



UvA-DARE (Digital Academic Repository)

“Tell me why”: Using natural language justifications in a recipe recommender system to support healthier food choices

Starke, A.D.; Musto, C.; Rapp, A.; Semeraro, G.; Trattner, C.

DOI

[10.1007/s11257-023-09377-8](https://doi.org/10.1007/s11257-023-09377-8)

Publication date

2024

Document Version

Final published version

Published in

User Modeling and user-adapted Interaction

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Starke, A. D., Musto, C., Rapp, A., Semeraro, G., & Trattner, C. (2024). “Tell me why”: Using natural language justifications in a recipe recommender system to support healthier food choices. *User Modeling and user-adapted Interaction*, 34(2), 407–440. <https://doi.org/10.1007/s11257-023-09377-8>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)



“Tell Me Why”: using natural language justifications in a recipe recommender system to support healthier food choices

Alain D. Starke^{1,2} · Cataldo Musto³ · Amon Rapp⁴ · Giovanni Semeraro³ · Christoph Trattner²

Received: 2 June 2022 / Accepted in revised form: 21 July 2023 / Published online: 24 October 2023
© The Author(s) 2023

Abstract

Users of online recipe websites tend to prefer unhealthy foods. Their popularity undermines the healthiness of traditional food recommender systems, as many users lack nutritional knowledge to make informed food decisions. Moreover, the presented information is often unrelated to nutrition or difficult to understand. To alleviate this, we present a methodology to generate natural language justifications that emphasize the nutritional content, health risks, or benefits of recommended recipes. Our framework takes a user and two recipes as input and produces an automatically generated natural language justification as output, based on the user’s characteristics and the recipes’ features, following a knowledge-based recommendation approach. We evaluated our methodology in two crowdsourcing studies. In Study 1 ($N = 502$), we compared user food choices for two personalized recommendation approaches, based on either a (1) single-style justification or (2) comparative justification was shown, using a no

✉ Alain D. Starke
a.d.starke@uva.nl

Cataldo Musto
cataldo.musto@uniba.it

Amon Rapp
amon.rapp@unito.it

Giovanni Semeraro
giovanni.semeraro@uniba.it

Christoph Trattner
christoph.trattner@uib.no

¹ Amsterdam School of Communication Research, University of Amsterdam, P.O. Box 15791, 1001 NG Amsterdam, The Netherlands

² SFI MediaFutures & DARS Research Group, University of Bergen, Lars Hilles gate 30, 5008 Bergen, Norway

³ Swap Research Group, University of Bari Aldo Moro, Piazza Umberto I 1, 70121 Bari, Italy

⁴ Department of Computer Science, University of Torino, C. so Svizzera 185, 10149 Torino, Italy

justification baseline. The recommendations were either popularity-based or health-aware, the latter based on the health and nutritional needs of the user. We found that comparative justification styles were effective in supporting choices for our health-aware recommendations, confirming the impact of our methodology on food choices. In Study 2 ($N = 504$), we used the same methodology to compare the effectiveness of eight different comparative justification strategies. We presented pairs of recipes twice to users: once *without* and once *with* a pairwise justification. Results indicated that justifications led to significantly healthier choices for first course meals, while strategies that compared food features and emphasized health risks, benefits, and a user's lifestyle were most effective, catering to health-related choice motivations.

Keywords Recommender systems · Natural language processing · Food · Explanations · User study · Health

1 Introduction

Food choices are the result of a context-dependent, multi-aspect process (Furst et al. 1996; Starke et al. 2021). While people's general food preferences in part determine short-term decisions (Köster 2009), a significant part of our eating habits is strongly influenced by contextual factors (Cadario and Chandon 2020). Many decisions are made at the point of purchase (Bialkova and van Trijp 2011; Bialkova et al. 2014). For example, foods presented at eye-level sight in supermarkets are more likely to be purchased (Kroese et al. 2016), just like food products with visually attractive packaging (Bialkova et al. 2014; Cadario and Chandon 2020). Such food decisions are often made routinely (Kalnikaitė et al. 2013) and are based on heuristics and so-called System 1 thinking (Kahneman 2011), rather than longer-term contemplation.

Online food choices are typically made in the context of information-filtering and retrieval systems (Starke et al. 2021). Food recommender studies have examined different approaches to cater toward a user's appetite (Freyne and Berkovsky 2010; Trattner and Elswailer 2019; Elswailer et al. 2022), but have paid little attention to how users can be supported to nourish themselves more healthily, despite evidence that commonly recommended popular internet-sourced recipes tend to be unhealthy (Trattner and Elswailer 2017). Consumers tend to be overwhelmed with information when making a decision (Kalnikaitė et al. 2013), which cannot be alleviated by changing the recommended context. At best, studies have considered specific dietary constraints (e.g., allergies) and nutrient intake to generate healthier recommendations (Schäfer et al. 2017; Schäfer and Willemsen 2019), or have leveraged human biases to steer user preferences toward specific recipes, for example by using visually attractive images (Elswailer et al. 2017; Starke et al. 2021).

In the recommender context, we argue that users can be supported to make healthier choices by using justifications concerning why a set of recommendations is presented. Specifically for the food domain, justifications of recommendations that elaborate on the nutritional content of different recipes can steer user choices away from the common popularity-based recommendations (Musto et al. 2020). An open question is to what extent justifications can affect user preferences if items are already personalized

(Starke et al. 2020), as well as whether user preferences can be affected if that user has made prior choices.

In this paper, we present an approach inspired by knowledge-based natural language generation strategies (Reiter and Dale 2000), to produce justifications for different recipe recommendations. Recent developments in natural language justification strategies show their merit in improving the transparency of the recommendation process, increasing users' trust and affecting their decision-making processes (Tintarev and Masthoff 2012; Nunes and Jannach 2017). The proposed framework 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. It draws upon general knowledge about health risks and benefits related to food consumption to generate justifications. Within the framework, eight different *justification strategies* are implemented through two different *justification styles*, based on the combination of different informative content and features. In particular, we generate *comparative* justifications of recommendations, which juxtapose the main characteristics of two recipes into a single natural language sentence. For instance, such a justification could compare the fiber content of two recipes. This taps into consumer research on the effectiveness of comparative evaluations of item attributes (Bettman et al. 1998), compared to a separate representation of that information (i.e., a "Single" justification).

We evaluate the effectiveness of the eight implemented justification strategies and two justification styles to support healthy food choices. We examine this across two different studies, asking users in each study to choose between popularity-based and health-based recommendations. In the first study ($N = 502$), we examine which natural language justification *style* is most effective in steering users toward healthier recipe choices. Building upon preliminary findings, in the second study ($N = 504$), we examine which natural language justification *strategy* is most effective in promoting healthy recipe choices. In doing so, we adopt a strict baseline, where we first present a recommendation pair to users with no justification, immediately followed by the same pair but accompanied by one of our eight justification strategies. Such preference or choice reversal is hard to achieve (Zhu et al. 2012), as people tend to stick to the status-quo when making a decision (Kahneman 2011). Finally, in both we inquire on *why* users have chosen either the healthy or popular recipe, as a user's motivation could help us to understand how to design better justifications (Tintarev and Masthoff 2012). We posit the following research questions:

[RQ1]: Which natural language justification *styles* are most effective in steering user preferences toward healthier recipes, and for which types of meals?

[RQ2]: Which natural language justification *strategies* are most effective in steering user preferences toward healthier recipes, and for which types of meals?

[RQ3]: To what extent can users' self-reported motivation predict healthy recipe choices?

As we will show in the following, it emerged that users preferred healthier recipes over popularity-based recommendations if comparative-style justifications are presented, as well as for specific types of justification strategies.

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. Moreover, we examine (iv) which justification strategies are most effective in affecting user choices.

2 Related work

The idea of providing intelligent information systems with *explanation* facilities has been studied since the early 1990s (Johnson and Johnson 1993). It was introduced in the area of recommender systems in the 2000s (Herlocker and Konstan 2001), only re-gaining attention due to the recent *General Data Protection Regulations (GDPR)*, which prescribed to increase the transparency of underlying algorithms. This particularly applies to recommender systems, since explanation strategies have shown to positively affect both a user's acceptance of and trust in presented recommendations (Sinha and Swearingin 2002; Cramer et al. 2008).

Explanations in recommender systems can have different aims (Tintarev and Masthoff 2012). For example, explanations can educate users or improve the efficiency of decision-making (Jannach et al. 2010; Tintarev and Masthoff 2011). For our current work, we identify *persuasiveness* as the main aim (Symeonidis et al. 2008). We specifically aim to promote healthy food choices through our justifications, which is novel to food recommender research (Trattner and Elsweiler 2019). The persuasive explanation aim is touted in other domains as useful to convince users to try or buy a recommended item, such as a product on Amazon or a movie on Netflix (Gkika and Lekakos 2014; Tintarev and Masthoff 2012).

With respect to the *information content* exploited to generate justifications, we frame our approach as being at the intersection between *content-based* and *knowledge-based* methods (Jannach et al. 2010). It is based on user characteristics and food features, along with general knowledge on food consumption. Taken together, they justify our health-aware recommendation by emphasizing health risks and benefits. This is related to studies where health risks are highlighted in a smoking cessation application in a recommender context (Hors-Fraile et al. 2016, 2022); although no evidence is provided concerning the effectiveness of such information (Hors-Fraile et al. 2016), a knowledge-based health recommender did lead to better results than a hybrid recommender (Hors-Fraile et al. 2022). Conversely, our work fills this knowledge gap by evaluating the impact of justifications, including health risks and benefits, on user food choices.

The effectiveness of justifications can be better understood by cognitive processing and decision-making theories. For one, dual-process theory emphasizes that people's behavior is determined by two diverse processes or systems: a non-conscious process that relates to spontaneous, heuristic-based thinking (i.e., "System 1"), and a reflective process that relies on rational and conscious decision-making (i.e., "System 2") (Hagger 2016; Kahneman 2011). This duality in cognition is also described by the Elaboration Likelihood Model (Petty et al. 1997), which is an information processing theory of persuasion that describes changes in a person's attitude as the result of two diverse "routes." Under the *central route*, the recipient of the persuasion attempt

(e.g., the user) is thinking rationally about the message, drawing upon prior experience and knowledge to carefully evaluate all of the information presented. In contrast, the *peripheral route* of persuasion relies on simple cues and heuristics to judge the relevance and validity of a persuasive message.

Fast decisions without much deliberation seem to be common in low-stake recommender domains, such as movies (Gomez-Uribe and Hunt 2015). These choices are typically the result of a simple association or inference process without much cognitive effort (Petty et al. 1997), activating the peripheral route. Affect is associated with peripheral activation in food choices, as certain emotions tend to be associated to specific foods (Gutjar et al. 2015), and foods are chosen based on their visual appeal (Elseweiler et al. 2017; Starke et al. 2021). Although peripheral activation is likely when an explanation is absent (e.g., when only showing images and ingredients), we argue that providing a justification *why* specific recipes are presented would increase the likelihood of activating the central rather than the peripheral route. In our study, it can be considered as a cognitively oriented healthy eating nudge (Cadario and Chandon 2020), making users reflect about the contents of recipes.

