# Recommender Systems and Nudges for Healthier Food Choices

**Ayoub El Majjodi**

Dissertation for the degree of philosophiae doctor (PhD)

at the University of Bergen, Norway

2025

Dissertation date:

| | |
|---|---|
| Year: | 2025 |
| Title: | Recommender Systems and Nudges for Healthier Food Choices |
| Author: | Ayoub El Majjodi |

# Scientific environment

The research presented in this dissertation was conducted at the Research Centre for Responsible Media Technology & Innovation (MediaFutures) and the Behavioral Data Analytics & Recommender Systems Research Group, hosted by the Department of Information Science and Media Studies at the University of Bergen, Bergen, Norway. It was supervised by Prof. Christoph Trattner and co-supervised by Ass. Prof. Alain D. Starke.

# Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, Prof. Christoph Trattner and Ass. Prof. Alain D. Starke, for their invaluable guidance, continuous support, and insightful feedback throughout this PhD journey. Their expertise and encouragement have been instrumental in shaping my research and helping me navigate the challenges of academia.

I extend my sincere appreciation to the members of my PhD jury for their time, thoughtful evaluations, and constructive feedback. Additionally, I would like to acknowledge all the instructors who contributed to my academic training, and for providing me with a strong foundations in the field.

My heartfelt thanks go to my colleagues and friends, whose support, collaboration, and stimulating discussions have made this journey more rewarding. Being part of MediaFutures for four years has been an incredible experience, and I deeply appreciate the opportunities, research environment, and camaraderie it has provided.

Finally, I am forever grateful to my family for their unwavering love, encouragement, and sacrifices. To my parents, brothers, and sisters, your belief in me has been my greatest source of strength, and without you none of this would have been possible.

This thesis is a reflection of the collective support and inspiration I have received, and I dedicate it to everyone who has been part of this journey.

# Abstract

Recommender systems are widely used to address the challenge of information overload by presenting users with the most relevant content through personalization techniques. In the food domain, decision-making is particularly complex due to the multifaceted nature of food choices, which are influenced by a range of individual, contextual, and environmental factors. Despite this complexity, recommender systems have shown considerable promise in modeling real world food preferences and supporting users in navigating food related decisions. Their adoption in food applications has been steadily increasing, reflecting the growing importance and difficulty of making informed, personalized, and health-conscious dietary choices. Nonetheless, these systems have often been shown to generate predominantly popular food options, which tend to be less healthy. As users interact with such systems, they are repeatedly exposed to unhealthy choices, which in turn reinforces their preferences for these items. This feedback loop causes algorithms to prioritize popular yet nutritionally poor options, ultimately amplifying unhealthy eating behaviors with potential negative implications for public health. At the same time, digital nudging has emerged as a promising strategy for influencing user behavior in subtle and non intrusive ways. However, limited research has investigated how digital nudges and recommender systems can be effectively combined particularly in user-centered settings aimed at supporting informed decision-making and promoting behavioral change.

To address this gap, this thesis adopts a Design Science Research (DSR) methodology to design, implement, and evaluate food recommender systems augmented with digital nudges. The research is documented across several peer-reviewed manuscripts and supported by both offline algorithmic evaluations and online user experiments. These studies examine how various preference elicitation methods, nudging techniques, and user characteristics such as food knowledge and dietary goals interact to shape user experience and behavior.

The findings reveal that several nudging techniques warrant further investigation in the context of food recommender systems, particularly through user-centric evaluation approaches. Moreover, while digital nudges can support healthier food choices, their effectiveness varies depending on personalization, user familiarity, and system design. Interestingly, non-personalized recommendations with clear nutritional labeling were often more effective in encouraging healthy decisions than personalized options. Additionally, the interplay between preference elicitation methods, user knowledge, and nudging strategies significantly influenced user choices, interactions, and overall experience.

This thesis contributes to the fields of recommender systems and persuasive technologies by demonstrating how digital nudges and system design features jointly influence health related decision-making. It emphasizes the importance of user-centric evaluation and lays the foundation for future research on adaptive nudging, long-term behavior change, and real-world deployments in food-related digital platforms.

# Sammendrag

Anbefalingssystemer er mye brukt for å møte utfordringen med informasjonsoverskudd ved å presentere brukere for det mest relevante innholdet gjennom personaliseringsteknikker. Innen matdomenet er beslutningstaking spesielt kompleks på grunn av den sammensatte naturen til matvalg, som påvirkes av en rekke individuelle, kontekstuelle og miljømessige faktorer. Til tross for denne kompleksiteten har anbefalingssystemer vist betydelig potensial i å modellere virkelige matpreferanser og støtte brukere i å navigere matrelaterte beslutninger. Bruken av slike systemer i matapplikasjoner har økt jevnt, noe som gjenspeiler den økende betydningen og vanskeligheten ved å ta informerte, personaliserte og helsefremmende kostholdsvalg.

Likevel har disse systemene ofte vist seg å generere hovedsakelig populære matvalg, som ofte er mindre sunne. Når brukere samhandler med slike systemer, blir de gjentatte ganger eksponert for usunne alternativer, noe som igjen forsterker preferansene for disse valgene. Denne tilbakemeldingssløyfen fører til at algoritmene prioriterer populære, men ernæringsmessig dårlige alternativer, noe som til slutt forsterker usunne spisevaner med potensielle negative konsekvenser for folkehelsen. Samtidig har digital nudging (digital dytting) dukket opp som en lovende strategi for å påvirke brukeradferd på subtile og ikke-påtrengende måter. Det finnes imidlertid begrenset forskning på hvordan digitale nudger og anbefalingssystemer effektivt kan kombineres, særlig i brukerfokuserte kontekster som har som mål å støtte informert beslutningstaking og fremme adferdsendring.

For å møte dette kunnskapshullet benytter denne avhandlingen en Design Science Research (DSR)-metodologi for å designe, implementere og evaluere mat-anbefalingssystemer forsterket med digitale nudger. Forskningen er dokumentert i flere fagfellevurderte artikler og støttes av både offline algoritmiske evalueringer og online brukerstudier. Disse studiene undersøker hvordan ulike metoder for preferanseinnhenting, nudging-teknikker og brukerkarakteristikker som matkunnskap og kostholdsmål samvirker for å forme brukeropplevelse og adferd.

Funnene viser at flere nudging-teknikker fortjener videre utforskning i konteksten av mat-anbefalingssystemer, særlig gjennom brukerorienterte evalueringsmetoder. Videre kan digitale nudger støtte sunnere matvalg, men deres effektivitet varierer avhengig av personalisering, brukerens kjennskap og systemets utforming. Interessant nok var ikke-personaliserte anbefalinger med tydelig ernæringsmerking ofte mer effektive for å fremme sunne valg enn personaliserte alternativer. I tillegg hadde samspillet mellom preferanseinnhentingsmetoder, brukerkunnskap og nudging-strategier en betydelig innflytelse på brukerens valg, interaksjoner og totale opplevelse.

Denne avhandlingen bidrar til feltene anbefalingssystemer og persuasive teknologier ved å vise hvordan digitale nudger og systemdesignfunksjoner i fellesskap påvirker helserelaterte beslutningsprosesser. Den fremhever viktigheten av brukerorientert evaluering og legger grunnlaget for fremtidig forskning på adaptiv nudging, langsiktige adferdsendringer og implementering i virkelige matrelaterte digitale plattformer.

# Outline

This PhD project investigates the design, implementation, and evaluation of food recommender systems enhanced with digital nudges to support users in making healthier food choices. It contributes to the broader field of recommender systems by exploring how these technologies can be optimized to facilitate informed decision-making.

The research involves the development and empirical testing of novel system designs that combine personalized recommendations with digital nudging techniques, especially the nutritional food labels and food explanations. These elements are embedded within the system's user interface to guide and influence food choices towards more healthy and informed choices.

Adopting a Design Science Research (DSR) methodology, the project has produced multiple system prototypes, which have been evaluated through real-world user studies. To date, the research has resulted in five peer-reviewed scientific publications, with an additional manuscript under review. Collectively, these contributions provide both empirical findings and theoretical advancements to the fields of recommender systems, user modeling, and digital behavior change. This thesis is structured as fellow: **Chapter 1** introduces the research project, contextualizes it within the field of food recommender systems, outlines the research gap, and presents the main motivations and contributions of the thesis. **Chapter 2** presents relevant background literature and defines key concepts related to recommender systems and digital nudges in the context of decision-making. **Chapter 3** describes the research methodology employed and explains how it was applied to address the proposed research questions. **Chapter 4** summarizes and discuss the main findings presented in the included manuscripts, hightlith the main limitions and provide the possible future research directions.

The following manuscripts are included in this thesis as part of the fulfilment of the requirements for the dissertation [1]:

- Ayoub El Majjodi, Sohail Ahmed Khan, Alain D. Starke, Mehdi Elahi, and Christoph Trattner. "Integrating Digital Food Nudges and Recommender Systems: Current Status and Future Directions." *IEEE Access. 2025. (under review).*

- Ayoub EL Majjodi, Alain D. Starke, and Christoph Trattner. "Nudging towards health? examining the merits of nutrition labels and personalization in a recipe recommender

[1] The candidate co-authored the manuscripts and was actively involved in all phases of the research, including formulating research questions, study design, data collection, analysis, interpretation, and manuscript preparation. A signed declaration of co-authorship statement is submitted in accordance with Faculty guidelines.

system." *In Proceedings of the 30th ACM conference on user modeling, adaptation and personalization, 2022.*

- Alain D. Starke, Ayoub El Majjodi, and Christoph Trattner. "Boosting Health? Examining the Role of Nutrition Labels and Preference Elicitation Methods in Food Recommendation." *In Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, ACM Recommender System, 2022.*
- Ayoub EL Majjodi, Alain D. Starke, Mehdi Elahi, Christoph Trattner. "The Interplay between Food Knowledge, Nudges, and Preference Elicitation Methods Determines the Evaluation of a Recipe Recommender System." *In Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, ACM RecSys, 2023.*
- Giovanni Castiglia, EL Majjodi Ayoub, Alain D. Starke, Fedelucio Narducci, Yashar Deldjoo, Federica Calò. "Nudging towards health in a conversational food recommender system using multi-modal interactions and nutrition labels." *In Fourth Knowledge-aware and Conversational Recommender Systems Workshop (KaRS). ACM RecSys, 2022.*
- Ayoub El Majjodi, Sohail Ahmed Khan, Alain D. Starke, Mehdi Elahi, Christoph Trattner. "Advancing Visual Food Attractiveness Predictions for Healthy Food Recommender Systems." *In The 6th Workshop on Health Recommender Systems, ACM RecSys, 2024.*

# Contents

# List of Figures

# Acronyms

**BIs** Behavioural Indicators. 27

**CB** Content-Based. 13, 25, 33

**Cb** Constraint-based. 33

**CF** Collaborative Filtering. 10, 12, 25

**DSR** Design Science Research. vii, ix, xi, xv, 23, 24, 36

**eq** Equation. 10, 17

**KB** Knowledge-Based . 15, 34

**MAE** Mean Absolute Error. 16

**NMF** Non-Negative Matrix Factorization. 12

**OSAs** Objective System Aspects. 26, 27

**PAs** Perception Aspects. 26, 27

**RMSE** Root Mean Squared Error. 16

**RQ** Research Question. 5

**SVD** Singular Value Decomposition . xiii, 12

**UCAs** User Characteristics Aspects. 27

**UXAs** User Experience Aspects. 26, 27

# Chapter 1

# Introduction

## 1.1 Introduction and Motivation

The phenomenon of choice overload arises when individuals struggle to make decisions due to an overwhelming number of available options. This challenge became particularly evident with the advent of the digital age, where users faced an unprecedented volume of information. In response, recommender systems emerged in the mid-1990's as a technological solution to assist users in navigating vast choice spaces (Goldberg et al., 1992; Long et al., 2025). By analyzing past user behaviors, these systems predict and suggest the most relevant options for each user, thereby mitigating decision fatigue (He et al., 2024; Chen et al., 2013) and supporting people in decision making process. Recommender systems have evolved significantly, demonstrating success across various domains, including e-commerce, entertainment, and healthcare (Jannach and Jugovac, 2019).

The distinct characteristics of the food domain extend beyond its fundamental role as a means of survival; food holds profound personal, social, and cultural significance. As the adage suggests (Capaldi, 1996), "we are what we eat." On an individual level, dietary choices are closely linked to personal lifestyles, which in turn align with specific types or groups of foods. Furthermore, factors such as income, socio-economic status, education, and knowledge play a crucial role in shaping food preferences. Additionally, food consumption is deeply embedded in social structures, as dietary habits are both shaped and constrained by social class and interpersonal influences. Indeed, the culinary traditions of a given place or culture often serve as reflections of its unique history, values, and way of life (Montanari, 2006).

In contemporary society, technological advancements have significantly transformed food-related decision-making. Digital platforms enable the widespread sharing of food images, recipes, cooking tutorials, restaurant recommendations, and dietary experiences, thereby expanding the online food choice space. However, this vast array of options can overwhelm users, making it difficult to determine what to eat (Galef Jr, 1996; Vasan et al., 2025). For instance, a pregnant woman may encounter an unsuitable online diet plan, an athlete may struggle to personalize a nutrition regimen, and a couple may face challenges in selecting a local restaurant. In such situations, individuals often rely on external recommendations, such as word of mouth, to navigate their food choices based on the experiences of others with similar preferences and needs. Recently, the food domain has also recommendation technology to facilitate user interactions with food choices (Trattner and Elsweiler, 2017b).

A food recommender system, as a general-purpose framework, is inherently aligned with the principles of traditional recommender system approaches (Freyne and Berkovsky, 2010a). These systems are designed to deliver personalized content, such as recipes (Pecune et al., 2020), food menus (Asani et al., 2021), and nutritional advice (Toledo et al., 2019), to users or groups of users (Berkovsky and Freyne, 2010) based on their preferences and needs. Recommendations relevancy, is derived through the analysis of user profiles, which may include various information (e.g., demographic data, food knowledge) as well as historical interaction patterns with the system.

