Modeling and User-Adapted Interaction* 18, 5 (2008), 455–496.

[11] Mehdi Elahi, Mouzhi Ge, Francesco Ricci, David Massimo, and Shlomo Berkovsky. 2014. Interactive Food Recommendation for Groups.. In *Recsys posters*. Citeseer.

[12] David Elsweiler, Morgan Harvey, Bernd Ludwig, and Alan Said. 2015. Bringing the" healthy" into Food Recommenders.. In *Decision Making and Recommender Systems - Proceedings of the 2nd International Workshop on Decision Making and Recommender Systems*. 33–36.

[13] David Elsweiler, Christoph Trattner, and Morgan Harvey. 2017. Exploiting Food Choice Biases for Healthier Recipe Recommendation. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Shinjuku, Tokyo, Japan) *(SIGIR '17)*. Association for Computing Machinery, New York, NY, USA, 575–584. https://doi.org/10.1145/3077136.3080826

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

[15] Gerhard Friedrich and Markus Zanker. 2011. A taxonomy for generating explanations in recommender systems. *AI Magazine* 32, 3 (2011), 90–98.

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

[17] Sofia Gkika and George Lekakos. 2014. The Persuasive Role of Explanations in Recommender Systems.. In *BCSS@ PERSUASIVE*. 59–68.

[18] Kristian J Hammond. 1986. CHEF: A model of case-based planning.. In *AAAI*. 267–271.

[19] J. Herlocker and J. Konstan. 2001. Content-Independent Task-Focused Recommendation. *IEEE Internet Computing* 5, 6 (2001), 40–47.

[20] Santiago Hors-Fraile, Francisco J Núñez Benjumea, Laura Carrasco Hernández, Francisco Ortega Ruiz, and Luis Fernandez-Luque. 2016. Design of two combined health recommender systems for tailoring messages in a smoking cessation app. *arXiv preprint arXiv:1608.07192* (2016).

[21] J.Z. Ilich, J.A. Vollono, and R.A. Brownbill. 1999. Impact of Nutritional Knowledge on Food Choices and Dietary Intake of College Students. *Journal of the American Dietetic Association* 99, 9, Supplement (1999), A89. https://doi.org/10.1016/S0002-8223(99)00707-5

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

[23] Renata Januszewska, Zuzanna Pieniak, and Wim Verbeke. 2011. Food choice questionnaire revisited in four countries. Does it still measure the same? *Appetite* 57, 1 (2011), 94–98.

[24] Hilary Johnson and Peter Johnson. 1993. Explanation Facilities and Interactive Systems. In *Proceedings of the 1st international conference on Intelligent user interfaces*. ACM, 159–166.

[25] Jinhyun Jun, Juhee Kang, and Susan W. Arendt. 2014. The effects of health value on healthful food selection intention at restaurants: Considering the role of attitudes toward taste and healthfulness of healthful foods. *International Journal of Hospitality Management* 42 (2014), 85 – 91. https://doi.org/10.1016/j.ijhm.2014.06.002

[26] 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, 4-5 (2012), 441–504.

[27] Nadja Leipold, Mira Madenach, Hanna Schäfer, Martin Lurz, Nada Terzimehic, Georg Groh, Markus Böhm, Kurt Gedrich, and Helmut Krcmar. 2018. Nutrilize a Personalized Nutrition Recommender System: an Enable Study. *HealthRecSys@ RecSys* 2216 (2018), 24–29.

[28] Edwin A Locke and Gary P Latham. 2002. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *Am Psychol* 57, 9 (Sep 2002), 705–717. https://doi.org/10.1037//0003-066x.57.9.705

[29] Jerko Markovina, Barbara J Stewart-Knox, Audrey Rankin, Mike Gibney, Maria Daniel V de Almeida, Arnout Fischer, Sharron A Kuznesof, Rui Poínhos, Luca Panzone, and Lynn J Frewer. 2015. Food4Me study: Validity and reliability of Food Choice Questionnaire in 9 European countries. *Food Quality and Preference* 45 (2015), 26–32.

[30] Linda D Meyers, Jennifer Pitzi Hellwig, Jennifer J Otten, et al. 2006. *Dietary reference intakes: the essential guide to nutrient requirements*. National Academies Press.

[31] S. A. Munson and S. Consolvo. 2012. Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. 25–32. https://doi.org/10.4108/icst.pervasivehealth.2012.248691

[32] Cataldo Musto, Marco de Gemmis, Pasquale Lops, and Giovanni Semeraro. 2020. Generating post hoc review-based natural language justifications for recommender systems. *User Modeling and User-Adapted Interaction* (2020), 1–45.