Another hallmark of the current work lies in the development of a justification framework, designed specifically for the *food* domain. As discussed in Tran et al. (2018), studies that evaluate the impact of explanations and justifications 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 in Leipold et al. (2018), where a very simple explanation strategy based on food features is integrated with a food recommender system, but the impact on food choices is not evaluated. Another simple explanation interface is presented in Elahi et al. (2014), 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 Elahi et al. (2014), 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 presented (e.g., separately or comparatively) affects user preferences (Bettman et al. 1998). A remotely similar approach is adopted by Chen and Wang (2017), who introduce a user interface where different recommendations are presented together with their distinctive features, obtained automatically from user reviews. However, in contrast with Chen and Wang (2017), 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 Friedrich and Zanker (2011), labelling it as a *black box* methodology. Hence, the explanation strategy is not aware of the underlying recommendation model, generating a *post hoc* explanation that is independent of the recommender algorithm. Post hoc explanations provide reliable and effective explanations that are typically preferred by final users (Musto et al. 2019, 2020). We evaluate this framework by implementing two food recommender approaches: one that identifies *popular*

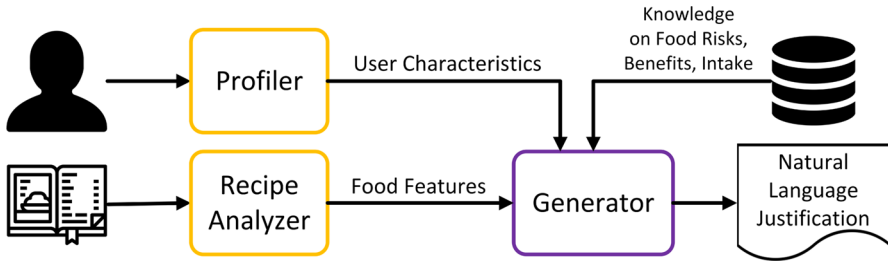


Fig. 1 Schematic workflow to generate natural language justifications, based on user and recipe features and food knowledge, to be incorporated in food recommendations

recipes and one that selects *healthier* recipes. More details about the algorithms will be provided in the methodology.

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 and Cotton (2017): 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.

3 Methodology

3.1 Natural language justification workflow

Figure 1 depicts the workflow to generate natural language justifications. It shows three main components.

The PROFILER module collects user’s characteristics. It adopts a holistic user profiling approach used in other studies (Cena et al. 2018a, b, 2020), including one on taste-based food recommendations and health-related scenarios (Polignano et al. 2020). holistic user models (HUMs) (Musto et al. 2020, 2021) rely on the intuition of modeling a profile of the user by combining heterogeneous data points and mapping them to a set of facets that describe the user. These facets include *affect* (e.g., a user’s current mood), *contextual constraints* (e.g., time and willingness to pay for meals), *demographics* (i.e., age, gender), *self-reported health data* (e.g., BMI, lifestyle self-evaluation, stress), and *weight-related goals*. Table 1 outlines the seven user aspects used, which are encoded in each user profile. Note that preferences were also inquired upon by asking about favorite ingredients, assuming that this was both related to the overall preferences and specific taste-related preferences.

In a similar vein, the RECIPE ANALYZER extracts the main food features of the recommended recipes (e.g., ingredients, nutrients). These include the nutritional content of food, expressed in nutrients (i.e., fats, fibers, proteins), calorie content, and a Food Standards Agency (FSA) recipe health score. The FSA score is an aggregate health

Table 1 User characteristics obtained by the PROFILER module in our natural language justification workflow

User aspect	Factors
<i>Affect</i>	Mood (positive, negative, neutral)
<i>Behavioral data</i>	Level of physical activity
<i>Constraints</i>	Cooking time, cost constraints
<i>Demographics</i>	Gender, age, height, weight
<i>Domain knowledge</i>	Cooking experience, available time, cost constraints
<i>Goals</i>	Losing weight (binary)
<i>Health data</i>	Lifestyle, BMI, amount of sleep, stress
<i>Preferences</i>	Food preferences and restrictions (lactose-free, vegan, etc.)

score that captures the nutritional content of a recipe, based on the serving weight and the weight per 100 g of nutrients: sugar, fat, saturated fat, and salt (Howard et al. 2012; Trattner and Elswailer 2017; Starke et al. 2021). In addition, the recipe analyzer also extracts contextual features of the recipes, such as cooking time and preparation difficulty. All these data are crawled from online sources (e.g., recipe web-sites, such as GialloZafferano¹) and publicly available knowledge bases.

Finally, the GENERATOR outputs the justification, also based on knowledge about health-related food risks and benefits. The final output comprises eight different *justifications strategies*, each emphasizing different recipe characteristics or user features. The generation process follows the principles of Natural Language Generation systems (Reiter and Dale 2000), completely *automated* and *unsupervised*, thus not requiring any human intervention.

On the basis of this setting, our framework generates its output by following two different justification *styles*: *single* and *comparative*. It takes as input two different recipes. On the one hand, by following the first justification style, both of them are processed separately and each recipe is provided with a different justification. On the other hand, a comparative justification contrasts the characteristics of the two recipes and is automatically generated by the algorithm.

To generate justifications, the GENERATOR module also relies on general *food knowledge*. It uses a food knowledge base that comprises *facts* related to the daily intakes of nutrients, as well as food consumption benefits and risks. Such knowledge relies on general guidelines concerning food consumption, such as government publications, academic studies, and commonsense knowledge. In particular, for each of the nutrients—sugar, carbohydrates, fats, proteins, fibers—around ten facts are encoded. For instance, “Consuming too much sugar increases the risk of diabetes,” “High protein intake improves muscle development,” and “High sodium intake increases health pressure.” In total, we have encoded around 150 facts in our knowledge base, which are used in several justification strategies.

¹ <https://www.giallozafferano.com/>.

3.2 Overview of the justification strategies

We defined *eight* different justification strategies. These are outlined in Table 2 along with the relevant characteristics and features. To define and select the justification strategies, we used two criteria:

(1) The set of justification strategies should elicit mainly (i) the central route of persuasion, i.e., encourage the user to reflect on her food choices, thinking rationally about the information provided; or (ii) both the central route and the peripheral route, i.e., based on cues aimed at activating non-conscious processes; or, to a lesser extent, (iii) the peripheral route of persuasion. In this way, we could compare different forms of persuasion, attempting to understand their effectiveness. We privileged the central route because we mainly embrace a cognitively oriented healthy eating nudge approach, which encourages users to reflect on their food choices. However, defining also strategies leveraging the peripheral route could give us insights on how a justification, which in principle should act on the conscious level of persuasion by providing information on the target behavior, could be combined with “nudges” that elicit unconscious processes.

(2) The formulation of the single justification strategies should tackle specific factors that, either consciously (via the central route) or unconsciously (via the peripheral route), may possibly affect behavior change, as pinpointed by behavior change theory. To this aim, we relied on five widely accepted theories of behavior change: The Health Belief Model (HBM) (Rosenstock 1974; Taylor et al. 2007) and the Theory of Planned Behavior (TPB) (Ajzen 1991) which pinpoint the role of attitudes and beliefs in driving human actions; the goal-setting theory, which shows that people make decisions and take action in line with their set goal (Locke and Latham 2002); the social cognitive theory (SCT) (Bandura 1986), which posits that behavior is affected by e.g., efficacy expectations (or self-efficacy) and the behavior of others; and the Transtheoretical Model of behavior change (TTM) (Prochaska and Velicer 1997), which describes change as a six-stage process through which an individual progresses. We chose these theoretical frameworks because they are the most widely used theoretical frameworks in technology-based interventions for behavior change (Orji and Moffatt 2018; Pinder et al. 2018; Rapp et al. 2019; DiSalvo et al. 2010; Stowell et al. 2018).

The strategies exploit different information sources and follow a pre-set structure that is filled in dynamically, based on the workflow components depicted in Fig. 1. The text outputs from the PROFILER, RECIPE ANALYZER, and food knowledge components are concatenated using adverbs and conjunctions by the GENERATOR.

While most justification strategies in Table 2 put emphasis on health, some variety is included. The *Description* strategy contrasts both recipes neutrally, providing context on a recipe’s origin. The *Popularity* strategy is based on social cognitive theory which highlights that people may imitate the behavior of others and choices that appear to be popular in order to be accepted by others (Bandura 1986). The strategy contrasts each recipe’s popularity score on the food community platform GialloZafferano², where they were initially uploaded. This strategy prioritizes the popularity-based recommendation over the healthy recommendation, in part encouraging peripheral processes of

² <https://www.giallozafferano.it/>.

Table 2 Overview of the eight comparative justification strategies used in our experimental evaluation

Just. Strategy	Information source		Food features	Example
	User features	Comparative justification		
<i>Description</i>	None	General Information	Seafood risotto is a classic first dish of Italian cuisine, perfect for special occasions and for all seasons! Chickpea soup is a very simple and tasty first course. A poor farmer's recipe, that is prepared in very few steps!	
<i>Popularity</i>	None	Popularity	Seafood Risotto is more popular than Chickpea Soup in the community	
<i>Food features</i>	Preferences, restrictions	Ingredients, nutrients	Seafood Risotto has a higher amount of proteins (22.7 g vs. 18.2 g), and a lower amount of fibers (1.2 g vs. 12.4 g) than chickpea soup	
<i>Food goals</i>	Dietary goals	Calories	Seafood Risotto has fewer calories than Chickpea Soup (308 vs. 356).	
<i>Health benefits</i>	BMI, Mood Phys., Activity, Sleep, Stress	Ingredients, nutritional information	Seafood Risotto is better to reach your goal of losing weight Seafood Risotto has a higher amount of proteins (22.7 g vs. 18.2 g), and a lower amount of fibers (1.2 g vs. 12.4 g) than chickpea soup. The intake of many proteins reduces hunger. Given your current weight, this can be helpful. However, the intake of fiber reduces cholesterol	
<i>Health risks</i>	BMI, mood, sleep, stress, physical activity	Ingredients, nutritional information	Seafood Risotto has a higher amount of proteins (22.7 g vs. 18.2 g), and a lower amount of fibers (1.2 g vs. 12.4 g) than chickpea soup. The intake of too many proteins can lead to constipation and dehydration	
<i>User Lifestyle</i>	Personal lifestyle	FSA Health Score	According to the FSA score, chickpea soup is healthier than Seafood Risotto. Please consider this, given the importance you attributed to a healthy lifestyle	
<i>User skills</i>	Cooking experience	Level of difficulty	Chickpea soup is easier to prepare than Seafood Risotto. It should be more adequate to your cooking skills, which are low	

While the *Description* and *Popularity* strategies did not incorporate user features, all others did and were designed to promote healthy recipe choices. In our example, the seafood risotto represents the popular recipe, while the chickpea soup represents the healthy recommendation

persuasion by creating a majority or bandwagon effect (Elsweiler et al. 2017), where the pressure of "peers" may act unconsciously, as also employed by Starke et al. (2020); Zhu et al. (2012).

The strategies related to the recipe's *Food Features* and the user's *Food Goals* support central route persuasion processes.