The conventional functioning of a food recommender system is underpinned by three primary stages (Elsweiler et al., 2022). Firstly, contextual analysis, which involves understanding the user's preferences, contextual factors, and dietary restrictions. The second step involves algorithmic formulation, where an appropriate recommendation algorithm is selected and applied to generate relevant suggestions based on user data. Finally, presentation design creates an intuitive and effective user interface for presenting the system's recommendations. These processes, when seamlessly integrated, enable the food recommender system to provide contextually relevant and personalized food-related content that aligns with the specific preferences and needs of the user. Figure 1.1 presents an example of a food recommender system.

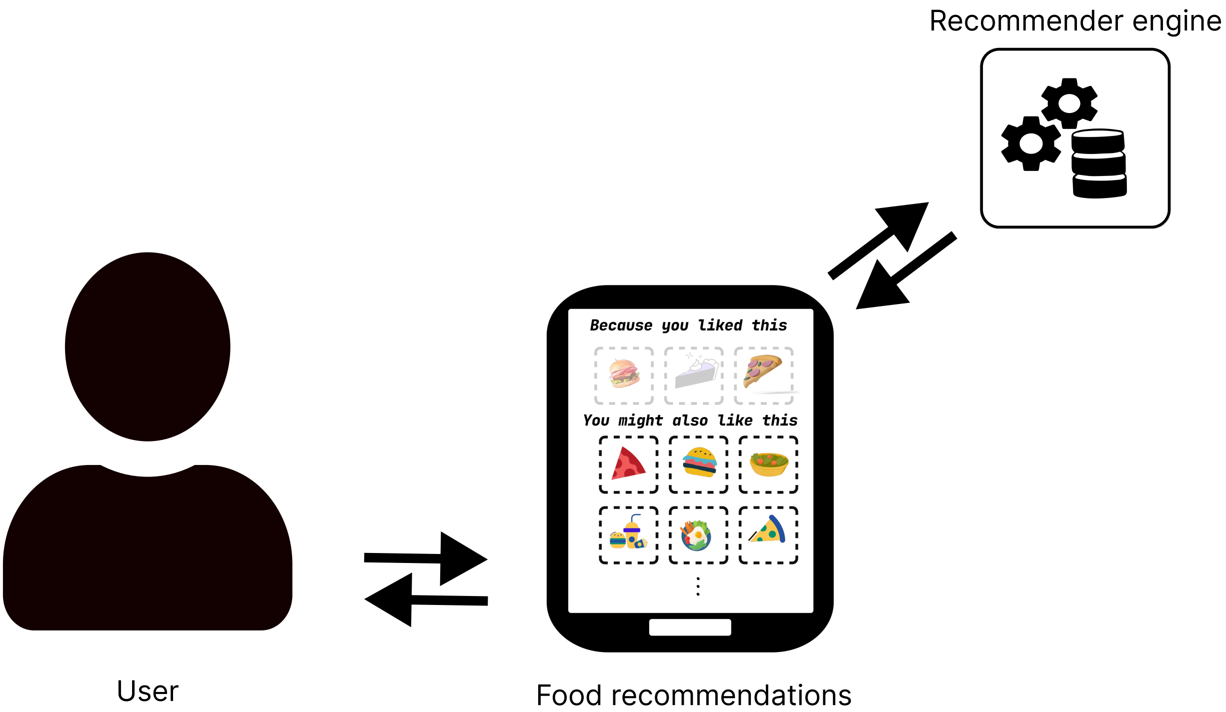



Figure 1.1: An illustration of a user engaging with a traditional food recommender system.

In relation to the food consumption these days, it is noticeable that there has been an increase of lifestyle-related illnesses (Warensjo et al., 2010), such as diabetes and obesity [1], which are the cause of many chronic diseases (Haththotuwa et al., 2020). This problem can be improved by making appropriate food choices and healthy lifestyle.

[1]Global Obesity, WHO

On the other hand, existing recommender systems have been shown to enhance the user experience with food, not only by addressing the inherent complexity of food choice but also by facilitating the discovery of new culinary options (Nguyen, 2016; Starke et al., 2023; Alshahrani et al., 2024). For instance, Allrecipes [2], the most widely used recipe website worldwide, provides trusted recipe recommendations to over 60 million users each month. In addition, to more than 15 million users actively contribute weekly. Notably, the search term "what to eat" has reached its highest levels in the past decade [3], highlighting the growing demand for personalized food suggestions, underscoring the growing demand for personalized food suggestions.

A major concern with food recommender systems is their tendency to prioritize user preferences and engagement over nutritional values (Lin et al., 2014; Teng et al., 2012; Freyne and Berkovsky, 2010b), often leading to unhealthy food suggestions (Trattner et al., 2018). By relying on user past behaviors and popular trends (Trattner et al., 2019), these systems may reinforce poor eating habits rather than promoting balanced and healthy choices (Dickinson et al., 2018; Rokicki et al., 2018; Camargo et al., 2022).

To address this challenge, various approaches have been developed to integrate nutritional needs into recommendation systems (Yang et al., 2017), while others incorporate health-aware dimensions (Jiang et al., 2019). However, research indicates that online recipe platforms predominantly feature unhealthy recipes, and many recommendation algorithms inadvertently reinforce this bias (Trattner et al., 2019).

A promising strategy to counteract this issue is post-filtering (Trattner and Elsweiler, 2017a), which refines recommendations to promote healthier options (Trattner and Elsweiler, 2017b; De Croon et al., 2021). Yet, this approach presents a trade-off: while it steers users toward better dietary choices, it may also diminish engagement and satisfaction, as health-conscious suggestions are often perceived as less appealing (He et al., 2024; Pecune et al., 2022; Starke et al., 2023). This raises a critical question:

**How can we guide users toward healthier food choices without compromising personalization?**

Digital nudges present a promising approach to this challenge. These nudges involve subtle design interventions within the user interface that can influence user behavior in a digital choice environment (Weinmann et al., 2016). By targeting the presentation phase of recommender systems, digital nudges could potentially be integrated into food recommenders (Elsweiler et al., 2022), helping to guide users toward informed choices while still maintaining personalization.

Digital nudges are grounded in nudge theory, introduced by Thaler and Sunstein (2009), and have been shown to improve offline decisions within several domains (e.g., health, wealth). At its core, nudging is grounded in two primary concepts: choice architecture and behavioral paternalism. Choice architecture refers to the context or environment in which individuals make decisions, for instance, a GPS system that guides users toward a destination can be considered a form of a nudge. Behavioral paternalism, on the other hand, emphasizes that nudges should be designed to support individuals in making more informed and beneficial

[2] `Allrecipes.com`

[3] Food Search Trend, Google Trends

choices than they might make in the absence of such interventions (Thaler, 2018). In this sense, the ultimate goal of a nudge is to serve the best interests of the user.

While digital nudging and recommender systems both aim to guide user decisions, their integration has received limited attention in the literature (Jesse and Jannach, 2021). However, preliminary studies suggest that the incorporation of digital nudges into recommendation systems may significantly enhance user engagement (Chiam et al., 2024), informed decisions (Karlsen and Andersen, 2019), and overall system effectiveness (Chiam et al., 2024), in various domains (Haque et al., 2023; Sobolev, 2021; Alves et al., 2024).

Despite the growing interest in digital nudging and recommender systems, their combined application in the domain of food recommendations remains particularly scarce (Forberger et al., 2024). This is notable given the strong potential of food recommender systems to influence daily dietary habits (Elsweiler et al., 2022; Tran et al., 2021). Existing studies have largely focused on nutritional optimization or preference modeling, often overlooking the role of subtle interface-level interventions that could encourage healthier food choices without compromising personalization or user experience. This highlights the need for further exploration into how digital nudges can be effectively designed and integrated into food recommender systems to support informed and healthier decision-making. Given the relatively low technical barrier to implementing digital nudges and their potential to meaningfully influence food choices, further investigation is both timely and necessary. For example, a digital nudge integrated into a food recommender system might involve highlighting healthier options with a subtle visual cue, such as a green border or a tag. This could encourage healthier decisions without explicitly changing the recommendation algorithm itself. An illustration of such a nudge is shown in Figure 1.2.

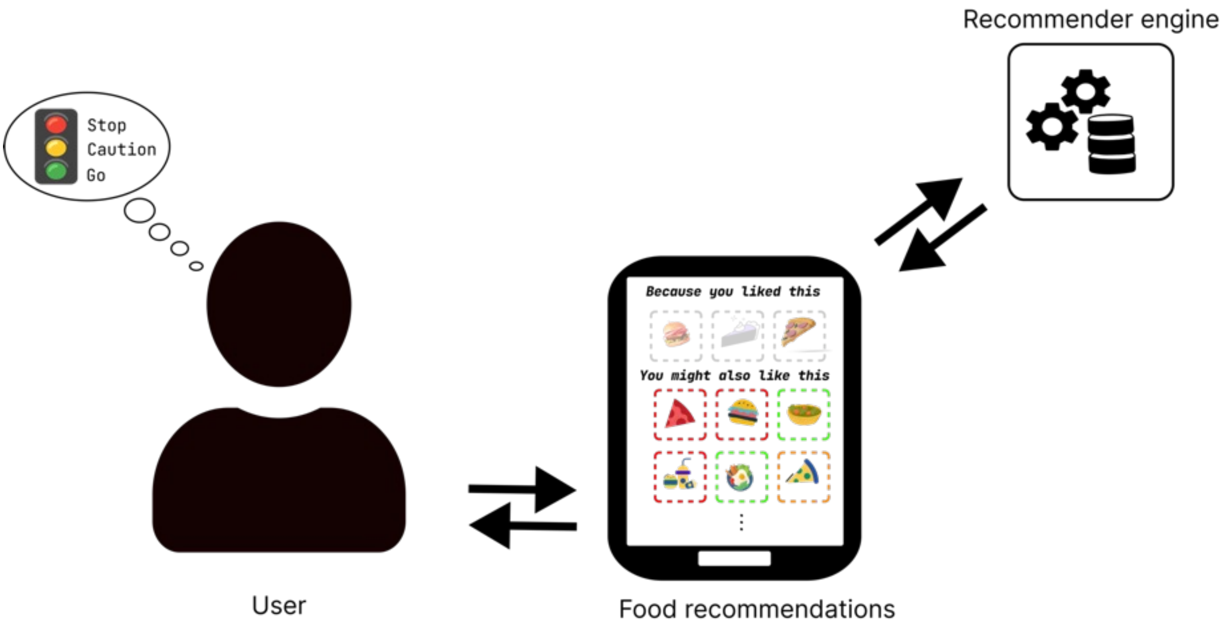



Figure 1.2: An illustration of a user interacting with a food recommender system enhanced by digital nudges, where colored borders indicate the healthiness of each item: green border for the healthiest choices, orange for moderately healthy options, and red for less healthy items.

## 1.2 Problem Statement & Research Questions

Food recommender systems assist users in discovering food items they are likely to enjoy by narrowing the search space and personalizing results. However, many existing systems tend to prioritize popular recipes, which are often less healthy according to various nutritional metrics. As a result, these systems may unintentionally reinforce unhealthy eating habits.

In contrast, digital nudges have been shown to support behavior change by helping users make more informed decisions across several domains, including health and finance. Integrating digital nudges into food recommender systems offers a promising approach to encouraging healthier food choices without compromising personalization or user satisfaction.

Despite this potential, the combination of digital nudging and food recommender systems for support user in making healthier and informed food choices remains insufficiently investigated, and presents several open research questions.

The aim of this thesis is to explore the integration of digital nudges into food recommender systems to support users in making healthier and more informed food choices.

The following research questions (RQs) are addressed in this thesis:

- **RQ1: Which types of digital nudges, when applied to food recommender systems, effectively improve users' food choices?**
- **RQ2: How do digital nudges influence different aspects of the recommender system and the overall user experience in the context of food recommendations?**
- **RQ3: How can digital nudges be effectively integrated into personalized food platforms to support healthier decision-making?**

## 1.3 Objective and Contributions

The primary objective of this thesis is to investigate and evaluate the integration of digital nudges into food recommender systems, with the overarching aim of promoting healthier and more informed decision-making among users. As the demand for personalized food recommendations grows, there is an increasing need to ensure that these systems do not solely focus on preference satisfaction but also encourage better and healthier nutritional choices. This research addresses critical gaps in the current literature by systematically assessing the effectiveness of digital nudges when embedded within food recommendation contexts and by analyzing their impact on user behavior, preferences, and overall user experience. The outcomes of this research is detailed and disucsssed accorss the scientific manuscipts[4]. These efforts culminated in the following key contributions:

- **Categorizing and determining digital nudging techniques worth investigation within food recommender system**

[4] It is essential to clarify that the first manuscript (see Manuscript 1), which is a review fo the literature , does not constitute the primary scientific contribution of this thesis. Consequently, the following manuscripts are not derived from the findings of the first one.

This project provides a systematic categorization of digital nudging techniques most commonly addressed in the context of promoting healthier food choices within recommender systems. Building upon this foundation, the thesis identifies the techniques most suitable for integration and empirical evaluation within a user-centered food recommendation environments. Through an extensive review of interdisciplinary research across recommender systems and digital nudging, it establishes a curated set of strategies with strong potential to positively influence dietary decision-making. This structured categorization not only maps the current landscape of digital nudges but also lays the groundwork for the subsequent future design, implementation, and evaluation of nudging interventions. The contribution places particular emphasis on user-centric evaluation to guarantee that the proposed nudging strategies align with users' needs, foster meaningful behavior change, and enhance satisfaction within food recommender systems. In addressing these objectives, the following contribution responds to the first research question (see RQ1: 1.2) and is presented in the first research manuscript (see Manuscript 1, 4.7.3).