[33] Cataldo Musto, Fedelucio Narducci, Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. 2019. Linked open data-based explanations for transparent recommender systems. *International Journal of Human-Computer Studies* 121 (2019), 93–107.

[34] Cataldo Musto, Marco Polignano, Giovanni Semeraro, Marco de Gemmis, and Pasquale Lops. 2020. Myrror: a platform for holistic user modeling. *User Model. User Adapt. Interact.* 30, 3 (2020), 477–511. https://doi.org/10.1007/s11257-020-09272-6

[35] 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, UMAP 2020, Genoa, Italy, July 12-18, 2020*, Tsvi Kuflik, Ilaria Torre, Robin Burke, and Cristina Gena (Eds.). ACM, 333–337. https://doi.org/10.1145/3340631.3394880

[36] Ingrid Nunes and Dietmar Jannach. 2017. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction* 27, 3-5 (2017), 393–444.

[37] Department of Health. 2013. Guide to creating a front of pack (FoP) nutrition label for pre-packed products sold through retail outlets.

[38] John Prescott, Owen Young, L O'neill, NJN Yau, and R Stevens. 2002. Motives for food choice: a comparison of consumers from Japan, Taiwan, Malaysia and New Zealand. *Food quality and preference* 13, 7-8 (2002), 489–495.

[39] Ehud Reiter and Robert Dale. 2000. *Building natural language generation systems.* Cambridge university press.

[40] Irwin M. Rosenstock. 1974. Historical Origins of the Health Belief Model. *Health Education Monographs* 2, 4 (1974), 328–335. https://doi.org/10.1177/109019817400200403 arXiv:https://doi.org/10.1177/109019817400200403

[41] 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.* 157–161.

[42] Hanna Schäfer and Martijn C Willemsen. 2019. Rasch-based tailored goals for nutrition assistance systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces.* 18–29.

[43] Benjamin Scheibehenne, Linda Miesler, and Peter M Todd. 2007. Fast and frugal food choices: Uncovering individual decision heuristics. *Appetite* 49, 3 (2007), 578–589.

[44] Rashmi Sinha and Kirsten Swearingen. 2002. The Role of Transparency in Recommender Systems. In *CHI'02 extended abstracts on Human factors in computing systems.* ACM, 830–831.

[45] Alain Starke. 2019. RecSys Challenges in achieving sustainable eating habits.. In *HealthRecSys@ RecSys.* 29–30.

[46] Alain Starke, Martijn Willemsen, and Chris Snijders. 2017. 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.* 65–73.

[47] Alain D. Starke, Martijn C. Willemsen, and Christoph Trattner. 2020. Nudging Healthy Choices in Food Search Through Visual Attractiveness. *Frontiers in Artificial Intelligence.* Preprint, osf.io/7w5dc (26 Nov. 2020).

[48] Alain D Starke, Martijn C Willemsen, and Chris CP Snijders. 2020. Beyond "one-size-fits-all" platforms: Applying Campbell's paradigm to test personalized energy advice in the Netherlands. *Energy Research & Social Science* 59 (2020), 101311.

[49] David Taylor, Michael Bury, Natasha Campling, Sarah Carter, Sara Garfied, Jenny Newbould, , and Tim Rennie. 2007. *A Review of the use of the Health Belief Model (HBM), the Theory of Reasoned Action (TRA), the Theory of Planned Behaviour (TPB) and the Trans-Theoretical Model (TTM) to study and predict health related behaviour change.* Technical Report. University of London.

[50] Nava Tintarev and Judith Masthoff. 2012. Evaluating the Effectiveness of Explanations for Recommender Systems. *UMUAI* 22, 4-5 (2012), 399–439.

[51] 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.

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

[53] 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.* 489–498.

[54] Alina Tudoran, Svein Ottar Olsen, and Domingo C. Dopico. 2009. The effect of health benefit information on consumers health value, attitudes and intentions. *Appetite* 52, 3 (2009), 568 – 579. https://doi.org/10.1016/j.appet.2009.01.009

[55] Lav R Varshney, Florian Pinel, Kush R Varshney, Debarun Bhattacharjya, Angela Schörgendorfer, and Y-M Chee. 2019. A big data approach to computational creativity: The curious case of Chef Watson. *IBM Journal of Research and Development* 63, 1 (2019), 7–1.

[56] J Wardle, K Parmenter, and J Waller. 2000. Nutrition knowledge and food intake. *Appetite* 34, 3 (Jun 2000), 269–275. https://doi.org/10.1006/appe.1999.0311