The *Food Features* strategy is based on the TTM, which notices that consciousness raising, that is the increasing of knowledge about aspects related to the behavior be changed, may encourage people progress toward behavior change (Prochaska and Velicer 1997). The strategy informs users about specific nutrients of both recipes, aiming to overcome poor nutrient intake and low food knowledge levels (Ilich et al. 1999; Wardle et al. 2000), based on a neutral lexicalization of the characteristics, such as "X contains more proteins and fats than Y, but fewer carbohydrates."

The *Food Goals* strategy relies on goal-setting theory (Locke and Latham 2002), which shows that people make decisions and take action in line with their set goal: reminding people of these goals is particularly effective if the goals are important to them and are self-set rather than assigned to them Munson and Consolvo (2012). Accordingly, Table 2 shows how nutritional food features per recipe are linked to a user's self-set goals, contrasting them ('X has more calories than Y'), and highlighting the recipe with fewer calories if a user pursues weight loss goals.

Two other justification strategies are based on the HBM and aim to highlight *Health Benefits* and *Health Risks*. HBM points out that health-related behaviors and choices are affected by: (i) the perceived susceptibility to illness or health problems and the perceived severity of the consequences associated with the state or condition, (ii) the perceived benefits of a health behavior (Rosenstock 1974; Taylor et al. 2007). Both justification strategies link nutrient intake information to health benefits or risks, which is split in three parts: (i) macro-nutrient selection, (ii) retrieving nutrient-specific food knowledge, highlighting either health benefits or risks, (iii) connecting relevant user characteristics to the nutrient-specific knowledge. For example, if the user reported to be overweight, the justification could highlight a risk related to heart diseases. Both pairwise strategies contrast the different levels of nutrients in two different sentences, each linking food characteristics to health benefits or risks, aiming to elicit emotional as well as reflective responses, activating both central and peripheral route processes.

The two final justification strategies are based on the user's self-reported lifestyle and skills and are aimed at eliciting the central route.

The *User Lifestyle* strategy relies on the Theory of Planned behavior, which states that human behavior is a consequence of one's behavioral intention, which is in turn explained by e.g., one's attitude and subjective norm (Jun et al. 2014). Attitudes may in turn be affected by values (Ajzen and Fishbein 2005; Ateş 2020) While a value may be defined as a desirable and fundamental standard that guides people's actions (Jun et al. 2014), health value is "the degree to which individuals value their health" (Tudoran et al. 2009). In the food domain, it has been shown that people's perceived health values positively affect their choices and actions toward low-fat or low-calories menu items (Jun et al. 2014). The strategy connects the comparative nutritional evaluation of both recipes (in the form of an FSA health score (Trattner and Elsweiler 2017)) to a user's personal values, such as the importance of maintaining a healthy lifestyle. The

value-attitude-behavior model explains that both values and attitudes affect behavior (Jun et al. 2014; Tudoran et al. 2009).

In a similar vein, the *User Skills* strategy is grounded in social cognitive theory and, in particular, in the construct of self-efficacy, which captures the belief in one’s capabilities to execute a course of action (Bandura 1997). People who report higher levels of self-efficacy tend to execute more difficult tasks (Elsweiler et al. 2017; Bandura 1986), because they are more confident that they will successfully execute the task; conversely, people with low self-efficacy may select less difficult activities and give up the accomplishment of difficult tasks (Zimmerman 2000; Zulkosky 2009; Schunk 1996). Bandura (Elsweiler et al. 2017; Bandura 1986) hypothesized that self-efficacy impacts on choice of activities, effort, and persistence. In our study, we link the user’s self-reported *cooking experience* to each recipe’s “*level of difficulty*.”

3.3 Food recommendation algorithms and dataset

For our experimental evaluation, which spans across two studies, we use two personalized food algorithms to retrieve recipes. The first personalized algorithm optimizes for a recipe’s health, which is referred to as the *Healthy* algorithm or health-aware algorithm. Healthy recipes are retrieved based on a variety of user characteristics, such as food goals and dietary constraints (Musto et al. 2020). The second algorithm retrieves popular recipes, based on given website ratings stored in the dataset, and is thus referred to as the *Popular* algorithm. Since our natural language justification framework is decoupled from both algorithms, we consider them as independent parameters in our experimental manipulation.

The recipes used for our NLP framework were sampled from a database of 4,671 Mediterranean-style recipes. The used dataset is available online, along with processing scripts.³ The recipes have been obtained from the popular food community platform GialloZafferano and translated to English. The recipes contain information about their name, category, preparation difficulty, as well as their ingredients, (macro-)nutrients, calories, rating count, and average website rating. Moreover, they also include several binary tags, such as *vegetarian*, *vegan*, *lactose-free*, and *low-nickel*.

4 Study 1: Examining the effectiveness of different justification styles

4.1 Method

In Study 1, we examined the merits of our natural language framework. We investigated the effectiveness of different justification styles (RQ1), comparing user choices for either the healthy or popular recipe recommendation across trials with no justification, a single-style justification, or a pairwise justification. We did so across three meal types, using eight different justification strategies throughout, exploring [RQ2] as well.

³ <https://osf.io/hn3et/>.

Rice soup with leeks and potatoes

Click to show ingredients

Spaghetti with burrata cream and eggplant

Click to show ingredients

Rice soup with leeks and potatoes is as easy to prepare as Spaghetti with burrata cream and eggplant.

Given the current information, which recipe you choose? Left side recipe

Why did you choose this recipe?

Remember: 1 Star means completely disagree, 5 Stars mean completely agree

It matches my food tastes and preferences	★	★	★	★	★
It seems savory and tastier	★	★	★	★	★
It helps me to eat more healthily	★	★	★	★	★
It would help me to lose/gain weight	★	★	★	★	★
It seems easier to prepare	★	★	★	★	★

Fig. 2 The study's interface for two first course meals. The recipe displayed on the left is our healthy-algorithm recommendation, the one displayed 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

4.1.1 Participants

In total, we analyzed a sample of 502 US-based participants (43.8% Male) in an experimental evaluation.⁴ They were recruited through Amazon MTurk, being required to have a hit rate of 98% and a minimum of 500 approved hits.⁵ Participants were required to be fluent in English. Most of the participants were employed (81.3%; 2.6% was student) and between 30 to 40 years (37.1%), whereas only 15.7% was between 20 and 30 years and 17.9% was between 40 and 50 years. More than 55% of the participants declared that they had a weight loss goal, whereas only (9.1%) had a weight gain goal. The majority of the participants completed the provided tasks between 5 and 10 min. They were reimbursed with 0.5 USD.

⁴ Although 504 participants had completed the study, two were dropped due to multiple missing values on choice variables.

⁵ Participants with such a hit rate likely generate high-quality data and meet attention checks.

4.1.2 Procedure

First, the participants were asked questions about demographics, health and well-being, dietary restrictions, food preferences, and experience with home cooking, which were needed to model their profile (see Table 1 for an overview of the feature of the model). Then, the profiler (cf. Fig. 1) generated three pairs of recommendations (see an example in Fig. 2, where the left recommendation is based on our *healthy* food recommender, whereas that on the right is generated using a *popularity-based* algorithm), which were presented sequentially to the participant: first, two first course meals, then, two second courses and, finally, two desserts. For each pair, participants were required to choose either i) the left-hand side recipe, ii) recipe the right-hand side recipe, iii) or neither. The participants were not aware which recipe was the *healthy* recommendation, or if there was any at all. Participants who chose one of the two recipes were subsequently asked to indicate the reason behind their choice, whether it was based on factors such as the recipe's taste, healthiness, or ease of preparation.

4.1.3 Research design

To examine whether healthy recipe choices could be supported with different justification styles (RQ1), we designed three between-subject conditions. The participants were either presented no justification for the prompted recipes (i.e., the baseline), a justification style focusing on each recipe separately (i.e., "Single Justification"), or a justification style comparing the two recipes (i.e., "Comparative Justification"). Moreover, to explore the merits of different justification strategies (RQ2) the conditions in which single-style or comparative justifications were presented, were subject to eight within-subject conditions (see Table 2). This way, one participant could be presented three different single justifications (e.g., popularity, food goals, and health risks), while another participant would be prompted three other comparative justifications (e.g., User Lifestyle, Food Features, Health Benefits), or no explanation at all for each recipe. Figure 2 provides an example of a "User Skills" justification, displayed within the red box.

4.1.4 Measures

To address [RQ1], we considered the effect of different justification styles on the percentage of healthy recommendations chosen by the participants. To this aim, we compared the "No Justification" baseline either with any justification style separately, that is "Single" and "Comparative" justifications, or across the different justification strategies listed in Table 2. The effectiveness of each justification style was compared 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). To address [RQ2], Different justification strategies were compared between the no explanation baseline and the comparative style, as the results showed that the comparative style was the most effective justification style.

Moreover, to address [RQ3], we examined participants' motivations for choosing one of the two presented recipes. The participants were required to indicate on 5-point

scales to what extent a certain motivation was applicable, as well as to report the reason why they had chosen one of the recipe. Motivation items are depicted in Fig. 2, and were related to a match with the user's preferences, weight loss or gain goals, healthy eating goals, the recipe's taste, and a recipe's ease to prepare. The user preferences herein were related to the overall evaluation of the recommendations, while other motivations related to specific aspects (e.g., a recipe's taste).

Finally, we discuss the set of user characteristics that users were asked to disclose. These measures were employed by the Profiler to produce healthy recommendations (see Table 1). Besides obtaining data on food preferences and demographics (i.e., age, gender, BMI), we asked users to report whether they had any food goal (i.e., weight loss, weight gain, or no goals), and to rate the healthiness or their lifestyle and the importance for them of having such a lifestyle (5-point scales). The participants were also required to rate how frequently (5-point scale) they make healthy food choices, use websites with recipes, look at the nutritional values of food, and engage in home cooking. Furthermore, the participants were asked about their current levels of sleep, physical activity, and mood (3-point scales), and whether they were depressed or stressed ("yes" or "no"). Finally, we asked them about their food knowledge, as they had to indicate their cooking experience (5-point scale) and cost and time constraints for cooking.

4.1.5 Manipulation check

We checked whether the health-aware recommendations could actually be considered as healthier than the popular recommendations. We assessed recipe healthiness through the "WHO Score," which was first used in a digital recipe context by Howard et al. (2012). It captured recommended daily intake levels for six nutrients and calories in a score between 0 and 7 (Organization 2003). We confirmed that the health-aware recommendations yielded higher WHO scores for each meal type than the popular recommendations: For first courses (health-aware: 4.21; popular: 2.30), second courses (health-aware: 2.65; popular: 1.61), and desserts (health-aware: 2.94; popular: 1.66). The only nutrient for which the popular recommendations were slightly healthier than the health-aware ones was sugar, as the popular recipes tended to be high in fat and saturated fat but somewhat lower in sugar.

4.2 Results

We examined user choices through three different analyses.⁶ We did so in three ways. First, we examined whether presenting *any* explanation, through two different styles, affected user preferences for healthy recommendations. Second, we examined preferences for each of our eight justification strategies. Third, we investigated more specifically why users had either chosen healthy or popular recipes.