- **Design and evaluation of new digital nudges within food recommender systems for healthier food choices.**

  The integration of digital nudges into recommender systems has garnered increasing attention across various domains, due to the potential of both technologies to facilitate positive behavioral change (Jesse and Jannach, 2021; Jesse et al., 2021). For instance, research has demonstrated that visually appealing images can enhance the healthiness of user choices, while the layout of recommendation lists significantly influences user experience (Starke et al., 2021c, 2022b). Building on these insights, this thesis presents the design, implementation, and integration of novel digital nudges within functional food recommender system prototypes, with the aim of supporting improved user decision-making and encouraging healthier food choices. The nudging techniques explored include food nutritional labels and explanatory nutritional messages, which are embedded within different types of food recommender systems. These systems are categorized into non-personalized and personalized approaches, encompassing collaborative filtering, content-based filtering, conversational interfaces, and knowledge-based systems. By integrating these nudges, the thesis contributes into how each system type can be enhanced to support informed and healthier food choices.

  The evaluations conducted assess the effectiveness of the selected nudges in two primary areas: their ability to help users make more informed decisions about their food choices, and their impact on the overall user experience. The analysis goes beyond simply measuring decision-making efficiency, extending to user satisfaction, engagement, and the perceived helpfulness of the nudging strategies. This comprehensive evaluation aims to provide insights into how different recommender system architectures, combined with targeted nudges, can influence both the cognitive aspects (e.g., perceived effort) and emotional aspects (e.g., choice satisfaction) of the user's decision-making process, ultimately enhancing their overall interaction with food recommendation platforms. Accordingly, this contribution addresses the second research question (see RQ2: 1.2) and is discussed across several manuscripts (see manuscripts: Manuscript 2, Manuscript 3, Manuscript 4, Manuscript 5, 4.7.3)

- **Evaluation of preference elicitation methods and user experience within food Recommender systems integrating digital nudges**

  The user experience of a recommender system plays a critical role in assessing its effectiveness in fulfilling users' functional, cognitive, and emotional needs (Knijnenburg et al., 2012; Pu et al., 2011). Furthermore, user experience is strongly correlated with the system components such as preference elicitation methods and the user behaviors, as the extent to which a system facilitates or supports behavioral change is influenced by its ability to assist users in efficiently identifying and selecting items that meet their goals and preferences (Tintarev and Masthoff, 2007; Pommeranz et al., 2012). In the context of food recommender systems, the impact of integrating digital nudges on user experience has received relatively limited attention (Trattner and Elsweiler, 2017b; Elsweiler et al., 2022). Addressing this gap, the present thesis investigate how digital nudges, integrated into food recommender systems with various preference elicitation techniques, influence key aspects of user interaction, such as satisfaction and overall experience. Specifically, it examines how the interaction between system components, whether objective elements like preference elicitation performance or subjective factors like user perceptions, significantly impacts the user experience. Using a user-centric evaluation framework, this work provides a holistic understanding of how different nudging techniques, preference elicitation strategies, and user knowledge shape the overall user experience, including the role of the user in evaluating these components. It offers valuable insights into the interplay between recommendation approaches, digital nudges, and user knowledge, and how these factors influence user behavior and decision-making. This contribution helps answer the second and third (see RQ2, RQ3: 1.2) research questions, and is specifically discussed in the Manuscript 3, Manuscript 4, Manuscript 5, Manuscript 6.

- **Insights into the integration of nudges in personalized food platforms**

  This thesis advances our understanding of user behaviors within personalized systems enhanced by digital nudges. The research paves the way for integrating nudging techniques into recommender systems by identifying which types of nudges hold the most promise for further exploration in the context of food recommendations. Furthermore, the user-centric evaluations provide a theoretical foundation for understanding how the interacting components of recommender systems shape and influence the user experience. The thesis also explores new and effective areas beyond recommender system accuracy, focusing on better reflecting user preferences in personalized recommendations. It emphasizes addressing biases in those preferences and harnessing interface elements and the presentation phase of recommender systems to guide and positively influence user choices towards healthier and more informed decision-making. Moreover, the contribution highlight the relevant aspect that effect user decision within a food recommender system, and design element influence their decision for instance food image factors (Manuscript 6, 4.7.3). The insights from both the literature and the user-centric experiments contribute to answering the overall thesis research question (see Section: 1.1), are presented and discussed across the published research manuscripts (see Manuscripts, 4.7.3).

# Chapter 2

# Background

This chapter provides the reader with a comprehensive exploration of the theoretical foundations of recommender systems and their underlying approaches. It also examines the principles behind digital nudges and various nudging techniques, followed by an overview of how recommender systems and digital nudges can be integrated to enhance decision-making.

## 2.1 Recommender Systems

Recommender systems, as the name suggests, are systems designed to provide recommendations and suggestions to users. The field of recommender systems has emerged from various disciplines, including cognitive science (Rich, 1979), approximation theory, information retrieval (Salton, 1989), and consumer choice modeling (Wind and Lilien, 1993). Since the mid-1990's, it has evolved into an independent research domain, primarily focusing on solving problems related to rating estimation and prediction.

The first paper on recommender systems conceptualized them as an assimilation of social relationships, where people receive suggestions from peers and like-minded individuals for video recommendations (Hill et al., 1995). This approach was later termed collaborative filtering. Research interest in this field has led to various formulations of recommender systems, of which we adopt the following definition proposed by Burke et al. (Burke, 2002):

**Definition**

A recommender system is any system that produces individualized recommendations as output or guides the user in a personalized way to interesting or useful objects in a large space of possible options.

Recommender systems operate on data sources that comprise three sets (Ricci et al., 2010). The first is the item set, which includes the objects being recommended. These items can vary in complexity and value, where a positive value indicates relevance and usefulness to the user, while a negative value suggests an unsuitable recommendation, often resulting from a suboptimal selection. The second is the user set, representing the individuals who interact with the system. To enhance personalization, recommender systems (RSs) utilize various types of user information, with the choice of features largely dependent on the recommendation technique employed. The third is the transaction set, which records the interactions between users, items, and the system as a whole. Transactions function as log-like data that capture essential information generated through human-computer interactions, enabling the system to refine and improve recommendations over time.

Recommender systems are designed to model user interactions with items, addressing the challenge of identifying the most relevant items for users, and conversely, determining the users most likely to engage with specific items. This problem can be mathematically formulated as follows:

**Formulation**

Let $\mathbb{U}$ denote the set of all users and $\mathbb{I}$ the set of all possible items.
Let $f$ represent the utility function that measures the suitability of item $i$ to the users $u$, i.e., $f : \mathbb{U}\mathbf{x}\mathbb{I} \rightarrow \mathbb{R}$ , where $\mathbb{R}$ is totally ordered set.
For each user $u$ in $\mathbb{U}$, the objective is to select an item $i'$ in $\mathbb{I}$, that maximize the user's utility function such as:

$$\forall u \in \mathbb{U}, i'_{\mathbb{I}} = argmax f(u, i) \tag{2.1}$$

To implement their core function, recommender systems employ appropriate approaches based on the source data, context, and application domain. Traditionally, these approaches fall into four main categories (Adomavicius and Tuzhilin, 2005; Ricci et al., 2010): collaborative filtering, content-based filtering, knowledge-based methods, and hybrid techniques.

### 2.1.1 Collaborative filtering

Collaborative filtering (CF) recommendation systems generate recommendations by leveraging user-item interactions (Herlocker et al., 2000). This approach relies on the assumption that users with similar preferences will rate items similarly. The core process involves predicting ratings for items based on past user interactions, which are stored in a user-item matrix. This matrix consists of users as rows, items as columns, and their respective ratings as values. The system identifies users with similar preferences and estimates ratings for a new user by analyzing the interactions of similar users. Formally, the utility function (eq 2.1) $f(u, i)$ for item $i$ and user $u$ is estimated based on the utility values $f(u_j, i)$ assigned to $i$ by similar users $u_j$ in $\mathbb{U}$.

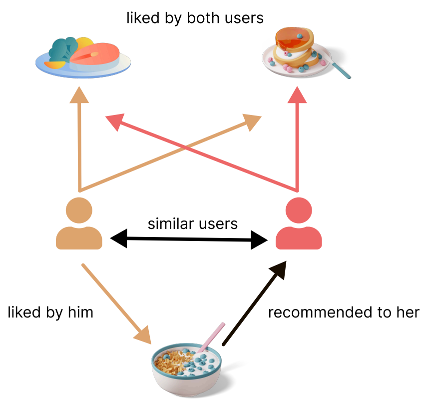



Figure 2.1: Collaborative recommendations example

For example, in food recommendation applications, a CF system predicts a user's expected rating for a recipe by identifying similar users, those who have demonstrated comparable preferences through past ratings. The system then recommends recipes that have been highly rated by these similar users. Figure 2.1 shows an example of a collaborative recommendation between two user.

Collaborative filtering systems can be categorized into memory-based and model-based approaches (Chen et al., 2018).

#### 2.1.1.1 Memory-based CF

Memory-based collaborative filtering algorithms operate on the entire user-item rating matrix to compute similarities and identify items highly rated by similar users. Memory-based recommendation systems can be further categorized into user-based and item-based approaches. Figure 2.2 shows the process of memory-based CF based on user and items. User-based

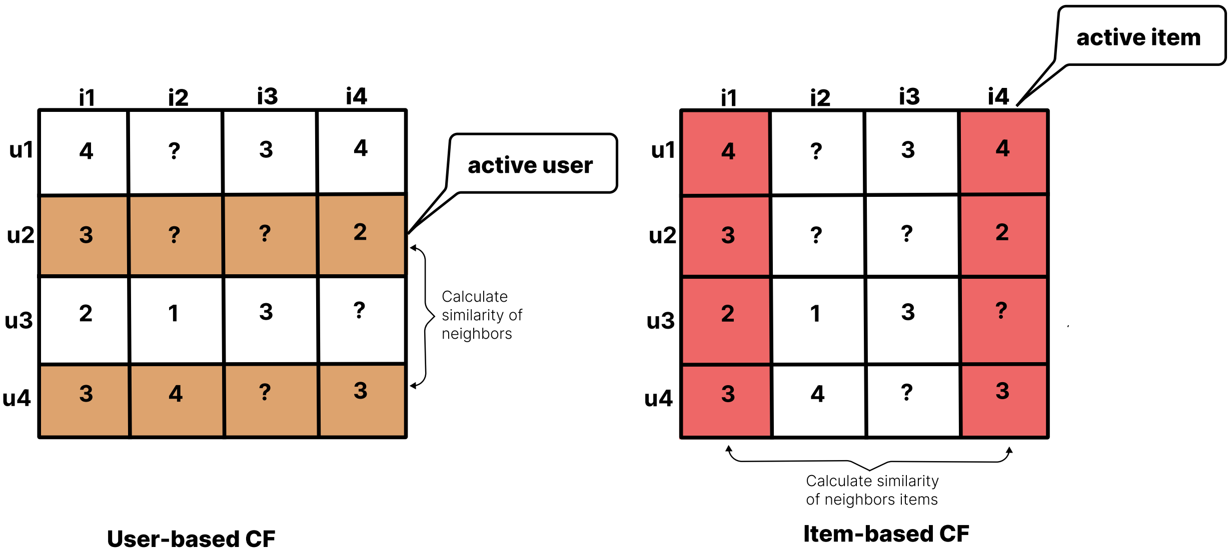



Figure 2.2: The process of user-based and items-based collaborative-filtering (CF) techniques.

methods operate on the assumption that users with similar historical ratings share similar interests. To predict a target user's missing ratings for specific items, these methods leverage the ratings provided by similar users for those items. First, similarity functions (e.g., cosine similarity, person similarity) are used to calculate the similarity between the active user and other users. Based on these similarity values, a set of neighboring users is selected. Finally, the active user's ratings are predicted by leveraging the historical preferences of these similar neighbors (Adomavicius and Tuzhilin, 2005).

Item-based approaches follow a similar process to user-based recommendations. First, item similarities are computed based on the user-item rating matrix. Using these similarity values, the algorithm selects a set of neighboring items. Finally, unknown ratings for a target item are predicted based on the ratings of its similar neighbors, generating a recommendation list (Schafer et al., 2007). In general, collaborative filtering recommender algorithms are executed according to the following steps:

- Compute similarity: as an example of cosine similarity being one of the most commonly used methods. This method calculates a similarity metric based on users' rating vectors patterns for items or vice versa, as illustrated in Equation 2.2.

$$sim_{u,v} = cos(\vec{r_u}, \vec{r_v}) = \frac{\vec{r_u}.\vec{r_v}}{||\vec{r_u}||_2 \times ||\vec{r_v}||_2} = \frac{\sum_{i \in I_{uv}} r_{ui}.r_{vi}}{\sqrt{\sum_{i \in I_u} r_{ui}^2}\sqrt{\sum_{i \in I_v} r_{vi}^2}} \tag{2.2}$$

  where $sim_{uv}$ represents the similarity between users $u$ and $v$, $\vec{r_u}$,$\vec{r_v}$ represent the ratings vector of $u$ and $v$, $||\vec{r_u}||$, $||\vec{r_v}||$ are second-norm of $u$ and $v$, respectively, $r_{ui}$ and $r_{vi}$

represent the ratings of $u$ and $v$ on the item $i$, $\mathbb{I}_u$ and $\mathbb{I}_v$ are sets of items rated by users $u$ and $v$ respectively, and $\mathbb{I}_{uv}$ represents the set of items commonly rated by both $u$ and $v$. The same procedure was applied for computing similarities between items.

- Finding neighbors: Nearest neighbors refer to the users or items most similar to the active user or item, based on similarity measures. The selection is typically guided by a predefined similarity threshold, ensuring that only the most relevant neighbors are considered.
- Predict ratings: The rating prediction phase builds on the previous steps and involves estimating the potential rating of an active user for a new item or predicting item's rating by a new user. The prediction is computed as follows:

$$\hat{r}_{ui} = \overline{r}_u + \frac{\sum_{v \in \mathbb{N}_u} sim_{u,v}(r_{vi} - \overline{r}_v)}{\sum_{v \in \mathbb{N}_u} |sim_{uv}|} \tag{2.3}$$

Where $\mathbb{N}_u$ denotes the set of neighbors of user u, and $\overline{r}_u$, $\overline{r}_v$ represent the average ratings from $u$ and $v$ respectively.

#### 2.1.1.2 Model-based CF

Model-based CF algorithms leverage user rating data to develop predictive models capable of estimating user preferences. A widely adopted approach for implementing model-based CF is latent factor modeling, which decomposes the user-item rating matrix into two low-rank matrices: the user feature matrix and the item feature matrix (Chen et al., 2018). Among various matrix factorization techniques, Singular Value Decomposition (SVD) has emerged as a particularly effective method due to its ability to uncover latent patterns within the data (Zhou et al., 2015; Sarwar et al., 2000). SVD-based approaches have been extensively studied and widely applied in recommendation systems, demonstrating superior performance in terms of accuracy and scalability. Additionally, SVD effectively addresses key challenges in collaborative filtering, including large-scale data processing, rating matrix sparsity, and the cold-start problem (Bokde et al., 2015). The popularity of SVD surged following its instrumental role in the Netflix Prize competition (Bennett et al., 2007), where it significantly advanced the state of recommendation algorithms. While alternative factorization techniques such as Non-Negative Matrix Factorization (NMF) have been widely explored (Huang et al., 2016), SVD remains the cornerstone of latent factor models in recommendation systems (Zhao, 2024; El Majjodi et al., 2020).

**2.1.1.2.1 Singular Value Decomposition (SVD)** Singular Value Decomposition (SVD) is a matrix factorization technique widely utilized in model based recommendation systems (Ma, 2008). It decomposes a user-item interaction matrix into latent feature representations of both users and items, enabling the prediction of missing values by leveraging these learned factors.

Assuming that user/item sparse rating matrix denoted by $R$, SVD decompose that matrix into $U$, $S$ and $V$, such as:

$$R_{m.n} = U_{m.m} S_{m.n} V^T_{n.n} \tag{2.4}$$

where:

- $U$ (user feature matrix) represents users in terms of latent factors.
- $S$ (diagonal matrix) contains singular values, ranking the importance of these factors.
- $V^T$ (item feature matrix) represents items in the same latent space.

These factors capture meaningful patterns, such as which users have similar tastes and which items are alike. By keeping only the most significant singular values, SVD reduces noise and improves predictions. Once trained, SVD reconstructs an approximation of $R$ and produce $\hat{R}$ with predicted missing ratings. This allows recommendation systems to suggest items even for users who have rated very few items (Chen et al., 2018).

### 2.1.2 Content based

A content-based recommender system (CB) is a recommendation approach that identifies items similar to those that a user has previously liked. Primarily rooted in information retrieval (IR), CB methods analyze item attributes and user content to generate personalized recommendations. The core of CB filtering lies in evaluating object metadata, leveraging both item characteristics and user-related profile attributes to identify meaningful similarities. Figure 2.3 presents the key component of a (CB) recommender system. In general (Lops et al., 2011), three main components interact to generate recommendations within a content-based framework:

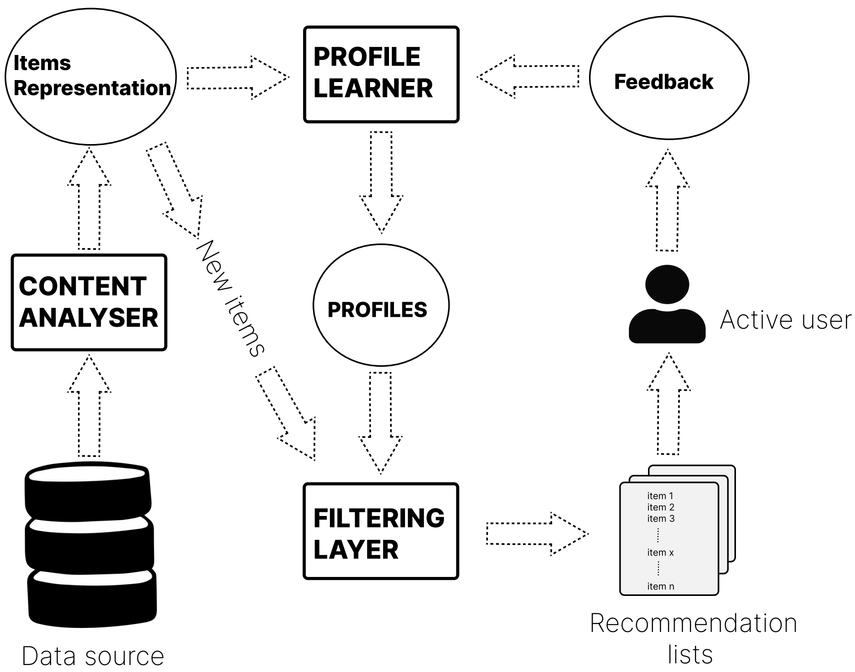



Figure 2.3: The key component of a content-based filtering approach, where the content analyzer generates item representations, which are then matched with learned user profiles through a filtering algorithm, to generate personalize recommendation.

- **Content Analyser**: This component is responsible for the preprocessing of data sources to retrieve relevant information about items within the database. The primary steps include content representation and the structuring of information sources (e.g., text, web pages, news articles, movies, food items) to extract features essential for subsequent processing. The feature extraction phase involves transforming data from its original information space into latent feature spaces, and extracting data into a machine-readable format (e.g. keywords vector, bag, or words). Using different techniques depending on the data types (e.g., TF-IDF(Qaiser and Ali, 2018), Word2Vect (Church, 2017), BERT (Koroteev, 2021), etc)

- **Profile Learner**: This core component gathers user data to extract preferences and construct a user profile. The profile, typically represented as a model, encapsulates user characteristics, interactions, and preferences, including liked and disliked items. Statistical and machine learning techniques are applied to structure and produce the user profile model, preparing it for effective preference extraction and analysis (Rocchio, 1971).

- **Filtering model**: The core function of this component is to match the user profile with relevant items from the available dataset. This process takes place within a feature space, where both user preferences and item attributes are represented as vectors. The distance or similarity between the user profile and the item representations is computed, often using metrics such as cosine similarity or euclidean distance. These calculations help quantify how closely an item aligns with the user's preferences. Based on these computed distances, the filtering model generates a ranked list of recommendations, with items that are most similar to the user's profile appearing at the top.

In content-based (CB) recommendation systems, several key components interact to generate personalized recommendations. The content analyzer plays a fundamental role in this process by extracting and processing information from various sources using advanced information retrieval techniques (Van Meteren and Van Someren, 2000; Mitra and Chaudhuri, 2000). This component constructs a structured dataset where items are represented as feature vectors, capturing essential attributes that define each item. These structured representations are stored in the item representation space, serving as the foundation for similarity-based recommendations (Philip et al., 2014).

The profile learner, then maintains and continuously updates user profiles. The user profile comprises explicit and implicit information, including demographic details, stated preferences, and historical interactions with items. User interactions such as likes, dislikes, comments, shares, and reviews contribute valuable feedback that refines the profile over time (Holte and Yan, 1996). By leveraging this feedback, the profile learner dynamically updates user profile models, ensuring that recommendations align with evolving user interests and preferences (Lops et al., 2010; Kaya, 2018). Each active user is thus associated with a personalized profile that adapts based on their engagement with the system.

The final step carried by the filtering component which responsible for generating recommendations through evaluating the relevance of new or existing items to the active user. Given a newly introduced item representation, this component predicts its potential relevance by comparing the item's feature representation with the user's profile data (Zanardi and Capra, 2008). Typically, the filtering component employs ranking strategies to prioritize items based on their similarity to the user's preferences (Yao, 1995). These strategies may involve similarity measures such as cosine similarity, Euclidean distance, or machine learning-based ranking models to ensure that the most relevant content is surfaced for the user (Pazzani and Billsus, 2007).

### 2.1.3 Knowledge based

Knowledge-based (KB) recommendation systems utilize structured domain knowledge about users and items to generate precise and personalized recommendations. Unlike collaborative or content-based approaches, KB recommenders explicitly gather user preferences and requirements for each item (Burke, 2000). These systems incorporate interactive user feedback to refine suggestions, ensuring alignment with user needs while supporting broader item exploration. The retrieval process is guided by well-defined knowledge attributes that describe the characteristics of a given item or service (e.g., car recommendations), allowing for more accurate and relevant recommendations (Aggarwal, 2016).

Knowledge-based (KB) recommender systems are employed when users have limited or no prior interaction with the system, allowing them greater control over the recommendation process. This control stems from the need to specify detailed requirements, particularly in complex and dynamically evolving domains. These systems are especially suitable contexts such as (Burke, 2000; Aggarwal, 2016):

- When users explicitly specify their requirements through an interactive component of the system.
- When obtaining ratings for a specific type of item is challenging due to the complexity of attributes and the wide range of available options.
- When items are time-sensitive, as ratings and reviews may become outdated due to the continuous evolution of products and services.

Knowledge-based (KB) recommendation systems can be categorized into two types based on user interaction methodology and the underlying knowledge structure used to facilitate the interaction. The primary categories are:

- *Constraint-based recommender systems*: In this approach, users specify their requirements and constraints using item attributes. Additionally, domain-specific rules, defined by system owners and domain experts, are presented to guide the matching process between user preferences and available products or services (Felfernig et al., 2015). If too few results are returned, users may relax certain constraints; conversely, if too many results are returned, they may refine their criteria. This interactive search process continues until the user reaches a satisfactory outcome. For example, in a car recommendation system, a user might specify that only cars produced in the last five years should be displayed.
- *Case-based recommender systems*: In case based recommenders, users drive the recommendation process by specifying a target or anchor point (Jannach et al., 2010). To retrieve items similar to the user-defined target, similarity metrics are applied to item attributes. These metrics are typically defined by domain experts to ensure relevance and accuracy. The retrieved results can then be iteratively refined, with users modifying the target criteria interactively to explore alternative recommendations.

In summary, knowledge-based (KB)) recommender systems offer a highly effective approach to personalized recommendations by combining domain-specific knowledge with user inputs. These systems stand out by incorporating explicit user preferences and constraints, enabling precise and tailored recommendations. By integrating expert knowledge with user-defined criteria, they foster an interactive, informed recommendation process, enhancing the overall user experience. These systems are especially valuable in domains where user input and domain expertise are crucial to decision-making.

## 2.2 Recommender system evaluation

The effectiveness of a recommender system can be evaluated using various metrics, which depend on the type of filtering algorithm employed. Statistically, several metrics are available to assess the performance of recommender algorithms (Ricci et al., 2010). Accuracy, for instance, refers to the proportion of correctly predicted recommendations out of the total possible recommendations.

Common statistical evaluation metrics include Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), both of which compare predicted ratings with actual user ratings. MAE, a widely used metric, measures the average deviation of predicted ratings from the actual user ratings and is computed as follows (Claypool et al., 1999):

$$MAE = \frac{1}{N} \sum_{u,i} |\hat{r}_{u,i} - r_{u,i}| \tag{2.5}$$

Where $\hat{r}_{u,i}$ is the predicted rating for user $i$ on item $i$, $r_{u,i}$ is the actual rating and $N$ is the total number of ratings on the items set. A lower (MAE) indicates higher accuracy in predicting user preferences. Similarly,(RMSE) places greater emphasis on large absolute errors and is computed as follows (Cotter and Smyth, 2000):

$$RMSE = \sqrt{\frac{1}{N} \sum_{u,i} (\hat{r}_{u,i} - r_{u,i})^2} \tag{2.6}$$

Beyond statistical measures, decision-support accuracy metrics such as precision and recall provide additional insights (Ricci et al., 2010). Precision quantifies the proportion of recommended items that the user has consumed, reflecting the rate of relevant recommendations. Recall, on the other hand, is defined as the proportion of consumed items in the recommendation list relative to the total number of items the user has consumed. Researchers often denote these metrics as *precision@N* and *recall@N*, where N represents the size of the recommendation list (Isinkaye et al., 2015).

$$Precision = \frac{Correctly\ recommended\ items}{Total\ recommended\ items} \tag{2.7}$$

$$Recall = \frac{Correctly\ recommended\ items}{Total\ useful\ recommended\ items} \tag{2.8}$$

Additionally, the F-measure combines precision and recall into a single metric, simplifying the comparison of algorithms across datasets. It is computed as follows:

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{2.9}$$

Most statistical measures for evaluating recommender systems are applied in offline settings, focusing on accuracy and error metrics to assess algorithmic performance (Shani and Gunawardana, 2011). However, after developing and deploying a recommender system, online evaluation becomes essential for measuring how effectively the system delivers recommendations to active users and how it influences user behavior (Chen and Liu, 2017). The literature presents several frameworks that assess recommender systems within a user-centric perspective (Knijnenburg et al., 2012; Pu et al., 2011), as statistical metrics alone do not fully capture the broader impact of recommender algorithms on user engagement, satisfaction, and overall user experience (Knijnenburg and Willemsen, 2015).

Despite extensive research and evaluation across different domains and settings, there is no definitive evidence that a single algorithm consistently outperforms all others. Performance varies based on factors such as user behavior, data sparsity, and the specific evaluation metrics used. Consequently, selecting an appropriate algorithm often requires empirical testing and adaptation to the given recommendation scenario.

## 2.3 Food Recommender System

Food recommender systems assist users in discovering relevant food options by analyzing past behaviors, similar to traditional recommender systems. Their significance is widely recognized, particularly due to their impact on users' health. These systems play a crucial role in helping individuals make informed food choices, whether for inspiration, meal planning, grocery shopping, or restaurant selection. Unlike other recommendation domains, food recommendations is highly complex, shaped by personal taste, health considerations, cultural influences, and contextual factors (Trattner and Elsweiler, 2017b). Moreover, poor dietary decisions contribute to global health issues such as obesity and diabetes, highlighting the importance of recommendation systems for the food domains.