⁶ The collected data for this study, as well as the analysis scripts can be obtained via <https://osf.io/vytdx/>.

4.2.1 Single and comparative justifications styles (RQ1)

We studied whether participants were more likely to choose healthier recipes if justifications were presented underneath it. We used a one-way ANOVA to examine choices made across all types of meals. A Shapiro-Wilk test for normality showed no evidence for non-normality of the dependent variable ($W = 1.00, p = 1.00$).⁷ The healthy recommendation was revealed to be chosen more often as long as any justification was presented underneath it (47.6% of choices, $SD = 0.50\%$), compared to the "No Justification" baseline ($M = 38.1\%, SD = 0.49\%$): $F(1, 1504) = 12.14, p < 0.001$. This suggested that justifications helped to steer user preferences toward the health-aware recommendation.

We further differentiated between the effects of presenting "Single" and "Comparative" styles. To do so, we performed a two-way ANOVA with two conditions dummies for "Single" and "Comparative" justification styles. Although users were not more likely to choose the healthy recommendation when being presented a "Single Justification" (43.0% of choices, $SD = 0.50\%, p = 0.13$), compared to the baseline (38.1%), they were more likely to do so when facing a "Comparative Justification" ($M = 51.1\%, SD = 0.50\%$): $F(1, 1503) = 18.24, p < 0.001$. This suggested that comparative justifications were particularly effective in supporting users choices for the healthy recommendation.

Further analyzes teased apart these effects by differentiating across the three meal types, as this would be consistent with previous research indicating that preferences differed across meal types (Musto et al. 2020). Using multiple one-way ANOVAs, we found that depicting any justification increased the number of choices for healthy recommendations for first courses ($F(1, 500) = 4.83, p < 0.05$) and desserts ($F(1, 500) = 4.43, p < 0.05$), but found no such effect for second course meals ($F(1, 500) = 3.03, p = 0.08$).⁸ We further inspected these effects by discerning between "Single" and "Comparative" justification styles per meal type, performing multiple two-way ANOVAs. This revealed that while "Single" justifications did not significantly boost healthy recommendation choices in any dish type (all p -values > 0.1), "Comparative" justifications did do so: for first courses ($F(1, 499) = 5.37, p < 0.05$), second courses ($F(1, 499) = 6.33, p < 0.05$), and desserts ($F(1, 499) = 6.61, p < 0.05$). This gave us further evidence that justifications comparing popular and healthy recommendations were more effective in steering participants' preferences toward healthy recommendations, than separate justifications per recipe.

To understand the results from the different ANOVAs, please refer to Fig. 3. Illustrated are recipe choices per meal type (from left to right: first course, second course, dessert), for which we examined the percentage of the chosen options per meal type: neither recipe, the popularity-based recommendation, or the health-aware recommendation. For first course meals and desserts, it was clear that the "Single" justification

⁷ For all ANOVAs, a dichotomous dependent variable was analyzed. The use of different statistical methods, such as the Kruskal-Wallis equality-of-populations rank test or a logistic regression, led to similar or identical results regarding the p -values of the independent variables. Hence, we considered ANOVA an appropriate method of analysis.

⁸ Performing a Repeated Measures ANOVA that included "meal type" as a categorical variable did not affect the main effects of the explanation styles.

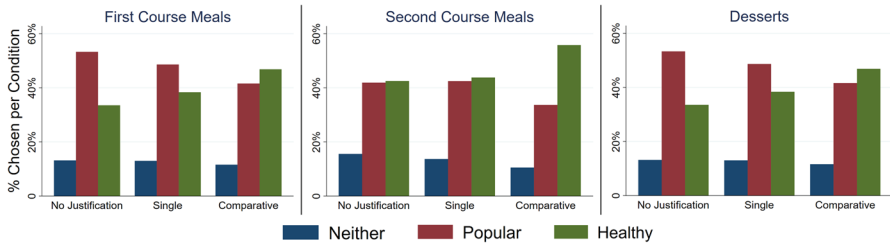


Fig. 3 Percentages of choices per condition, per meal type. Depicted are choices for neither recipe (in blue), the Popular recipe (in red), and the Healthy recommendation across three different meal types. Conditions are the three different justification styles: No justification, single justifications,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

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 choices for healthy recommendations.

4.2.2 Effectiveness of justification strategies (RQ2)

The previous subsection highlighted that pairwise justifications were the most effective in steering participants’ preferences toward healthy recommendation. Here, we examine the effectiveness of specific justification strategies (cf. Table 2) to promote our healthy recommendations.⁹

We examined the effectiveness across all meal types, as well as per separate type of meal. Table 3 outlines 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.

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).

⁹ 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.

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

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

The first model examines choices across all meal types ($N = 1071$), the other models concern meal type-specific analyses ($N = 357$ for each model). The denoted "Pseudo R^2 " is McFadden's pseudo R^2 (McFadden 1973): $1 - \frac{\log(Lc)}{\log(Lnull)}$, where "Lc" denotes the maximized likelihood value of the current model and "Lnull" of the baseline model. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

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 points out which strategies did not affect participants' preferences between the "Comparative" and "No Explanation" conditions. Both giving comparative descriptions of the contents of the recipe (e.g., the ingredients) and comparing whether the recipes match with the participant's lifestyle—for each meal type, did not affect participants' preferences. Furthermore, comparative justifications of food goals did not influence choices about desserts, whereas highlighting health benefits and risks did not affect choices about first course meals.

4.2.3 Choice motivation (RQ3)

Finally, we investigated *why* the participants had chosen one of the proposed recipes (RQ3). We performed four logistic regression analyses that compared cases in which either the popular or healthy recommendation was chosen, while ignoring cases in which neither recipe was chosen. Table 4 shows a model that includes a participant's choice motivation across all meal types, as well as three meal-specific models. Significant, positive effects in Table 4 indicate reasons why the healthy recommendation

Table 4 Four logistic regression models, each predicting user choices for the *Healthy Recommendation*

Choice motivation	All meal types β (S.E.)	First course β (S.E.)	Second course β (S.E.)	Dessert β (S.E.)
Ease of preparation	-.079 (.054)	-.26 (.099)**	-.030 (.098)	-.047 (.096)
Fitted food goals	.060 (.062)	-.0031 (.11)	.21 (.11)*	-.13 (.12)
Fitted Preferences	.14 (.070)*	-.052 (.13)	.52 (.13)***	-.20 (.12)
Health	.41 (.063)***	.78 (.12)***	.13 (.11)	.47 (.12)***
Taste	-.47 (.072)***	-.54 (.14)***	-.58 (.13)***	-.22 (.11)
Intercept	.037 (.32)	.52 (.62)	. - 52 (.55)	.62 (.56)
Pseudo R^2	.0628	.134	.0671	.0545

Models either included choices across all meal types ($N = 1339$), 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). R^2 is McFadden's pseudo- R^2 (McFadden 1973). *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

was chosen, while significant negative effects provided evidence as to why a popular recommendation was chosen. The best model fit was observed for the first course meal 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 ($\beta = .41$, $p < 0.001$), as well as for first course meals ($\beta = .78$, $p < 0.001$) and desserts ($\beta = .47$, $p < 0.001$). In contrast, tastiness was related to popular meal choices: averaged across meal types ($\beta = -.47$, $p < 0.001$), as well as for first course ($\beta = -.54$, $p < 0.001$) and second course meals ($\beta = -.58$, $p < 0.001$). 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 ($\beta = .13$, $p < 0.05$), in particular for second course meals ($\beta = .52$, $p < 0.001$). Second course healthy recipes were also chosen more often because a match in food goals: $\beta = 0.21$, $p < 0.05$. In contrast, easiness was negatively related to choosing healthy first course recommendations ($\beta = -.26$, $p < 0.01$), 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.

4.3 Conclusion

This study explored the effectiveness of different kinds of justifications aimed at explaining health-aware recommendations. With regard to justification styles (RQ1), the study results show that participants preferred popular recipes when no explanation is presented, whereas they preferred health-aware recommendations when a justification is paired with the suggestion. Among the different justification styles presented, we first discovered that comparative justifications are more effective in encouraging healthy choices than single justifications. This falls in line with previous research that

emphasizes that individuals tend to make comparative judgments rather than combining two independent observations (Köster 2009). Furthermore, we have explored the effectiveness of different justification strategies (RQ2), finding that comparing two recipes features and their related health risks better promotes healthy food choices. Finally, we have also shown what drives users' choices in selecting healthier recommendations (RQ3), and whether the reasons differ per meal type. For most meal types, we discovered that popularity-based choices are driven by taste motivations, while choices for our health-aware recommendations are tied to health-related reasons.

This said, the contrast between "No Justification" and "Justification" scenarios is usually evaluated in between-subject designs (i.e., A/B tests) or in a within-subject design across multiple, heterogeneous sets (Symeonidis et al. 2008; Tintarev and Masthoff 2012). In contrast, examining changing preferences for the same set of recommendations is uncommon (Ekstrand and Willemsen 2016; Starke 2019), for this is harder to measure. To date, only Zhu et al. (2012) examined whether a recommender could reverse user choices within a single study due to majority-based social explanations (e.g., "108 people prefer this one" vs "8 people prefer this one"). Users were first presented pairs of items without any explanation, after which later in the study the same pairs were presented again, but this time with social explanations. The explanation was presented alongside furniture products, baby photos, and other items from various domains. They found that 14.1% of the users switched toward the item with the majority norm if it was presented quickly after the first trial, while this percentage was higher (22.4%) if there was more time between trials. We follow this approach of preference reversal in Study 2.

5 Study 2: Investigating recipe choices for different justification strategies

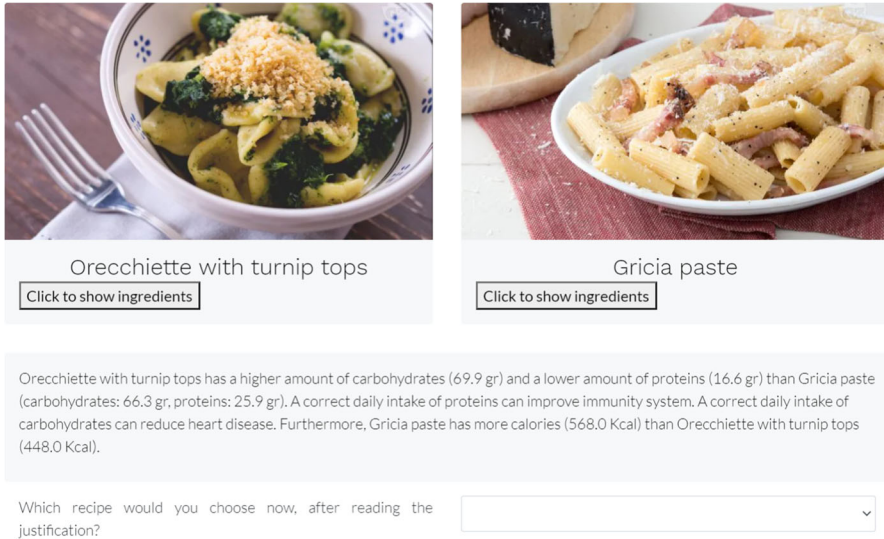
For Study 2, we considered a stricter study setup than in Study 1, following the work of Zhu et al. (2012). We examined whether back-to-back trials with and without justifications lead to choice reversal across a recommendation pair. In doing so, we assessed the effectiveness of eight different justification strategies across three different meal types. Note that all relevant processing scripts and datasets are available in our repository: <https://osf.io/hn3et/>.