Food recommender systems present a context-dependent challenge that requires considering multiple factors throughout the recommendation process. Generating accurate and relevant suggestions for users involves accounting for various players inherited from the broader field of recommender systems and can modeled following the formulation in (see eq 2.1). These key players include:

- User: The active individual seeking food recommendations, either for personal use or on behalf of a group. For instance, a user may create a profile to receive recommendations tailored to their family or household. In such cases, food recommender systems must account for multiple preferences and find a balanced recommendation that satisfies the entire group.
- Items: The food items that can be recommended, categorized into two main types: basic foodstuffs and food products. Basic foodstuffs include ingredients such as tomatoes, carrots, and sugar, which can be grouped into broader categories like groceries or recipes containing these ingredients. Food products, such as yogurt, are commercially

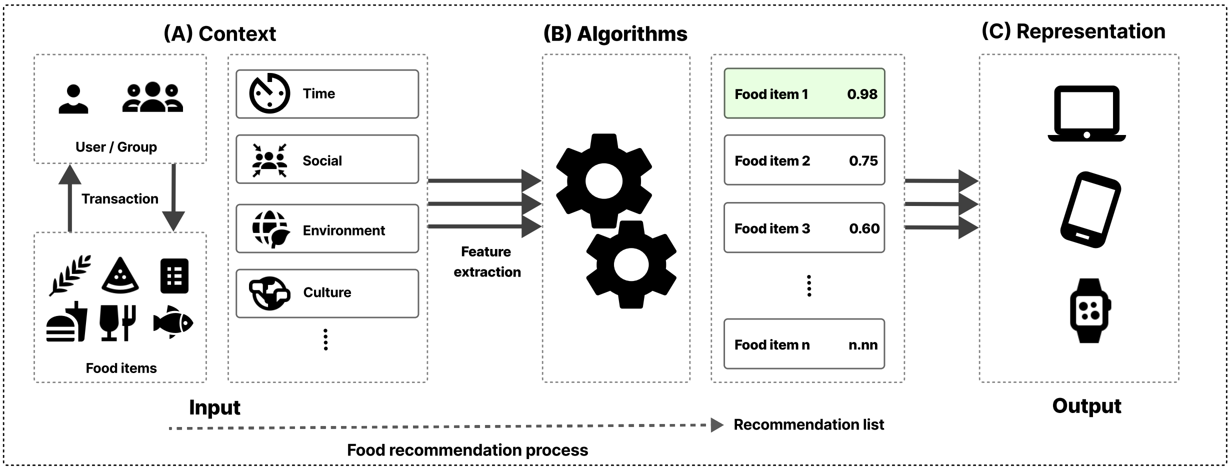



Figure 2.4: Traditional food recommender system framework, wherein inputs encompassing food attributes and user factor are processed by a recommendation algorithm to identify and present the most relevant food options to the active user.

available items with predefined compositions. Additionally, a third category meals consists of various basic ingredients and food products combined into a complete dish.

The food recommender system is a process in which multiple interacting factors contribute to facilitating the transaction between users and food items, ultimately generating relevant recommendations for active users (Elsweiler et al., 2022). The recommendation generation process within a food recommender system is illustrated in Figure 2.4. The main factors include:

- **Context** : Food choices are strongly influenced by context, which encompasses all factors associated with food consumption. In the food recommendation framework, context includes both internal factors, such as ingredients and nutritional properties, and external factors, such as availability and cost (Poelman and Steenhuis, 2019). Additionally, user-specific aspects, whether an individual or a group, along with environmental and social settings, further shape food preferences (Maia and Ferreira, 2018). Within the food recommendation process, context functions as a critical input feature, encapsulating user characteristics, food item attributes, environmental conditions, and historical user-item interactions.

- **Algorithms** : The algorithmic component is the core of a food recommender system, responsible for processing user and item features, along with other contextual inputs, to determine the most relevant recommendations for the active user (Elsweiler et al., 2022). A variety of algorithms and approaches are used (see Section 2.1), with the optimal choice being highly dependent on the specific problem context, dataset characteristics, and application constraints (Trattner and Elsweiler, 2017b). The primary objective of these algorithms is to learn and model user preferences, constructing a user profile that facilitates the prediction of food preferences and the generation of ranked recommendation lists.

- **Representation**: After the food recommendations are generated for the active user, the way these recommendations are presented plays a crucial role in shaping the user's final decision (Trattner et al., 2018). The user interface and presentation environment serve as the medium through which users interact with the recommended food items. Factors such as layout design, visual emphasis, ordering of recommendations, and the level of detail provided for each item can strongly impact user preferences and decision-making processes. The presentation phase is not merely a passive display of recommendations, it actively contributes to preference elicitation by shaping user perceptions and guiding their selections (Chen and Tsoi, 2011). For instance, arranging food items based on nutritional value, popularity, or personalized relevance can lead users toward healthier or more suitable options (Starke et al., 2022b). Similarly, the inclusion of images, ingredient lists, health labels, and user reviews can enhance engagement and affect choice confidence.

## 2.4 Digital Nudges

### 2.4.1 Thinking and Decision Making

The theory of human cognition, introduced in the 1970's, offers key insights into individual thought processes (Wason and Evans, 1974). It distinguishes between two modes of thinking: one that is fast, automatic, and effortless, and another that is slow, deliberate, and analytical. This distinction is well-supported by experimental evidence in cognitive psychology (Denes-Raj and Epstein, 1994; Epstein et al., 1996). Daniel Kahneman (Daniel, 2017) further refined this theory, labeling these modes as 'System 1' and 'System 2' thinking in his work Thinking, Fast and Slow. System 2, or controlled thinking, involves more conscious, effortful cognitive processing. It involves a higher degree of effort and control, often triggered by the necessity to solve a problem or make a decision requiring careful consideration and analysis. System 2 is typically associated with logical reasoning, analytical thinking, and deliberative mental tasks. On the contrary, System 1 is linked with emotions, intuition, and spontaneous judgments, responsible for quick decisions made with minimal deliberation (Loo, 2023). While most human behaviors are attributed to System 2 thinking, a substantial body of research indicates that System 1 dominates roughly 95% of the time, highlighting its pervasive influence on human cognition and behavior.

Relying solely on deliberate thinking and thorough reasoning for every daily decision can overwhelm the human brain, given the estimated 35,000 decisions individuals make daily (Roberts, 2025). Hence, there's a critical necessity for using System 1 thinking, which aids in handling the cognitive load during daily tasks and activities. However, for this system to operate swiftly and subconsciously, it heavily relies on shortcuts, also known as heuristics.

The term "heuristics" originates from the Greek word meaning "to discover." The concept was first introduced into psychological discourse by Nobel laureate Herbert Simon, an economist and cognitive psychologist, in the 1950's. Simon posited that while individuals aspire to make rational decisions, human judgment is constrained by cognitive limitations. In an idealized rational decision-making process, one would systematically evaluate all possible alternatives by assessing their respective costs and benefits (Vlaev, 2018).

However, decision-making in real-world contexts is bounded by various constraints, including the limited time available for deliberation and the finite amount of information accessible to an individual. Additional factors, such as cognitive capacity and the accuracy of perceptual judgments, further shape the decision-making process. Moreover, research from behavioral decision theory and consumer psychology suggests that preferences are not fixed, but rather constructed in the moment (Bettman et al., 1998). This notion of constructive consumer choice means that decisions are heavily influenced by contextual cues, such as how options are framed or ordered, what information is salient, or how much effort is required to choose.

Thaler and Sunstein (Sugden, 2009) related on the premise that individuals make a biased decision constrained by heuristics and context , which they would not have made if they had paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control. To address these cognitive and contextual limitations, they propose the role of a 'choice architect', an individual responsible for structuring the environment in which decisions are made to promote better outcomes (Moseley and Stoker, 2013).

### 2.4.2 Nudging and Digital Nudges

The concept of the choice architect is embedded in the term "Nudge", a theory based on facilitating design between decision problems and decision makers, so that individuals' choices align well with their interests. A nudge is defined as any aspect of the choice architecture that predictably alters people's behavior without forbidding any options or significantly changing their economic incentives. This term encompasses various techniques used to support people's behavior and overcomes decision biases and heuristics inherent in human thinking systems based on the choice architect. Nudges influence choice behavior through various mechanisms (Li and Chapman, 2013; Wildavsky, 2013), including (a) providing information (e.g., informing about energy consumption levels (Newell and Siikamäki, 2014)), (b) correcting misconceptions about social norms (e.g., nudging farmers into pro-environmental practices (Kuhfuss et al., 2016)), (c) modifying the salience of choices (e.g., making healthier food options more prominent in a cafeteria (Meeusen et al., 2023)), and (d) implementing default options (e.g., donation to charity (Goswami and Urminsky, 2016)). The core principle of nudge theory is to enhance the convenience or salience of the "better" choice for the decision-maker. A choice is considered "better" to the extent that it maximizes long-term health, wealth, and overall well-being.

In the digital age, choices and behaviors increasingly occur within digital environments, where decision-making is influenced by the design of online interfaces. Digital nudges, derived from nudge theory and popularized by Weinmann (Weinmann et al., 2016), utilize specific interface elements to subtly guide user choices and shape behavior in online contexts. By integrating behavioral science with information technology, digital nudges provide designers with tools to support users in making more informed and beneficial decisions (Caraban et al., 2019). These nudges can take various forms, such as adjusting the prominence of options, implementing default settings, or providing real-time feedback. Broadly, digital nudges can be categorized into three main types (Cadario and Chandon, 2020), each targeting different aspects of user interaction and decision-making:

- **Cognitive nudges** influence decision-making by structuring the presentation of information or choices to align with cognitive processes, such as heuristics and biases. These nudges aim to enhance decision quality by mitigating cognitive limitations and promoting rational judgment.

- **Affective nudges** leverage emotional responses to shape behavior. By incorporating cues that evoke specific emotions or associations, these nudges influence individuals' perceptions and decision-making processes, often enhancing engagement and motivation.
- **Behavioral nudges** directly modify the choice environment or decision architecture to guide user actions. By altering contextual factors, such as default settings, spatial arrangement, or timing of choices, these nudges facilitate desired behaviors with minimal cognitive effort.

Designing effective digital nudges relies on three key principles (Caraban et al., 2019; Schneider et al., 2018): technology, data, and experimentation. The first principle, technology, involves selecting the appropriate digital environment or platform for implementing the nudge, such as online food choice websites or wearable devices. Understanding the technological context ensures seamless integration into the user's decision-making process. The second principle, data, emphasizes the ethical and responsible collection of relevant information to personalize and optimize nudges. Choice architecture should identify and utilize necessary data in a way that benefits users while adhering to privacy and ethical standards. Finally, experimentation is crucial for assessing the effectiveness of digital nudges (Weinmann et al., 2016). Designers can leverage digital resources to test their applicability and measure the extent to which they influence behavior. By adhering to these principles, digital nudges can be thoughtfully designed to support users in making better decisions while maintaining ethical integrity.

In digital interface, such as recommender systems, choice architecture can becomes particularly powerful. Interface elements like option defaults, nutritional labels, or the visual prominence of healthier foods can all act as digital nudges. By aligning system design with how users actually think and choose, recommender systems can do more than just reflect preferences, they can help shape better ones, particularly in domains like food, health, and sustainability.

## 2.5 Food Recommender Systems and Digital Nudges

Recommender systems and digital nudges are both designed to influence user choices and behaviors. However, they differ in their primary mechanisms. Digital nudges primarily utilize subtle cues or prompts to guide decisions, whereas recommender systems focus on delivering personalized recommendations based on comprehensive user profiles.

Online food recommender systems employ straightforward approaches, utilizing user data to identify and recommend new items similar to those previously consumed. However, concerns have been raised about the healthfulness of the online food data, suggesting that personalized food recommender systems may not adequately support healthy decision-making (Trattner and Elsweiler, 2017a). Recently, recommender systems and digital nudges have been shown to impact decision-making (Starke et al., 2021c, 2022a). However, this combination has been less explored in the research literature (Jesse and Jannach, 2021). In the food domain, such a combination fosters several goals; recommender systems, on the one hand, personalize items to users using recommender techniques, while digital nudges serve as a cue to help the user choose healthier food items from the personalized item list.

The complete version of this section is discussed in our literature review, which discuss and analyses research efforts on integrating digital nudges with food recommender systems. Furthermore, the review highlights current work and outlines future research directions. The review can be found in Manuscripts (Manuscript 1, 4.7.3).

## 2.6 Conclusion

In this chapter, we have provided a comprehensive theoretical foundation to familiarize the reader with key concepts used throughout this thesis. We explored various approaches to recommender systems, with a particular focus on their application within the food domain. Additionally, we examined nudge theory and the main categories of digital nudges relevant to this research. Finally, we conducted a literature review to analyze the integration of recommender systems and digital nudges in supporting users in making healthier food choices.

In the next chapter, we shift our focus to a detailed explanation of the research methodology employed in this thesis. This includes a discussion of the research design, data collection methods, experimental setup, and analytical techniques used to investigate the integration of digital nudges with food recommender systems.

# Chapter 3

# Research Methodology

This chapter presents the unified methodological research framework adopted throughout this thesis. The study employs the Design Science Research (DSR) methodology, which is systematically structured to ensure uniformity and consistency across all research components. This approach facilitates the design and implementation of research ideas while enabling precise interpretation of results. Furthermore, (DSR) establishes clear and rigorous guidelines to enhance experimental reproducibility and support meaningful, robust comparisons across studies.

## 3.1 Introduction

A recommender system is designed for a specific domain, with its architecture, user interface, and recommendation techniques optimized for generating relevant and effective suggestions. Its development follows key principles, including problem definition, data acquisition, and iterative offline and online evaluation to enhance recommendation quality (Ricci et al., 2021; Jannach et al., 2010).

Our research aims to not only personalize food recommendations but also facilitate healthier choices through nudging techniques. A user-centric approach ensures the system is both effective and engaging, considering factors such as choice satisfaction, perceived effort, decision difficulty, understandability, and usability (Knijnenburg and Willemsen, 2015; McNee et al., 2006).

To evaluate its impact on behavior change, we adopt a Design Science Research (DSR) methodology, providing a structured framework for rigorous system development and assessment.

Design Science Research (DSR) methodology is rooted in engineering and follows a problem-solving paradigm. At its core, DSR aims to create innovations and establish the theoretical foundations, practices, technical tools, and artifacts that enable the systematic analysis, design, implementation, management, and utilization of information systems effectively and efficiently (Simon, 1988). The DSR strategy focuses on addressing information system challenges through a structured set of guidelines, emphasizing that knowledge and understanding of a design problem and its solutions emerge through the iterative process of building and applying artifacts (Hevner et al., 2004).

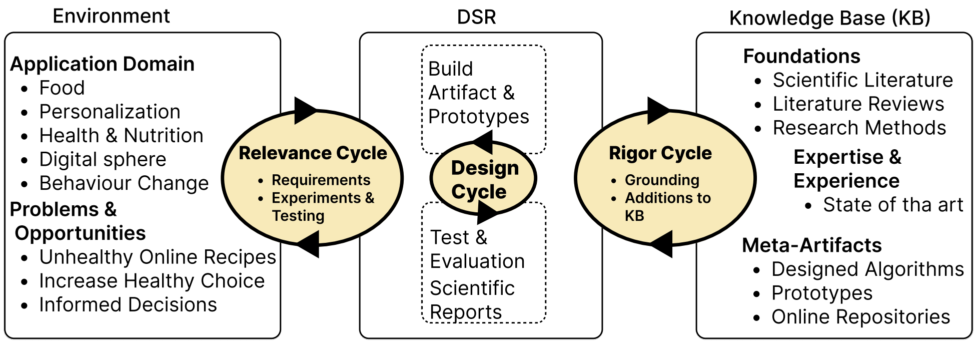



Figure 3.1: Design Science Research (DSR) Methodology: A cyclical process linking the relevance cycle, identifying the application domain and research opportunities, with the design cycle, where artifacts are developed and evaluated, and the rigor cycle, which provides the scientific and practical foundations for the research.

We adopted the information systems research framework (Hevner, 2007; Hevner et al., 2004), which emphasizes the design science through three main cycles. Figure 3.1 illustrates these cycles, outlining the research design methodology used to address our research questions.

**The Relevance Cycle** connects the contextual environment of the research project with Design Science activities, aiming to improve the environment through innovative artifacts. It defines the application domain (e.g., food, behavior change), identifies key problems and opportunities, and integrates environmental attributes into the research objectives.

A structured Design Science approach begins by identifying opportunities and challenges within the application environment (Hevner et al., 2004). In our research design, the environment component represents the application domain and its associated problems and opportunities, as presented in Figure 3.1.

**The Rigor Cycle** connects Design Science activities to the scientific knowledge base, ensuring a theoretically grounded and methodologically rigorous approach. It incorporates foundational research to establish theoretical foundations, domain expertise to define advancements in the field, and meta-artifacts as reference points within the domain. By integrating prior knowledge, the Rigor Cycle fosters innovation and novelty. The components of the knowledge base are illustrated in Figure 3.1.

**The Design Cycle** is central to any Design Science Research project, facilitating an iterative process of building, testing, and evaluating artifacts or processes. This cycle involves generating scientific report and design alternatives that will be systematically evaluated against predefined requirements until an optimal solution is achieved. These requirements are derived from both the environment and the scientific knowledge base, ensuring relevance and rigor.

## 3.2 Research Phases and Methodological Approach

Following the adopted design science research methodology, our research project was structured in several iterative phases, with each phase building upon the outcomes of the previous one. The process began with a comprehensive literature review, which established the state of the art based on existing expertise and prior research. This phase resulted in a review publication that detailed the research problem, identified research opportunities, and proposed potential solutions. These insights serve as a foundation for guiding future research and findings. The review can be found in the Manuscripts **??**.

### 3.2.1 Dataset

To conduct our research and address the research questions explored in each phase of the project, we utilized publicly available datasets. The evaluation process, which involved designing recommender algorithms for the food domain, required a dataset encompassing diverse food attributes, including ingredients, macronutrient composition, user ratings, and reviews. For this purpose, we adopted a widely used dataset in food recommender systems (Trattner and Elsweiler, 2019, 2017a; Starke et al., 2021a), sourced from AllRecipes.com [1], comprising an initial corpus of 58,263 recipes.

Depending on the requirements of each research phase, relevant subsets of recipes were extracted from the initial corpus, focusing on specific attributes (e.g., user ratings for CB, ingredients for CF). The datasets used in each research phase are documented in the corresponding manuscripts and contributions. Furthermore, additional datasets were generated through user experiments. To ensure the reproducibility of our results, all collected data have been made publicly available and are linked to each research contribution.

### 3.2.2 Algorithms and Offline Evaluation

An essential step in developing a recommender system is selecting the appropriate approach and algorithms. The recommendation algorithm constitutes the core computational component, responsible for predicting relevant items for active users (Konstan and Riedl, 2012). Choosing the right algorithm is context-dependent and requires significant time and effort due to the wide range of options that must be evaluated before identifying the most suitable approach (Cañamares et al., 2020).

Offline evaluation plays a crucial role in this selection process, allowing recommender algorithms to be assessed before deployment in user-facing environments (Castells and Moffat, 2022). Originally developed from experimental practices in machine learning and information retrieval, offline evaluation has evolved into a standard methodology for systematically testing and refining recommender systems.

In our research project, we employ an offline evaluation strategy to systematically select the most suitable recommender algorithm for integration with the chosen nudging techniques. For each research question, we begin by reviewing the literature to identify and examine relevant recommender approaches. Then the appropriate dataset is divided into two disjoint subsets: a training set used as input for the proposed algorithm and a test set serving as ground truth for evaluation metric computation. The algorithm is iteratively refined until optimal performance is achieved. All findings related to the developed algorithms are detailed in the respective manuscripts addressing each research question.

[1] https://www.allrecipes.com

Offline evaluation provides valuable insights into a recommender algorithm's ability to fulfill the system's defined tasks and objectives. However, it does not account for user satisfaction, acceptance, or overall experience. An algorithm may achieve high accuracy in identifying relevant items, yet the system may still fall short of user expectations due to factors such as usability, interpretability, and engagement (Konstan and Riedl, 2012; Knijnenburg and Willemsen, 2015; Ricci et al., 2021). Therefore, our evaluation extends beyond algorithmic performance to incorporate a user-centric evaluation framework, ensuring a more comprehensive assessment of the system's effectiveness and research goals.

### 3.2.3 User Centric Evaluation Approach

Evaluating the impact of developed algorithms directly on users is a fundamental aspect of research in the recommendation domain at large and within our project specifically (Pu et al., 2011). Beyond the algorithmic tasks of selecting, developing, and assessing recommendation model, alongside the careful design of nudging techniques, it is imperative to examine user behavior in a naturalistic setting (Konstan and Riedl, 2012). The most reliable approach to achieving this is through online evaluation, where user interactions with the system, encompassing both the recommendation algorithms and the user interface, are systematically analyzed (Knijnenburg and Willemsen, 2015). Furthermore, conducting field studies with real users provides essential empirical insights into user experience and the broader effectiveness of the system in real-world contexts, regarding the algorithm accuracy does not always correlate with user satisfaction with the system (McNee et al., 2006).

In order to precisely evaluate the impact of our interventions within a food recommender system and its influence on user decision-making, we adopted a user-centric evaluation framework. This framework is designed to provide a nuanced understanding of how different system components shape the user experience. We employed the well-established framework proposed by Knijnenburg et al. (Knijnenburg et al., 2012; Knijnenburg and Willemsen, 2015), which offers a structured approach to measuring user experience in recommender systems. This methodology enables us to systematically analyze the interaction between recommendation algorithms and user behavior, providing valuable insights into the effectiveness of our approach.

The framework provides both theoretical foundations and a structured methodology for conducting user experiments aimed at statistically measuring various aspects of user interaction within the recommender system. To achieve this, we adopted the following key aspects:

- **Objective System Aspects (OSAs)**: These are the system components currently under evaluation, including the recommender algorithm employed (e.g., collaborative filtering vs. content-based), the preference elicitation method used (e.g., constraint-based vs. case-based), and the system's generated recommendations.
- **Perception Aspects (PAs)** : refer to users' perceptions of (OSAs) in a recommender system, (e.g., the preceived effort). Measured through questionnaires, PAs help determine whether users notice and interpret system improvements, serving as mediators between OSAs and user experience. They clarify how and why system changes influence user satisfaction and interaction.
- **User Experience Aspects (UXAs)**: encompasses users' subjective evaluations of a

recommender system's quality, measured through questionnaires. It is categorized into three dimensions: **(1) System experience**, which assesses the perceived effectiveness of the recommender system itself; **(2) Process experience**, which evaluates the user's interaction with the system, including preference elicitation, browsing, and selection; and **(3) Outcome experience**, which captures satisfaction with the recommended choices, primarily measured through choice satisfaction and choice difficulty questionnaires.

- **Behavioural Indicators (BIs)**: are objective measures of user interaction with a recommender system, quantified through observable changes in user behavior. These metrics include the number of recommendations viewed, choices made, and time spent using the system. By tracking user clicks, selections, and ratings, behavioral indicators offer a robust foundation for assessing user experience and understanding the system's influence on behavior.

- **User Characteristics Aspects (UCAs)**: encompass personal characteristics, such as domain knowledge (e.g., subjective food knowledge), and situational characteristics, such as choice goals (e.g., eating goals). While the primary focus of user experiments is to examine the interaction effects of OSAs, PAs, UXAs, and BIs, these outcomes can also be influenced by UCAs. Typically assessed through pre-interaction questionnaires, UCAs exist independently of the system's influence and provide essential context for understanding user behavior and experience.

User-centric evaluation involves comprehensively understanding and statistically measuring user experience by quantifying the interactions and influence of each system aspect. In our project, we specifically measure the impact of nudging techniques integrated into food recommendations. This evaluation follows the guidelines proposed in the user-centric framework while aligning with the Design Science Research (DSR) methodology.

The first step in our process involves selecting the Objective System Aspects (OSAs) to be tested and evaluated by users. This is achieved by proposing a research model that includes system development, research questions, and hypotheses. These hypotheses predict how independent variables influence dependent variables and explore possible interactions between different aspects of the system (i.e., PAs, UXAs, BIs, UCAs) that can cause variations in user experience outcomes.

An essential aspect of the study is the design, which aligns with the proposed research model. Our user experiment aims to assess the recommendation algorithms (e.g., preference elicitation methods) and interface design (i.e., the nudge techniques). This is done by comparing it to a baseline condition, which serves as a reference point to link and compare the experiment results. The selection of an appropriate baseline may include non-personalized systems or state-of-the-art techniques. The experiment tests several conditions to determine which works best in relation to the research design and model.

A crucial phase after establishing the research model is participant recruitment. This step involves assembling a sample of users who will accurately represent the population and provide valid data for statistical evaluation. In our design, we primarily recruit participants through crowd-sourcing platforms such as Amazon Mechanical Turk [2] and Prolific [3]. Recruitment is conducted in an unbiased manner, following high-quality guidelines to ensure the integrity of the sample and the reliability of the results.

The final step in the framework involves statistically measuring user experience and the influence of system aspects within the recommender system. To quantify PAs, UXAs, and UCAs, pre-validated questionnaires were employed. Additionally, BIs were tracked using several metrics, such as FSA score (Starke et al., 2021c), user ratings, and feedback. We ensured that all the questionnaires used were statistically validated, meeting the preconditions for construct validity. Specifically, Confirmatory Factor Analysis (CFA) (Tavakol and Dennick, 2011) was employed to verify construct validity, while internal consistency was assessed through measures such as Cronbach's alpha (Harrington, 2009) and Average Variance Expained (AVE) (Harrington, 2009).

Furthermore, to evaluate differences between experimental conditions, a variety of statistical tests were applied, including t-tests, ANOVA test, and regression analysis. The overall user experience was assessed using Structural Equation Modeling (SEM) (Hair Jr et al., 2021), which enables the testing of all hypotheses while accounting for the interactions among the system aspects.

This user-centric evaluation framework, as outlined by Knijnenburg et al. (Knijnenburg et al., 2012; Knijnenburg and Willemsen, 2015), not only examines the individual effects of each variable within the recommender system but also provides a comprehensive understanding of how these factors collectively shape the user experience.

The framework enables a deeper understanding of the significant influences that nudging techniques and recommender algorithms have on users. Furthermore, each of our manuscripts offers a detailed presentation of research methods, evaluated system aspects, and the findings from the respective user experiments.

## 3.3 Ethical Consideration

This research project and methodology adhered to to the ethical guidelines set forth by the University of Bergen, Norway [4], and the regulatory standards for scientific research established by the university. Each user experiment was thoroughly reviewed to ensure compliance with these ethical principles, and it was determined that no further extensive ethical review was required. The studies was designed to avoid any misleading information, stressful tasks, or content that could potentially provoke extreme emotional responses from recommender system users. Furthermore, to safeguard participant privacy and confidentiality, all collected data were anonymized, ensuring that no personal identifying information was associated with the responses. These measures were taken to uphold the highest ethical standards in the conduct of the research and to protect the well-being of all participants involved.