5.1 Method

5.1.1 Participants

We invited users from the crowdsourcing platform Amazon Mechanical Turk to participate in a study on recipe recommendations and food enjoyment. Participants were required to be US-based and to have a hit rate of 98%, with a minimum of 500 approved hits,¹⁰ and were reimbursed with 0.5 USD. In total, 504 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

¹⁰ Such participants are more likely to generate high-quality data and to meet attention checks.



Orecchiette with turnip tops
Click to show ingredients

Gricia paste
Click to show ingredients

Orecchiette with turnip tops has a higher amount of carbohydrates (69.9 gr) and a lower amount of proteins (16.6 gr) than Gricia paste (carbohydrates: 66.3 gr, proteins: 25.9 gr). A correct daily intake of proteins can improve immunity system. A correct daily intake of carbohydrates can reduce heart disease. Furthermore, Gricia paste has more calories (568.0 Kcal) than Orecchiette with turnip tops (448.0 Kcal).

Which recipe would you choose now, after reading the justification?

Fig. 4 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. On the first trial, no justification is given but a list of ingredients per recipe. Depicted here is the second trial, presenting a pairwise “health benefits” justification underneath both recipes. Users were asked to choose one recipe or neither of them, and to provide reasons why they had chosen either recipe

loss goal (51.1%), while only 70 users (13.9%) had a weight gain goal. Participants were recruited throughout the USA, which may have varying levels of familiarity with Italian cuisine and a Mediterranean Diet (Lee et al. 2014).

5.1.2 Procedure

To provide personalized recipe recommendations, we first asked users to indicate their personal preferences regarding their eating habits and to disclose demographics. These included the different user features that were also used to generate the different justification strategies (cf. Table 2), including questions about a user's BMI, cooking experience (5-point scale), self-reported health (5-point scale), mood and well-being (3-point scales), as well as their dietary restrictions (e.g., no gluten or lactose) and general food preferences (i.e., input of ingredients a user liked).

Subsequently, we presented six pairs of recipe recommendations; one at a time. The PROFILER (cf. Figure 1) generated three recipe pairs based on a user's responses, which were each presented twice to a user. This included a pair of Mediterranean-style first course meals (Willett 2006), a pair of second course recipes, and a pair of desserts. Figure 4 shows an example set of first course recommendations, depicting the *healthy* recommendation on the left and the popularity-based recommendation on the right. Users were asked to choose the recipe they preferred the most, or neither of them. In addition, users were required to indicate on 5-point scales to what extent different reasons were underlying their choice, whether this was due to a recipe's ease of preparation, fit with user goals or preferences, health, or taste.

5.1.3 Research design

In line with Zhu et al. (2012), we presented each recipe pair twice to a user. While the first trial was presented with no justification, the second trial presented the same pair of recommendations *with* a pairwise justification. In doing so, we examine [RQ2], representing the peripheral route of the elaboration likelihood model by a recommendation scenario with no justifications. In contrast, decisions facing a pairwise justification require to interpret what is comparatively presented, encouraging the user to reflect on the information provided and, thus, eliciting central route processes. Hence, the current study juxtaposes these two scenarios, by initially asking users to choose a recipe from a pair of recommendations in the absence of any justification and, subsequently, re-visiting that choice when that same pair is presented again—accompanied by a justification. While the latter should take a more central route toward a user's elaboration, the justifications in the current study are situated on different points of the "peripheral-central continuum," supporting rational reflection to different degrees and also prompting information that elicits peripheral processes. Each justification strategy was randomly sampled from the eight strategies listed in Table 2.

5.2 Results

In the following, we examined [RQ2] and [RQ3]. We first reported the descriptive statistics of our "No Justification" baseline. Then, we examined how often users switched toward a different recipe when facing *any* justification strategy, before examining the effect of specific strategies (RQ2), and how different choice motivations related to healthy food choices (RQ3).

5.2.1 Baseline results and users switching to the healthy recommendation

To investigate whether justifications led users to swap their initial choices for the healthier recommendations (related to all research questions), we first examined user choices in the no justification baseline. Figure 5 depicts the distribution of recipe choices per meal type. For first course meals and desserts, the popular recipe was slightly favored, while the healthier recommendation was preferred for second course meals. Since popular recipes were typically preferred in other studies (Trattner and Elswiler 2017), this suggested that our health-aware recommendation pipeline was sufficiently personalized to the extent that many users already liked it—even without any justification.

By comparing Figs. 5 and 6, we examined whether user choices reversed for the same recipe pair after a justification was presented. By performing paired *t*-tests, we found that users were more likely to switch to the healthier recommendation when any justification was presented alongside first course meals, compared to no justification: $t(503) = -3.17$, $p < 0.01$. In contrast, we observed no differences in healthy recipe choices for second course meals: $t(503) = 0.24$, $p = 0.81$, nor for desserts: $t(503) = -0.24$, $p = 0.81$.

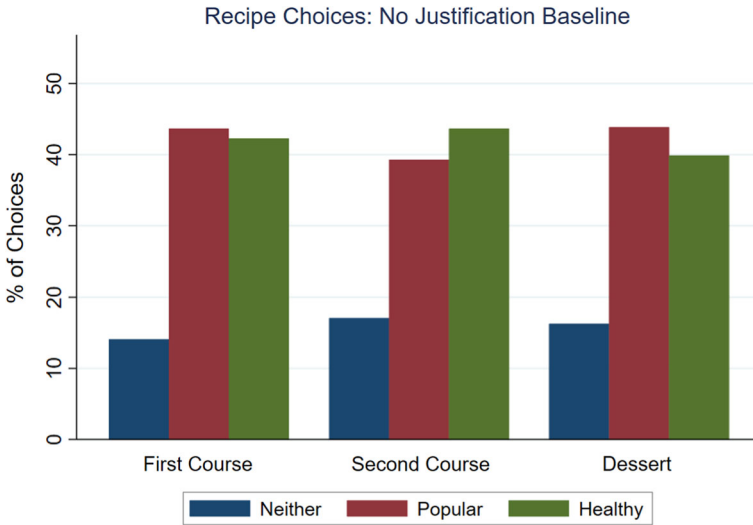


Fig. 5 Distribution of recipes chosen in the no justification baseline (i.e., the first choice made for a recipe pair), per meal type

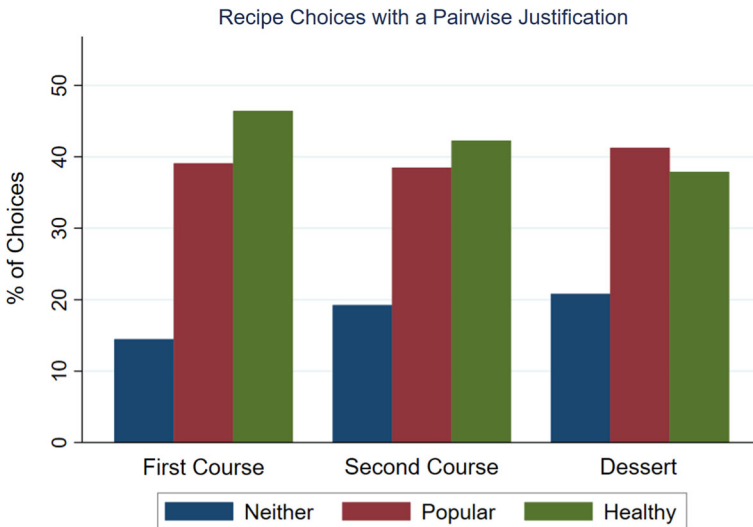


Fig. 6 Distribution of recipes chosen when a pairwise justification was presented (i.e., a recipe pair's second choice), per meal type

5.2.2 Specific justification strategies (RQ2)

We further investigated which justification strategies led users to reverse their choices toward the healthier recommendation (RQ2). We assessed whether the likelihood that a healthy recipe was chosen increased or decreased due to a specific justification strategy (i.e., reversing user choices), compared to the no justification baseline in the first trial.

Table 5 Random-effects logistic regression models (clustered at the user level), capturing different justification strategies that predict whether the healthy recipe is chosen from a recommendation pair

	First course β (S.E.)	Second course β (S.E.)	Dessert β (S.E.)
Description	.13 (.90)	-.10 (1.00)	-1.42 (1.06)
Popularity	-1.16 (1.06)	-.071 (.86)	-1.13 (1.04)
Food features	1.69 (.78)*	-1.58 (1.15)	1.12 (.96)
Food goals	1.30 (.93)	-.50 (1.00)	1.24 (.88)
Health benefits	2.21 (.73)**	.57 (.86)	-2.66 (.85)**
Health risks	3.25 (.94)**	-.21 (1.00)	.042 (.94)
User lifestyle	1.84 (.83)*	-1.65 (.86)	-.57 (.92)
User skills	.030 (.83)	-1.13 (1.00)	-1.79 (.87)*
Intercept	-4.86 (.40)***	-4.41 (.40)*	-5.03 (.39)***
Wald $\chi^2(8)$	28.92***	6.99	19.47*

Effects are relative to the effect in the "No Justification" baseline. Note that the Second Course model does not pass Wald's model test and can be disregarded. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

To this end, Table 5 reports three random-effects logistic regression models, one per meal type, of which the second course model is reported but disregarded, because it did not pass the Wald χ^2 test of model fit.¹¹

Table 5 shows that different justification strategies affected users' healthy choices for different meals. For first course meals, four different strategies increased the likelihood that a healthy recipe was chosen: a justification that described the features of both recipes ($\beta = 1.69, p < 0.05$), justifications that compared both recipe's nutrients and linked them to health benefits ($\beta = 2.21, p < 0.01$) and risks ($\beta = 3.25, p < 0.01$), and a justification on how a recipe could contribute to a user's lifestyle ($\beta = 1.84, p < 0.05$). This suggested that most of the justification strategies that highlighted nutritional aspects of recipes, and possibly linked these to user characteristics, were successful in reversing initial user choices and steering them toward healthier choices for first course meals.

Justifications were less successful in promoting healthy dessert choices. Table 5 shows that the strategies that affected the likelihood of healthy first course choices, did not do so for desserts. Instead, justification strategies on the recipes' health benefits ($\beta = -2.66, p < 0.01$) and preparation difficulty (i.e., user skills; $\beta = -1.79, p < 0.05$) decreased the likelihood that a healthy dessert was chosen. It seemed that our justification strategies were not appropriate for the dessert context, as users might have had more taste-related reasons for their choices, which was examined next.

¹¹ One reason that we did not observe any effects, nor could infer a significant model for the second course meals was the relatively large amount of healthy recipe choices in the No Justification baseline. This high proportion of healthy recipes was barely affected by any of the justification strategies.

5.2.3 Choice motivation (RQ3)

Finally, to contextualize our findings, we examined to what extent a user's motivation to choose the healthy recommendation changed after being presented *any* justification (RQ3). Table 6 describes six logistic regression models: three models that predicted healthy recipe choices *before* a justification was presented (denoted by β_{pre} ; one per meal type), and three models for *after* a justification was presented (denoted by β_{post}). Across all meal types, we found that health-related choice motivations positively affected the likelihood of healthy recipe choices “post-justification,” while this only applied to first course meals and desserts “pre-justification.” This suggested that our health-aware recommendations catered to users who were making health-motivated recipe choices, while the justification was important for second course meals. In contrast, none of the models showed a relation between preference-related, goal-related, and food characteristics-related motivation and healthy recipes choices, indicating that these motivations were not specifically linked to either recommendation.

Table 6 further suggests that addition of justifications seemed to put less emphasis on contextual factors. Whereas motivations related to taste (first course meals and desserts) and ease of preparation (first course) decreased the likelihood that a healthy recipe was chosen, these effects were no longer present “post-justification.” This suggested that the nutritional or health-related emphasis of most of our justifications was successful, arguably making users reflect on their initial food choice and tapping into the more central route of persuasion.

5.3 Conclusion

Study 2 analyzed users' changing preferences for the same set of recommendations provided, examining choice reversal in back-to-back trial with and without justification. We provided additional evidence for addressing our research questions, by evaluating the effectiveness of eight different justification strategies (RQ2), grounded in psychological literature, across three different meal types. The study results pointed out that pairwise justifications may encourage participants to reverse their choices toward healthier recipe recommendations, moving them away from popular recipes, but that this particularly applied to first course meals. Moreover, we discovered that different kinds of justifications may have different effect for different types of meals. Justification strategies tied to food features, health benefits and risks, and the participant's lifestyle are most effective with reference to first course meals. However, with reference to second course meals we found no effect, which might be due to the fact that this kind of meal was preferred by a large part of the participants in the pre-justification trial, leaving little room for improvement when introducing pairwise justifications.

With regard to the choice motivation of participants (RQ3), we found more evidence that users who are interested in health were more likely to choose the healthy recipe. This already applied to the pre-justification conditions for first course meals and desserts, but also post-justification for second course meals. In addition, we observed that other motivations that were present pre-justification, such as ease of preparation

Table 6 Six Logistic Regression models predicting healthy recipe choices using different choice motivations

	First course meals		Second course meals		Desserts	
	β_{pre} (S.E.)	β_{post} (S.E.)	β_{pre} (S.E.)	β_{post} (S.E.)	β_{pre} (S.E.)	β_{post} (S.E.)
Ease of preparation	-.19 (.088)*	-.13 (.090)	.075 (.087)	-.060 (.083)	-.11 (.089)	-.023 (.090)
Fitted food goals	.022 (.11)	-.011 (.11)	.049 (.11)	-.15 (.12)	.068 (.12)	.018 (.13)
Fitted preferences	-.15 (.14)	-.30 (.16)	.10 (.14)	-.043 (.13)	-.086 (.13)	-.12 (.13)
Food characteristics		-.066 (.12)		-.012 (.12)		-.025 (.12)
Health	.72 (.13)***	.71 (.11)***	0.017 (.11)	.32 (.10)**	.47 (.11)***	.39 (.099)***
Taste	-.35 (.14)*	-.24 (.14)	-.22 (.13)	-.20 (.13)	-.25 (.12)*	-.18 (.13)
Intercept	.048 (.56)	.49 (.55)	.16 (.54)	.62 (.56)	.053 (.56)	.018 (.55)
Pseudo R^2	.105***	.121***	.009	.024*	.089***	.063***

Reported are the β coefficients and standard errors before being presented a justification ("pre") and after being presented one ("post"), per meal type. Food Characteristics-related motivations were only inquired "post-justification." *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

and the taste of the recipes, were no longer important after seeing a justification, indicating that the justifications affected what mattered to users when choosing a recipe.

6 Discussion

We examined to what extent natural language justifications in a knowledge-based food recommender system can support healthier recipe choices. We have presented two studies in which we have predicted recipe healthiness by the style of justification used (Study 1; RQ1), by the justification strategy used (Study 1, but mostly Study 2; RQ2), and by a user's choice motivation (both studies; RQ3). The effectiveness of eight different justification strategies, which have been grounded in psychological literature, have been evaluated across three different meal types. In doing so, Study 2 has employed a research design with a stricter baseline, examining choice reversal in back-to-back trial with and without justification, to which we are among the first in recommender system research (Zhu et al. 2012) and the first in food recommender research (Trattner and Elswailer 2019; Musto et al. 2020).

The overall 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 (Elswailer et al. 2022; Freyne and Berkovsky 2010; Trattner and Elswailer 2019), we have not focused on recipes that users liked in the past, but we have considered 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.

Our results indicate that pairwise justifications can help to reverse and steer user preferences toward healthier recipe recommendations, moving away from the commonly-preferred popular recipes. However, it seems that different types of justifications might be effective for different types of meals. The use of justifications has led to the most preference reversals in first course meal choices, for which we have found that strategies related to food features, health benefits and risks, and a user's lifestyle are most effective. In terms of persuasiveness, we expect these strategies to have appealed to different parts of peripheral-central route continuum of the elaboration likelihood model (Petty et al. 1997), since the Health Benefits and Health Risks justifications comprise both emotional and reflective responses (Rosenstock 1974; Taylor et al. 2007), while Food Features and User Lifestyle mainly require longer-term contemplation. The justification effectiveness is also reflected in the reported choice motivation of users: whereas ease of preparation and taste-related reasons neg-

actively affected "pre-justification" healthy choices, we have only found health-related reasons to choose a healthy recipe "post-justification."

The lack of any effects due to our justifications for second course meals could be attributed to the relatively high proportion of choice for the healthy recommendations in the baseline. Since these were preferred by a large proportion of users in the pre-justification trial, this left little room for improvement by introducing pairwise justifications.

Furthermore, we find that dessert choices are mostly taste-related, which undermines the effectiveness of most health and nutrition-related justifications. Nonetheless, our analysis of choice motivations suggests that the justifications have put more emphasis on the health aspect, as taste-related motivations decreased post-justification. We expect that justifications will mostly resonate with users who have strict dietary restrictions or ambitious healthy eating goals.

A limiting factor to our study's design was that the same order of meal types was maintained across all participants, starting with first course meals and ending with desserts. It is possible that users facing their second or third pair of recipes were less likely to change their preferences when facing a justification for those meal types. Alternatively, users might have already opted for the healthier choice in the first place (e.g., for the second course meals), because the justification for the first course meal activated reflective cognitive processes (Petty et al. 1997), which could have spilled over into later trials. In that sense, the results for the first course meals are likely to be more representative than those for second course meals, as this meal type is also less familiar to non-Italian natives.

The extent to which users are familiar with Italian cuisine has not been measured in our studies. It is possible that their evaluation of Italian-style recipes is different from, for example, American-style recipes, for example due to differences in dietary intake styles (Willett 2006). Italian recipes could fall under a Mediterranean diet, which is, among others, characterized by a high intake of fruits, vegetables, whole grains, legumes, and nuts and a much more moderate intake of red meat and dairy products compared to a North American diet (Trichopoulou et al. 2014). While all participants in both studies are based in the USA, requiring fluency in English, their cultural and ethnical background is not known, nor is their knowledge on various cuisines. Regional differences exist in the USA regarding the dominance of the Italian cuisine (Lee et al. 2014), among others due to large-scale immigration from Italy around the turn of the 20th century (Levenstein 1985). While the implications of an American-Italian match in cuisine cannot be inferred from our results, it is clear that many Americans are familiar with Italian-style meals (Lee et al. 2014). Moreover, general attitudes toward Italian products are rather positive (Bonaiuto et al. 2021), which might have increased user favorability toward any Italian recipe. Follow-up studies could control for this match between participants and cuisine.

Another limiting factor to our findings is the extent to which the recommendations fit into one's diet. While shifting toward a healthier dinner meal can go a long way in terms of improving dietary intake (Dallacker et al. 2018; Neumark-Sztainer et al. 2014), it is not informative about one's eating habits throughout the rest of the day. In a similar vein, the extent to which longer-term preferences have been considered is minimal. For our approach, we have assumed that one's preferences as elicited

in our knowledge-based system apply to the current session and beyond, using the session-based approach of previous recipe recommenders (Elahi et al. 2015; Starke and Trattner 2021). While this has been appropriate to address our research questions regarding justifications (RQ1-RQ2), future research is required to examine whether such an intervention will lead to longer-term changes.

We recommend that follow-up studies explore the effectiveness of different justification strategies in a less controlled environment. Whereas the research design of the current study is suitable to point out specific effects, most food choices are not made between pairs of recipes, but rather in the context larger lists, such as in “more like this” recommendations on recipe websites or in the context multi-list food recommender interfaces (Starke and Trattner 2021).

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 (Bettman et al. 1998), 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 (Musto et al. 2020). 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 toward 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 (Elsweiler et al. 2022; Trattner and Elsweiler 2019).

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 (Starke et al. 2021). 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. Furthermore, 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.

Acknowledgements This work is in part funded by MediaFutures partners and the Research Council of Norway (grant number 309339). We also received support through the PNRR project FAIR - Future AI Research (PE00000013), Spoke 6 - Symbiotic AI (CUP H97G22000210007), under the NRRP MUR program funded by NextGenerationEU. Moreover, we would like to thank Andrea Montemurro and Ivan Paparella for the implementation efforts and the Backstreet Boys for inspiring some of the authors.

Author Contributions Starke was responsible for the analyses and contributed to writing throughout the entire manuscript and all re-submissions. Musto developed the paper's main idea and was responsible for the development of the natural language recommender and contributed to most sections in the manuscript. Rapp helped with the development of the recommender and contributed in writing to Sections 2 and 3. Semeraro helped with the finalization of the paper and supported the initial idea development. Trattner helped to fund the paper and its finalization.

Declarations

Ethics approval This paper or a similar version is not currently under review by a journal or conference. This paper is void of plagiarism or self-plagiarism as defined by the Committee on Publication Ethics and Springer Guidelines. All relevant materials for processing and analysis are available in our repository: <https://osf.io/hn3et/>.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ajzen, I., Fishbein, M.: The influence of attitudes on behavior. In: *The Handbook of Attitudes*. Volume 1: Basic Principles, pp. 173–221. Routledge, London (2005)
- Ajzen, I.: The theory of planned behavior. *Organizational Behavior and Human Decision Processes. Theories of Cognitive Self-Regulation* **50**(2), 179–211 (1991). [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T).