[2] https://www.mturk.com/
[3] https://www.prolific.com/
[4] University of Bergen, Ethical guidelines

## 3.4 Conclusion

This chapter presents a detailed explanation of the research methodology employed in our project, highlighting the key components of our approach and its significance in the context of recommender system research. We begin by providing an overview of the Design Science Research (DSR) methodology, outlining its core attributes and relevance to the field. Subsequently, we describe the methodological phases, which include dataset selection, offline evaluation of recommender algorithms, and the selection of the nudging techniques, all aligned with DSR guidelines.

A comprehensive examination of the user-centric evaluation approach is also provided, as it forms the foundation of the user experiments conducted in this study. This evaluation framework is based on well-established methodologies and is carefully aligned with the objectives of our research. Finally, we discuss the ethical guidelines adhered to throughout the research process to ensure compliance with best practices.

# Chapter 4

# Results and Discussions

Food recommender systems derive their mechanisms from recommendation techniques designed to reduce choice overload and deliver personalized user experiences (Elsweiler et al., 2022). This personalization improves the efficiency and relevance of food item discovery. However, due to the prevalence of unhealthy popular online food content and users' inherent preferences for such options, food recommender systems often unintentionally promote unhealthy choices, reinforcing poor dietary behaviors and ultimately impacting public health (Trattner and Elsweiler, 2017a; Trattner et al., 2019). On the other hand, digital nudges which adapt the theory of offline nudging into online settings, shown promising results in influencing user behavior through subtle interface elements in digital environments across various domains (Weinmann et al., 2016; Jesse and Jannach, 2021).

This PhD project explored the central question: How can we support users in making healthier food choices without compromising personalization?. This research investigates the integration of digital nudging, specifically through nutrition labels and food explanations, within the presentation layer of food recommender systems to foster healthier, more informed decision-making.

Using a design science methodology, this thesis involved a series of user-centric evaluations of multiple food recommender system approaches and prototypes. These systems were enhanced with various digital nudging strategies and assessed through controlled experiments and real-user studies. The contributions of this works are aligned with the proposed research questions and are presented through a series of scientific manuscripts and peer-reviewed publications. The first manuscript (Manuscript 1) investigates and categorizes existing digital nudging techniques within the context of food recommender systems. The other manuscripts (Manuscript 2, Manuscript 3, Manuscript 4, Manuscript 5, Manuscript 6) provide empirical insights into which nudges are most effective in supporting informed decision-making and enhancing user experience. Furthermore, the insights generated through this work offer valuable implications for both researchers and practitioners, not only within the food domain but also across other domains that employ recommender systems.

This research contributes to making recommender systems more effective tools for behavioral change. Furthermore, the insights generated through this work offer valuable implications for both researchers and practitioners, not only within the food domain but also across other domains that employ recommender systems.

## 4.1 First Manuscript

- **Summary:** The first manuscript (Manuscript 1) presents a comprehensive systematic review of digital nudging techniques integrated into food recommender systems to support healthier food choices. The manuscript critically examines the multifaceted nature of food decision-making by discussing established food choice theories and the interplay of individual, contextual, and item-specific factors. It explores how food recommender systems have successfully addressed the problem of choice overload and enabled content personalization by modeling user preferences. However, it highlights a notable limitation in current systems, such as the insufficient emphasis on the nutritional quality and healthfulness of recommended items. To address this gap, the manuscripts discuss digital nudges as complementary technique to guide the user towards more healthy food choices.

- **Contribution:** The manuscript contribute to bridging the gap between digital nudging and food recommender systems for healthier food choice, by systematically reviewing and categorizing nudging techniques applied within this domain. The findings reveal that integrating strategies such as defaults, explanations, and notifications can effectively promote healthier food choices and improve user experience. At the same time, the review highlights that several techniques remain underexplored, while others, such as feedback and incentives, may be less appropriate due to their potential to increase cognitive load.

  The findings emphasizes the need for user-centered evaluations to assess the behavioral impact of these interventions and identifies promising, yet underutilized, nudging strategies for future research.

## 4.2 Second Manuscript

- **Summary:** The second manuscript (Manuscript 2) investigates the intersection of digital nudging and personalization in food recommender systems, focusing on how nutrition labels affect healthy food choices. The Manuscript presents a design of a recipe recommender system to conduct an online study with 600 participants, comparing personalized collaborative filtering vs. non-personalized recommendations combined with different labeling strategies (No-label, Multiple Traffic Light, Nutri-Score). Results showed that personalization leads to less healthy choices, likely due to popularity bias, while nutrition labels reduced choice difficulty in personalized interfaces, though they did not significantly increase the healthiness of selected recipes.

- **Contribution:** This manuscript (Manuscript 2) contributes to our understanding of the effectiveness of personalized food recommendations when supported by digital nudges, such as cognitive informational nudges. The findings emphasize that personalization significantly influences user choices but tends to promote unhealthy recipes, likely due to the reinforcement of existing user preferences and popular recipes. While front-of-package nutrition labels were introduced as digital nudges, they did not significantly mitigate the unhealthiness of chosen recipes, suggesting that unhealthy preferences often outweigh nutritional cues, especially in personalized contexts. Interestingly, in non-personalized settings, the addition of nutrition labels led to a slightly higher proportion of healthy recipe selections, though not to a statistically significant degree. From a user experience perspective, the integration of nutrition labels into personalized recommendations improved perceived choice difficulty, indicating that digital nudges may enhance decision-making ease even if they don't alter outcomes.

Finally, the study highlights the role of individual differences, such as users' eating goals, in predicting the healthiness of their choices, pointing to the potential of tailoring recommender systems not just to preferences, but also to personal health objectives.

## 4.3 Third Manuscript

- **Summary:** This manuscript (Manuscript 3) investigates how food recommender systems can support healthier recipe choices through the use of boosting an approach that empowers users by helping them understand the nudge (e.g., explaining nutritional labels) before making the decision. A 2x2 online experiment with 244 participants evaluated two preference elicitation methods: collaborative filtering (CB) and constraint-based (Cb), each with and without nutrition labeling. The findings reveal that nutrition labels significantly improved the healthiness of recipe choices, regardless of the recommendation method used. However, user experience varied depending on both the preference elicitation method and the user's level of health consciousness. Users with higher health consciousness levels found constraint-based systems more effortful and less satisfying, whereas users with lower level experienced them more favorably. These differences were reflected in measures of perceived effort, choice difficulty, and satisfaction. However, in this specific use case, the presence of the digital nudge shows no significant effect on the user experience metrics.

- **Contribution:** The main contribution of this manuscript (Manuscript 3) lie in identifying the key factors that influence food choices and user experience within food recommender systems. It demonstrates that boosting can significantly support healthier food decisions when combined with personalized recommender interfaces. Users consistently chose healthier recipes when these explainable digital nudges were present, highlighting the importance of user comprehension in promoting informed, health-conscious choices. Furthermore, user perception closely related to the preference elicitation method and the user's level of health awareness of the user. Those with higher health awareness reported greater effort and lower satisfaction when interacting with constraint-based systems. In contrast, users with lower health consciousness responded more favorably to them, contrasting with the overall preference for collaborative filtering These findings underscore that the user experience is shaped by both the preference elicitation method and user traits, such as health consciousness.

  Designing food recommender systems should therefore, go beyond pure algorithmic optimization and prioritize interface design and information presentation to support diverse users' needs and enhance their experience.

## 4.4 Forth Manuscript

- **Summary:** The manuscript (Manuscript 4) presents an evaluation of digital nudges (nutrition labels) applied within two types of recipe recommender systems: content-based and knowledge-based (KB). These differ in how they elicit user preferences, by selecting individual recipes or specifying attribute-based needs, respectively. The study employed a 3×2 between-subjects online experiment with 360 participants, testing three labeling conditions (no label, Multiple Traffic Light) label, and full nutritional label) across two preference elicitation methods. Results showed that the knowledge-based method significantly improved the healthiness of user choices, while nutrition labels show only a marginal effect. Additionally, users' subjective food knowledge influenced how they evaluated the system, especially regarding perceived effort, satisfaction, and difficulty, depending on the interplay between the preference elicitation method and nutritional labels.

- **Contribution:** This manuscript (Manuscript 4) contributes to a deeper understanding of how various system aspects, specifically preference elicitation methods, user domain knowledge, and digital nudges, interact to shape user behavior and experience in food recommender systems. Through a user-centered evaluation, the study demonstrates that both the preference elicitation method and the user's level of food knowledge significantly influence the healthiness of recipe choices. Notably, knowledge-based methods, lead to healthier decision-making and greater user satisfaction. While the impact of cognitive nudges (i.e., nutrition labels) on the healthiness of choices is marginal, their influence on user satisfaction is moderated by food knowledge, suggesting that such nudges are more effective when users possess the ability to interpret them.

  These findings underscore the importance of considering domain knowledge in the evaluation and design of recommender systems. They further highlight the need to align interaction design, including preference elicitation and interface nudging, with user characteristics. Moreover, the manuscript contributes to the growing recognition that effective food recommender system design must move beyond algorithmic optimization, toward user-adaptive approaches that account for the interplay between system design elements and individual user traits to improve both behavior and experience.

## 4.5 Fifth Manuscript

- **Summary:** The manuscript (Manuscript 5) investigates how different interaction modalities (i.e., interface design) in a conversational food recommender system influence user behavior and experience, particularly in supporting healthier food choices. A between-subjects user study (N = 195) compares three modalities: (1) text-only, (2) multi-modal (text and recipe images), and (3) multi-modal with nutrition labels.

  The results show that the multi-modal with labels condition led to significantly longer interaction durations and slightly healthier recipe choices, although the latter not statistically significant. Importantly, users engaged more extensively with the labeled multi-modal system and evaluated it as more effective than the single modality alternatives.

- **Contribution:** The contribution of this manuscript (Manuscript 5) lies in demonstrating that user experience in a conversational food recommender system is heavily influenced by the interface design and the type of interaction modality employed. The study reveals that these modalities vary in their effectiveness and in how users evaluate them. Specifically, the multi-modal interaction with nutrition labels resulted in significantly longer user engagement. Importantly, this increased interaction time shown to be associated with higher perceived system effectiveness, suggesting that richer, multi-modal interfaces can enhance engagement without leading to user fatigue or frustration.

  Overall, the manuscript contributes to our understanding of how combining multiple interaction modalities with digital nudges can inform the design of more engaging and health-supportive conversational food recommender systems.

## 4.6 Sixth Manuscript

- **Summary:** This manuscript (Manuscript 6) investigates the image-based and user-based factors influencing perceived food image attractiveness in the context of healthy food recommender systems. In a user study with 192 participants, users provided both numerical ratings and textual justifications for food image attractiveness. Results show that deep learning-based visual features outperform traditional low-level features in predicting image attractiveness. Among user factors, only cooking skills and recipe website usage significantly correlate with higher attractiveness ratings. Furthermore, the perceived appearance and healthiness of the food images were key dimensions influencing user textual judgment.

- **Contribution:** The contribution of this manuscript (Manuscript 6) lies in emphasizing the relationship between food image features, user characteristics, and perceived attractiveness. It demonstrates that advanced deep learning models outperform traditional low-level image features in predicting attractiveness. However, low-level features such as colorfulness, brightness, and naturalness still offer valuable, interpretable insights, positively influencing attractiveness, while features like saturation and sharpness negatively impact it.

  On the user side, profile characteristics such as cooking skills and frequent recipe website usage significantly contribute to attractiveness judgments. Overall, the findings suggest that using visually attractive food images as nudges in recommender systems can influence users' perception, which may ultimately shape their food choices and overall experience.

  Furthermore, image appearance and perceived healthiness are critical dimensions in users' evaluation of food images. Leveraging these attributes to develop advanced attractiveness prediction models and image-based explanations holds promise for promoting healthier food choices in recommender systems.

## 4.7 Discussions

### 4.7.1 On the findings

This thesis explores a novel research direction within the domain of food recommender systems, focusing on the integration of digital nudges with personalized recommendation techniques to support users in making more informed and healthier food choices. The main findings are derived following Design Science Research (DSR) methodology and are documented in peer-reviewed manuscripts, offering valuable insights for both researchers and practitioners in designing systems that promote behavioral change.

While the results underscore the potential of such systems, they also highlight several promising techniques that warrant further investigation within user-centric evaluation settings. This is necessary to better understand the interaction dynamics and how they influence overall user experience (Manuscript 1).

Integrating digital nudges into food recommender systems has demonstrated clear potential in promoting healthier food choices. However, a surprising finding, is that annotating personalized recipes with nutritional labels shows a weaker impact on encouraging healthy decisions compared to non-personalized recommendations. Specifically, users made healthier choices when presented with randomly generated recipe options rather than personalized ones. This supports existing concerns that personalization may inadvertently reinforce unhealthy behaviors, likely due to the widespread availability of popular, yet less nutritious, recipes online and users' natural inclination toward such options (Trattner and Elsweiler, 2017a; Elahi et al., 2021; Trattner et al., 2019). These tendencies may be rooted in our evolutionary predisposition for calorie-dense foods and the development of unhealthy food preferences from an early age (Beckerman et al., 2017; Waynforth, 2010).

Conversely, users with specific dietary goals demonstrated more deliberate behavior and greater intention in their food choices. They also perceived nutritional nudges as more helpful in aligning with their health objectives (Manuscript 2). In so-called "boost" scenarios (Manuscript 3), where users were explicitly informed about the meaning and function of the nudges, they exhibited a stronger understanding and were more likely to apply the nutritional information effectively, leading to healthier decisions. This highlights the critical role of understandability and user knowledge in shaping how users interact with recommender systems (Gedikli et al., 2014). Understandability not only enhances user trust and engagement but also enables users to make more informed and autonomous choices, key principles in boosting strategies, which differ from nudging approaches (Grüne-Yanoff and Hertwig, 2016). When users comprehend why a recommendation is made and how the system operates, they can better evaluate the relevance of the recommendation to their goals, particularly in health-related contexts (Dietvorst et al., 2015; Tintarev and Masthoff, 2012). However, this enhanced transparency and user engagement can also introduce biases. For example, providing too much control or explanation may lead to confirmation bias, where users seek information that confirms their existing preferences, even if those preferences are unhealthy (Schwind and Buder, 2012). Similarly, users with prior beliefs about food or nutrition might interpret nudges in line with those beliefs, diminishing their effectiveness or reinforcing existing misconceptions (Hauser et al., 2018).