- Ateş, H.: Merging theory of planned behavior and value identity personal norm model to explain pro-environmental behaviors. *Sustain. Prod. Consump.* **24**, 169–180 (2020). <https://doi.org/10.1016/j.spc.2020.07.006>
- Bandura, A.: *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall, New York, NY (1986)
- Bandura, A.: *Self-Efficacy: The Exercise of Control*. Freeman, New York, NY (1997)
- Bettman, J.R., Luce, M.F., Payne, J.W.: Constructive consumer choice processes. *J. Consum. Res.* **25**(3), 187–217 (1998)
- Bialkova, S., van Trijp, H.C.: An efficient methodology for assessing attention to and effect of nutrition information displayed front-of-pack. *Food Qual. Prefer.* **22**(6), 592–601 (2011)
- Bialkova, S., Grunert, K.G., Juhl, H.J., Wasowicz-Kirylo, G., Stysko-Kunkowska, M., van Trijp, H.C.: Attention mediates the effect of nutrition label information on consumers' choice. Evidence from a choice experiment involving eye-tracking. *Appetite* **76**, 66–75 (2014)
- Biran, O., Cotton, C.: Explanation and justification in machine learning: a survey. In: *IJCAI-17 Workshop on Explainable AI (XAI)*, p. 8 (2017)
- Bonaiuto, F., De Dominicis, S., Ganucci Cancellieri, U., Crano, W.D., Ma, J., Bonaiuto, M.: Italian food? sounds good! made in Italy and Italian sounding effects on food products' assessment by consumers. *Front. Psychol.* **12**, 581492 (2021)
- Cadario, R., Chandon, P.: Which healthy eating nudges work best? A meta-analysis of field experiments. *Mark. Sci.* **39**(3), 465–486 (2020)
- Cena, F., Likavec, S., Rapp, A.: 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*, 1–26 (2018a). <https://doi.org/10.1007/s10796-017-9818-3>. Cited By 0; Article in Press
- Cena, F., Rapp, A., Musto, C., Lops, P.: Towards a conceptual model for holistic recommendations. In: *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization. UMAP '18*, pp. 207–210. Association for Computing Machinery, New York, NY (2018b). <https://doi.org/10.1145/3213586.3225248>
- Cena, F., Rapp, A., Musto, C., Semeraro, G.: Generating recommendations from multiple data sources: a methodological framework for system design and its application. *IEEE Access* **8**, 183430–183447 (2020). <https://doi.org/10.1109/ACCESS.2020.3028777>
- Chen, L., Wang, F.: Explaining Recommendations based on Feature Sentiments in Product Reviews. In: *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, pp. 17–28 (2017). ACM
- Cramer, H., Evers, V., Ramlal, S., Van Someren, M., Rutledge, L., Stash, N., Aroyo, L., Wielinga, B.: The effects of transparency on trust and acceptance of a content-based art recommender. *User Model. User-Adap. Inter.* **18**(5), 455–496 (2008)
- Dallacker, M., Hertwig, R., Mata, J.: The frequency of family meals and nutritional health in children: a meta-analysis. *Obes. Rev.* **19**(5), 638–653 (2018)
- DiSalvo, C., Sengers, P., Brynjarsdóttir, H.: Mapping the landscape of sustainable hci. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '10*, pp. 1975–1984. Association for Computing Machinery, New York, NY, USA (2010)
- Ekstrand, M.D., Willemsen, M.C.: Behaviorism is not enough: better recommendations through listening to users. In: *Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 221–224 (2016)
- Elahi, M., Ge, M., Ricci, F., Fernández-Tobías, I., Berkovsky, S., David, M., et al.: Interaction design in a mobile food recommender system. In: *CEUR Workshop Proceedings (2015)*. CEUR-WS
- Elahi, M., Ge, M., Ricci, F., Massimo, D., Berkovsky, S.: Interactive food recommendation for groups. In: *8th ACM Conference on Recommender Systems (RecSys) (2014)*. CEUR Workshop Proceedings
- Elsweiler, D., Hauptmann, H., Trattner, C.: Food recommender systems. In: *Recommender Systems Handbook*, pp. 871–925. Springer, New York, NY, USA (2022)
- Elsweiler, D., Hauptmann, H., Trattner, C.: Food recommender systems. In: Ricci, F., Rokach, L., Shapira, B. (eds.) *Recommender Systems Handbook*, pp. 871–925. Springer, New York, NY (2022). https://doi.org/10.1007/978-1-0716-2197-4_23
- Elsweiler, D., Trattner, C., Harvey, M.: Exploiting food choice biases for healthier recipe recommendation. In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '17*, pp. 575–584. Association for Computing Machinery, New York, NY, USA (2017). <https://doi.org/10.1145/3077136.3080826>

- Freyne, J., Berkovsky, S.: Intelligent food planning: personalized recipe recommendation. In: Proceedings of the 15th International Conference on Intelligent User Interfaces, pp. 321–324 (2010)
- Friedrich, G., Zanker, M.: A taxonomy for generating explanations in recommender systems. *AI Mag.* **32**(3), 90–98 (2011)
- Furst, T., Connors, M., Bisogni, C.A., Sobal, J., Falk, L.W.: Food choice: a conceptual model of the process. *Appetite* **26**(3), 247–266 (1996)
- Gkika, S., Lekakos, G.: The persuasive role of explanations in recommender systems. In: *BCSS@ PER-SUASIVE*, pp. 59–68 (2014)
- Gomez-Uribe, C.A., Hunt, N.: The netflix recommender system: algorithms, business value, and innovation. *ACM Trans. Manag. Inform. Syst. (TMIS)* **6**(4), 1–19 (2015)
- Gutjar, S., de Graaf, C., Kooijman, V., de Wijk, R.A., Nys, A., Ter Horst, G.J., Jager, G.: The role of emotions in food choice and liking. *Food Res. Int.* **76**, 216–223 (2015)
- Hagger, M.S.: Non-conscious processes and dual-process theories in health psychology. *Health Psychol. Rev.* **10**(4), 375–380 (2016). <https://doi.org/10.1080/17437199.2016.1244647>. (PMID: 27718880)
- Herlocker, J., Konstan, J.: Content-independent task-focused recommendation. *IEEE Internet Comput.* **5**(6), 40–47 (2001)
- Hors-Fraile, S., Benjumea, F.J.N., Hernández, L.C., Ruiz, F.O., Fernandez-Luque, L.: Design of two combined health recommender systems for tailoring messages in a smoking cessation app (2016). [arXiv:1608.07192](https://arxiv.org/abs/1608.07192)
- Hors-Fraile, S., Candel, M.J., Schneider, F., Malwade, S., Nunez-Benjumea, F.J., Syed-Abdul, S., Fernandez-Luque, L., de Vries, H.: Applying collective intelligence in health recommender systems for smoking cessation: A comparison trial. *Electronics* **11**(8), 1219 (2022)
- Howard, S., Adams, J., White, M.: Nutritional content of supermarket ready meals and recipes by television chefs in the united kingdom: cross sectional study. *BMJ* **345** (2012)
- Ilich, J.Z., Vollono, J.A., Brownbill, R.A.: Impact of nutritional knowledge on food choices and dietary intake of college students. *Journal of the American Dietetic Association* **99**(9, Supplement), 89 (1999). [https://doi.org/10.1016/S0002-8223\(99\)00707-5](https://doi.org/10.1016/S0002-8223(99)00707-5)
- Jannach, D., Zanker, M., Felfernig, A., Friedrich, G.: *Recommender Systems: An Introduction*. Cambridge University Press, New York, NY (2010)
- Johnson, H., Johnson, P.: Explanation Facilities and Interactive Systems. In: Proceedings of the 1st International Conference on Intelligent User Interfaces, pp. 159–166 (1993). *ACM*
- Jun, J., Kang, J., Arendt, S.W.: The effects of health value on healthful food selection intention at restaurants: considering the role of attitudes toward taste and healthfulness of healthful foods. *Int. J. Hosp. Manag.* **42**, 85–91 (2014). <https://doi.org/10.1016/j.ijhm.2014.06.002>
- Kahneman, D.: *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York, NY, USA (2011)
- Kalnikaitė, V., Bird, J., Rogers, Y.: Decision-making in the aisles: informing, overwhelming or nudging supermarket shoppers? *Pers. Ubiquit. Comput.* **17**(6), 1247–1259 (2013)
- Köster, E.P.: Diversity in the determinants of food choice: a psychological perspective. *Food Qual. Prefer.* **20**(2), 70–82 (2009)
- Kroese, F.M., Marchiori, D.R., de Ridder, D.T.: Nudging healthy food choices: a field experiment at the train station. *J. Public Health* **38**(2), 133–137 (2016)
- Lee, J.H., Hwang, J., Mustapha, A.: Popular ethnic foods in the united states: a historical and safety perspective. *Comp. Rev. Food Sci. Food Saf.* **13**(1), 2–17 (2014)
- Leipold, N., Madenach, M., Schäfer, H., Lurz, M., Terzimehic, N., Groh, G., Böhm, M., Gedrich, K., Kremer, H.: Nutrilize a personalized nutrition recommender system: an enable study. *HealthRecSys@ RecSys* **2216**, 24–29 (2018)
- Levenstein, H.: The American response to Italian food, 1880–1930. *Food Foodways* **1**(1–2), 1–23 (1985)
- Locke, E.A., Latham, G.P.: Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. *Am. Psychol.* **57**(9), 705–717 (2002). <https://doi.org/10.1037/0003-066x.57.9.705>
- McFadden, D.: Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (ed.) *Frontiers in Econometrics*, pp. 105–142. Academic Press, New York (1973)
- Munson, S.A., Consolvo, S.: 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, pp. 25–32 (2012). <https://doi.org/10.4108/icst.pervasivehealth.2012.248691>
- Musto, C., de Gemmis, M., Lops, P., Semeraro, G.: Generating post hoc review-based natural language justifications for recommender systems. *User Modeling and User-Adapted Interaction*, pp. 1–45 (2020)

- Musto, C., Trattner, C., Starke, A., Semeraro, G.: Towards a knowledge-aware food recommender system exploiting holistic user models. In: Kuflik, T., Torre, I., Burke, R., Gena, C. (eds.) Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2020, Genoa, Italy, July 12–18, 2020, pp. 333–337. Association for Computing Machinery, New York, NY, USA (2020). <https://doi.org/10.1145/3340631.3394880>
- Musto, C., Narducci, F., Lops, P., de Gemmis, M., Semeraro, G.: Linked open data-based explanations for transparent recommender systems. *Int. J. Hum. Comput. Stud.* **121**, 93–107 (2019)
- Musto, C., Polignano, M., Semeraro, G., de Gemmis, M., Lops, P.: Myrror: a platform for holistic user modeling. *User Model. User-Adap. Inter.* **30**(3), 477–511 (2020)
- Musto, C., Narducci, F., Polignano, M., De Gemmis, M., Lops, P., Semeraro, G.: Myrrorbot: A digital assistant based on holistic user models for personalized access to online services. *ACM Trans. Inform. Syst. (TOIS)* **39**(4), 1–34 (2021)
- Neumark-Sztainer, D., MacLehose, R., Loth, K., Fulkerson, J.A., Eisenberg, M.E., Berge, J.: What's for dinner? Types of food served at family dinner differ across parent and family characteristics. *Public Health Nutr.* **17**(1), 145–155 (2014)
- Nunes, I., Jannach, D.: A systematic review and taxonomy of explanations in decision support and recommender systems. *User Model. User-Adap. Inter.* **27**(3–5), 393–444 (2017)
- Organization, W.H.: Diet, Nutrition, and the Prevention of Chronic Diseases: Report of a Joint WHO/FAO Expert Consultation, vol. 916. World Health Organization, Geneva (2003)
- Orji, R., Moffatt, K.: Persuasive technology for health and wellness: State-of-the-art and emerging trends. *Health Inform. J.* **24**(1), 66–91 (2018)
- Petty, R.E., Heesacker, M., Hughes, J.N.: The elaboration likelihood model: Implications for the practice of school psychology. *J. Sch. Psychol.* **35**(2), 107–136 (1997). [https://doi.org/10.1016/S0022-4405\(97\)00003-4](https://doi.org/10.1016/S0022-4405(97)00003-4)
- Pinder, C., Vermeulen, J., Cowan, B.R., Beale, R.: Digital behaviour change interventions to break and form habits. *ACM Trans. Comput. Hum. Interact.* (2018). <https://doi.org/10.1145/3196830>
- Polignano, M., Narducci, F., Iovine, A., Musto, C., De Gemmis, M., Semeraro, G.: Healthassistantbot: a personal health assistant for the Italian language. *IEEE Access* **8**, 107479–107497 (2020)
- Prochaska, J.O., Velicer, W.F.: The transtheoretical model of health behavior change. *Am. J. Health Promot.* **12**(1), 38–48 (1997)
- Rapp, A., Tirassa, M., Tirabeni, L.: Rethinking technologies for behavior change: A view from the inside of human change. *ACM Trans. Comput.-Hum. Interact.* (2019). <https://doi.org/10.1145/3318142>
- Reiter, E., Dale, R.: Building Natural Language Generation Systems. Cambridge University Press, New York, NY (2000)
- Rosenstock, I.M.: Historical origins of the health belief model. *Health Educ. Monogr.* **2**(4), 328–335 (1974). <https://doi.org/10.1177/109019817400200403>
- Schäfer, H., Hors-Fraile, S., Karumur, R.P., Calero Valdez, A., Said, A., Torkamaan, H., Ulmer, T., Trattner, C.: Towards health (aware) recommender systems. In: Proceedings of the 2017 International Conference on Digital Health, pp. 157–161 (2017)
- Schäfer, H., Willemsen, M.C.: Rasch-based tailored goals for nutrition assistance systems. In: Proceedings of the 24th International Conference on Intelligent User Interfaces, pp. 18–29 (2019)
- Shunk, D.: Self-efficacy for learning and performance. In: Annual Conference of the American Educational Research Association, pp. 1–24 (1996). <https://doi.org/10.4108/icst.pervasivehealth.2012.248691>
- Sinha, R., Swearingen, K.: The Role of Transparency in Recommender Systems. In: CHI'02 Extended Abstracts on Human Factors in Computing Systems, pp. 830–831 (2002). ACM
- Starke, A., Asotic, E., Trattner, C.: “serving each user”: Supporting different eating goals through a multi-list recommender interface. In: Fifteenth ACM Conference on Recommender Systems, pp. 124–132 (2021)
- Starke, A.D., Trattner, C.: Promoting healthy food choices online: A case for multi-list recommender systems. In: HEALTHI'21: Joint Proceedings of ACM IUI 2021 Workshops (2021)
- Starke, A.D., Willemsen, M.C., Snijders, C.: With a little help from my peers: Depicting social norms in a recommender interface to promote energy conservation. In: Proceedings of the 25th International Conference on Intelligent User Interfaces, pp. 568–578 (2020)
- Starke, A.: The effectiveness of advice solicitation and social peers in an energy recommender system. In: 6th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, IntRS 2019, pp. 65–71 (2019). CEUR

- Starke, A.D., Willemsen, M.C., Trattner, C.: Nudging healthy choices in food search through visual attractiveness. *Front. Artif. Intell.* **4**, 20 (2021). <https://doi.org/10.3389/frai.2021.621743>
- Stowell, E., Lyson, M.C., Saksone, H., Wurth, R.C., Jimison, H., Pavel, M., Parker, A.G.: Designing and evaluating mhealth interventions for vulnerable populations: A systematic review. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. CHI '18, pp. 1–17. Association for Computing Machinery, New York, NY, USA (2018). <https://doi.org/10.1145/3173574.3173589>
- Symeonidis, P., Nanopoulos, A., Manolopoulos, Y.: Providing justifications in recommender systems. *IEEE Trans. Syst. Man Cybern.-A: Syst. Humans* **38**(6), 1262–1272 (2008)
- Taylor, D., Bury, M., Campling, N., Carter, S., Garfied, S., Newbould, J., Rennie, T.: 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 (2007)
- Tintarev, N., Masthoff, J.: Designing and evaluating explanations for recommender systems. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) *Recommender Systems Handbook*, pp. 479–510. Springer, New York, NY (2011)
- Tintarev, N., Masthoff, J.: Evaluating the effectiveness of explanations for recommender systems. *UMUAI* **22**(4–5), 399–439 (2012)
- Tran, T.N.T., Atas, M., Felfernig, A., Stettinger, M.: An overview of recommender systems in the healthy food domain. *J. Intell. Inform. Syst.* **50**(3), 501–526 (2018)
- Trattner, C., Elswailer, D.: Food recommendations. In: *Collaborative Recommendations: Algorithms. Practical Challenges and Applications*, pp. 653–685. World Scientific, UK (2019)
- Trattner, C., Elswailer, D.: 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, pp. 489–498 (2017)
- Trichopoulou, A., Martínez-González, M.A., Tong, T.Y., Forouhi, N.G., Khandelwal, S., Prabhakaran, D., Mozaffarian, D., de Lorgeril, M.: Definitions and potential health benefits of the Mediterranean diet: views from experts around the world. *BMC Med.* **12**(1), 1–16 (2014)
- Tudoran, A., Olsen, S.O., Dopico, D.C.: The effect of health benefit information on consumers health value, attitudes and intentions. *Appetite* **52**(3), 568–579 (2009). <https://doi.org/10.1016/j.appet.2009.01.009>
- Wardle, J., Parmenter, K., Waller, J.: Nutrition knowledge and food intake. *Appetite* **34**(3), 269–275 (2000). <https://doi.org/10.1006/appe.1999.0311>
- Willett, W.C.: The mediterranean diet: science and practice. *Public Health Nutr.* **9**(1a), 105–110 (2006)
- Zhu, H., Huberman, B., Luon, Y.: To switch or not to switch: understanding social influence in online choices. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2257–2266 (2012)
- Zimmerman, B.J.: Self-efficacy: an essential motive to learn. *Contemp. Educ. Psychol.* **25**(1), 82–91 (2000). <https://doi.org/10.1006/ceps.1999.1016>
- Zulkosky, K.: Self-efficacy: a concept analysis. *Nurs. Forum* **44**(2), 93–102 (2009). <https://doi.org/10.1111/j.1744-6198.2009.00132.x>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Alain D. Starke is an assistant professor persuasive communication at the Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Netherlands. He is also an adjunct associate professor at the University of Bergen, Norway, where he is affiliated to the SFI MediaFutures Centre for Responsible Media Technology. He has a MSc degree in Innovation Sciences from TU Eindhoven, where he also obtained a PhD in Human-Technology Interaction in 2019. His PhD research on energy recommender systems was funded through a Research Talent Grant from the Dutch Research Council, while his Postdoc research on psychologically-informed food recommenders was awarded a Niels Stensen Fellowship. In his work, he uses recommender systems to support behavioral change, about he has published in a variety of scientific disciplines. This includes journals, such as *Energy Research and Social Science*, *Nature Food*, *TiiS*, and *Trends in Food Science & Technology*, but also conferences, such as CHI, IUI, RecSys and UMAP.

Cataldo Musto is assistant professor at the Department of Informatics, University of Bari. He completed his Ph.D. in 2012 with a thesis entitled “Enhanced Vector Space Models for Content-based Recommender Systems.” His research focuses on the adoption of natural language processing techniques and models for fine-grained semantic content representation in recommender systems and user modeling platforms. He has been involved in various national and international research projects that dealt with natural language processing and recommender systems. From 2016 to 2019, he acted as project leader for a funded project regarding Semantic Holistic User Modeling for Personalized Access to Digital Content and Services. He regularly acts as a PC member on several top-tier conferences and co-organizes or co-chairs a number of workshops. Recently, he co-organized RecSys workshops about new trends in content-based RecSys (2016), UMAP workshops about Holistic User Modeling (2017, 2018 and 2019) and UMAP Workshop on Explainable User Modeling (2020). He is one of the authors of the textbook “Semantics in Adaptive and Personalized Systems: Methods, Tools and Applications,” edited by Springer.

Amon Rapp is an Associate Professor in the Department of Computer Science at the University of Torino, where he leads the Smart Personal Technology Lab at ICxT. He is a member of the SIOS (Smart Interactive Objects and Systems) group. His research is situated within the area of human-computer interaction and focuses on how people interact with self-tracking technologies, wearable devices, video games, gamified applications, and behavior change systems. Amon’s work has been published in international journals such as ACM Transactions on Computer-Human Interaction, IEEE Transactions on Human-Machine Systems, Human-Computer Interaction, Computers & Education, and the International Journal of Human-Computer Studies. He is an associate chair of the program committee of several top-tier international conferences like CHI and CSCW.

Giovanni Semeraro is full professor of computer science at University of Bari Aldo Moro, Italy. He leads the Semantic Web Access and Personalization (SWAP) “Antonio Bello” research group. He was one of the founders of AILC (Italian Association for Computational Linguistics) and on the Board of Directors till 2018. His research interests include machine learning; AI and language games; recommender systems; user modeling; intelligent information mining, retrieval, and filtering; semantics and social computing; natural language processing; the semantic web; personalization. He has been the principal investigator of University of Bari in several European national and regional projects. He is author of more than 400 publications in international journals, conference and workshop proceedings, as well as of 3 books, including the textbook “Semantics in Adaptive and Personalized Systems: Methods, Tools and Applications” published by Springer. He regularly serves in the PC of the top conferences in his area and was Program co-chair of RecSys 2015, General Co-chair of UMAP 2013 and Program Co-Chair of CLiC-it 2019. Since 2013, he has been the coordinator of the 2nd Cycle Degree Program in Computer Science at University of Bari. He is also the coordinator of the 1st edition of the Master in Data Science at University of Bari. He is a member of the Steering Committee of the National Laboratory of Artificial Intelligence and Intelligent Systems (AIIS) of the National Interuniversity Consortium for Informatics (CINI) and of the Steering Committee of the ACM Conference Series on Recommender Systems.

Christoph Trattner is a full professor at the University of Bergen and the Founder and the Center Director of the Research Centre for Responsible Media Technology & Innovation. He is also the founder and leader of the DARS research group at UiB and holds a Research Professor (Forsker I) position in NORCE one of Norway’s largest research institutions. He received a PhD (with distinction), an MSc (with distinction) and a BSc in Computer Science and Telematics from Graz University of Technology (Austria). Since 2012, his research is focused on improving diets in the context of food and media. He published over 100 scientific articles in top venues in this context of research. He is a former board member of Media City Bergen, Norway’s largest Media Cluster and a member of editorial boards including Springer’s Journal of Intelligent Information Systems, and AI and Ethics. Since June 2021, he is an appointed ACM Distinguished Speaker and ACM Senior Member.