Therefore, while boosting understandability and knowledge improves decision quality in many cases, system designers must find a balance to avoid overfitting to users' existing mental models or reinforcing unhealthy patterns.

Another central finding is that knowledge-based preference elicitation methods were more effective in guiding users toward healthier recipes compared to content-based methods. This mirrors findings from adjacent domains (Starke et al., 2021b; Knijnenburg and Willemsen, 2009), where users with domain expertise benefit more from attribute-based interactions due to their ability to articulate structured preferences. Similarly, food-knowledgeable users reported higher satisfaction and lower cognitive effort when interacting with knowledge-based systems, reinforcing the importance of aligning system complexity with user competence (Manuscript 4).

Moreover, user choice and the perceived attractiveness of food images are influenced by multiple factors, including both user intentions and the visual features of the images. Thus, the way content is presented shown to play an essential role in shaping user choices and supporting healthier, more informed decision-making. Overall, the findings demonstrate that preference elicitation methods, interface design, and digital nudges significantly affect user evaluation and experience within recommender systems. The interplay between user knowledge and the type of preference elicitation method emerged as particularly critical, emphasizing that adaptive and personalized interaction designs can enhance both the perceived usefulness and overall effectiveness of food recommender systems (Manuscript 4, Manuscript 5, Manuscript 6).

### 4.7.2 On the Limitations

The promising findings of this project pave the way for a deeper understanding and more effective evaluation of recommender systems and digital nudging interventions aimed at supporting users in making healthier and more informed decisions. While the use of the Design Science Research methodology offered a structured and rigorous framework for system development and evaluation, the reliance on online user experiments and crowdsourcing platforms may not fully capture real-world food decision-making behaviors. Such as, the tasks in which participants engaged were not always situated within typical real-life food decision-making contexts, for instance, being asked to choose a dinner recipe earlier in the day, which may have contributed to socially desirable responses. In everyday contexts, food choices are influenced by a wide range of external factors, such as time constraints, social dynamics, emotional states, and environmental cues, that are difficult to replicate in controlled experimental settings. Moreover, the inherent complexity of the food domain further limits the generalizability of the findings to real-life scenarios.

Another important limitation concerns users' understandability and familiarity with the digital nudges employed. For instance, the effectiveness of nutritional labels found to increase when users were explicitly introduced to their meaning and function. This suggests that the impact of such nudges may be constrained if users are unaware of or unfamiliar with the information being presented. Finally, although the insights presented are grounded in multiple manuscripts, the generalizability of the results is limited by the sample characteristics and the scope of the online user studies. Future research should aim to validate these findings in more diverse and ecologically valid settings and explore the long-term effects of repeated interactions with recommender systems on user behavior.

### 4.7.3 On the Future works

Building on the findings of this thesis, several avenues for future research emerge. First, there are numerous nudging techniques that merit further exploration within the context of recommender systems, particularly through user-centric evaluation settings. In addition, future studies should investigate the long-term impact of both nudges and preference elicitation strategies through longitudinal research designs, which can capture how user behaviors and preferences evolve over time and assess the sustained effectiveness of such interventions.

Second, future work should examine the integration of adaptive or personalized nudging mechanisms that dynamically respond to users' goals, behaviors, and contextual factors, such as time of day, emotional state, or health status. Incorporating diverse experimental setups, such as in-the-wild studies in which user make real food choices, or real-world deployments within food-related applications and grocery platforms, would further help evaluate the practical applicability and robustness of these systems.

Another important direction is to investigate the role of user knowledge and interface transparency in improving the interpretability and effectiveness of digital nudges, particularly for users unfamiliar with the nudging format (e.g., nutritional labels). Finally, expanding the demographic diversity of study participants and incorporating culturally informed food preferences will be crucial to enhancing the generalizability and inclusiveness of future recommender system designs aimed at supporting healthier and more informed decision-making.

# Summary

This PhD thesis investigated the interplay between recommender systems and digital nudges in promoting healthier food choices. The main findings shed light on how different recommendation approaches, nudging techniques, and levels of user knowledge shape user experience and influence behaviors. The chapters in this thesis provide a structured contribution for both researchers and practitioners working in the domain of food recommender systems.

The first chapter (Chapter 1) outlined the research gaps in current food recommender systems and presented the motivation for exploring the role of digital nudges in this context. The second chapter (Chapter 2) provided the theoretical background, introducing key concepts related to food recommender systems, digital nudging, and decision-making. The third chapter (Chapter 3) described the research methodology adopted to conduct the study and achieve the research objectives, while the fourth chapter (Chapter 4) presented and discussed the main findings, highlighting their implications and suggesting directions for future research.

The findings, documented across several scientific manuscripts (see, Manuscripts 4.7.3), underscore the complex role that personalization plays in food recommendation systems. While digital nudges such as nutritional labels showed promise in encouraging healthier behavior, their effectiveness varied depending on context and user familiarity with the nudging format. Notably, non-personalized recommendations were sometimes more effective than personalized ones, particularly when personalization reinforced preferences for less healthy, yet more popular, food options. Additionally, knowledge-based preference elicitation methods were found to be more effective than content-based approaches, especially for users with higher levels of food-related knowledge. User experience and evaluation also varied across preference elicitation methods, nudging strategies, and individual user knowledge.

This work contributes to the growing field of recommender systems by demonstrating how system design, nudging techniques, and user traits jointly shape user experience and behavioral outcomes. While the results are promising, they are limited by the scope of online user studies and the complexity of replicating real-world food decision-making environments.

Looking ahead, future research should investigate adaptive and context-aware nudging strategies, conduct longitudinal and in-the-wild studies, and explore the long-term behavioral impacts of intelligent food recommender systems. By addressing these areas, future work can further improve the design of personalized, responsible, and health-promoting food recommendation technologies.

# Bibliography


Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749. 10, 11

Aggarwal (2016). Knowledge-based recommender systems. *Recommender Systems: The Textbook*, pages 167–197. 15

Alshahrani, H., Mengash, H. A., Maashi, M., Kouki, F., Mahmud, A., and Al Duhayyim, M. (2024). Enhancing online food service user experience through advanced analytics and hybrid deep learning for comprehensive evaluation. *IEEE Access*. 3

Alves, G., Jannach, D., de Souza, R. F., Damian, D., and Manzato, M. G. (2024). Digitally nudging users to explore off-profile recommendations: here be dragons. *User Modeling and User-Adapted Interaction*, 34(2):441–481. 4

Asani, E., Vahdat-Nejad, H., and Sadri, J. (2021). Restaurant recommender system based on sentiment analysis. *Machine Learning with Applications*, 6:100114. 2

Beckerman, J. P., Alike, Q., Lovin, E., Tamez, M., and Mattei, J. (2017). The development and public health implications of food preferences in children. *Frontiers in nutrition*, 4:66. 36

Bennett, J., Lanning, S., et al. (2007). The netflix prize. In *Proceedings of KDD cup and workshop*, volume 2007, page 35. New York. 12

Berkovsky, S. and Freyne, J. (2010). Group-based recipe recommendations: analysis of data aggregation strategies. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 111–118. 2

Bettman, J. R., Luce, M. F., and Payne, J. W. (1998). Constructive consumer choice processes. *Journal of consumer research*, 25(3):187–217. 20

Bokde, D., Girase, S., and Mukhopadhyay, D. (2015). Matrix factorization model in collaborative filtering algorithms: A survey. *Procedia Computer Science*, 49:136–146. 12

Burke, R. (2000). Knowledge-based recommender systems. *Encyclopedia of library and information systems*, 69(Supplement 32):175–186. 15

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12:331–370. 9

Cadario, R. and Chandon, P. (2020). Which healthy eating nudges work best? a meta-analysis of field experiments. *Marketing Science*, 39(3):465–486. 20

Camargo, A. M. d., Botelho, A. M., Irmão, G. B., and Fiates, G. M. R. (2022). Analysis of recipes shared as 'healthy'in a popular brazilian website: A cross-sectional study. *International Journal of Environmental Research and Public Health*, 19(21):13914. 3

Cañamares, R., Castells, P., and Moffat, A. (2020). Offline evaluation options for recommender systems. *Information Retrieval Journal*, 23(4):387–410. 25

Capaldi, E. D. (1996). *Why we eat what we eat: The psychology of eating.* American Psychological Association. 1

Caraban, A., Karapanos, E., Gonçalves, D., and Campos, P. (2019). 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–15. 20, 21

Castells, P. and Moffat, A. (2022). Offline recommender system evaluation: Challenges and new directions. *AI magazine*, 43(2):225–238. 25

Chen, L., De Gemmis, M., Felfernig, A., Lops, P., Ricci, F., and Semeraro, G. (2013). Human decision making and recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 3(3):1–7. 1

Chen, L. and Tsoi, H. K. (2011). Users' decision behavior in recommender interfaces: Impact of layout design. In *RecSys' 11 Workshop on Human Decision Making in Recommender Systems*, pages 389–390. 19

Chen, M. and Liu, P. (2017). Performance evaluation of recommender systems. *International Journal of Performability Engineering*, 13(8):1246. 17

Chen, R., Hua, Q., Chang, Y.-S., Wang, B., Zhang, L., and Kong, X. (2018). A survey of collaborative filtering-based recommender systems: From traditional methods to hybrid methods based on social networks. *IEEE access*, 6:64301–64320. 10, 12, 13

Chiam, J., Lim, A., and Teredesai, A. (2024). Nudgerank: Digital algorithmic nudging for personalized health. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4873–4884. 4

Church, K. W. (2017). Word2vec. *Natural Language Engineering*, 23(1):155–162. 14

Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., and Sartin, M. (1999). Combining content-based and collaborative filters in an online newspaper. In *Proceedings of ACM SIGIR workshop on recommender systems*, volume 60, pages 1853–1870. Citeseer. 16

Cotter, P. and Smyth, B. (2000). Ptv: Intelligent personalised tv guides. *AAAI/IAAI*, 2000:957–964. 16

Daniel, K. (2017). *Thinking, fast and slow*. 19

De Croon, R., Van Houdt, L., Htun, N. N., Štiglic, G., Abeele, V. V., Verbert, K., et al. (2021). Health recommender systems: systematic review. *Journal of Medical Internet Research*, 23(6):e18035. 3

Denes-Raj, V. and Epstein, S. (1994). Conflict between experiential and rational processing: When people behave against their better judgement journal of personality and social psychology. *57 (2)*, pages 332–350. 19

Dickinson, K. M., Watson, M. S., and Prichard, I. (2018). Are clean eating blogs a source of healthy recipes? a comparative study of the nutrient composition of foods with and without clean eating claims. *Nutrients*, 10(10):1440. 3

Dietvorst, B. J., Simmons, J. P., and Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of experimental psychology: General*, 144(1):114. 36

El Majjodi, A., Elahi, M., El Ioini, N., and Trattner, C. (2020). Towards generating personalized country recommendation. In *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, pages 71–76. 12

Elahi, M., Kholgh, D. K., Kiarostami, M. S., Saghari, S., Rad, S. P., and Tkalčič, M. (2021). Investigating the impact of recommender systems on user-based and item-based popularity bias. *Information Processing & Management*, 58(5):102655. 36

Elsweiler, D., Hauptmann, H., and Trattner, C. (2022). *Food Recommender Food recommender Systems*, pages 871–925. Springer US, New York, NY. 2, 3, 4, 7, 18, 31

Epstein, S., Pacini, R., Denes-Raj, V., and Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of personality and social psychology*, 71(2):390. 19

Felfernig, A., Friedrich, G., Jannach, D., and Zanker, M. (2015). Constraint-based recommender systems. *Recommender systems handbook*, pages 161–190. 15

Forberger, S., Reisch, L. A., van Gorp, P., Stahl, C., Christianson, L., Halimi, J., De Santis, K. K., Malisoux, L., de Magistris, T., and Bohn, T. (2024). 'let me recommend…': use of digital nudges or recommender systems for overweight and obesity prevention—a scoping review protocol. *BMJ open*, 14(7):e080644. 4

Freyne, J. and Berkovsky, S. (2010a). Intelligent food planning: personalized recipe recommendation. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 321–324. 2

Freyne, J. and Berkovsky, S. (2010b). Recommending food: Reasoning on recipes and ingredients. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 381–386. Springer. 3

Galef Jr, B. G. (1996). Food selection: problems in understanding how we choose foods to eat. *Neuroscience & Biobehavioral Reviews*, 20(1):67–73. 1

Gedikli, F., Jannach, D., and Ge, M. (2014). How should i explain? a comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies*, 72(4):367–382. 36

Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70. 1

Goswami, I. and Urminsky, O. (2016). When should the ask be a nudge? the effect of default amounts on charitable donations. *Journal of Marketing Research*, 53(5):829–846. 20

Grüne-Yanoff, T. and Hertwig, R. (2016). Nudge versus boost: How coherent are policy and theory? *Minds and Machines*, 26(1):149–183. 36

Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., Ray, S., Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., et al. (2021). An introduction to structural equation modeling. *Partial least squares structural equation modeling (PLS-SEM) using R: a workbook*, pages 1–29. 28

Haque, M. S., Vainionpää, F., Tripathi, A., and Caraban, A. (2023). The present and future aspects of digital nudging and digital persuasion. In *PERSUASIVE (Adjunct)*. 4

Harrington, D. (2009). *Confirmatory factor analysis*. Oxford university press. 28

Haththotuwa, R. N., Wijeyaratne, C. N., and Senarath, U. (2020). Worldwide epidemic of obesity. In *Obesity and obstetrics*, pages 3–8. Elsevier. 2

Hauser, O. P., Gino, F., and Norton, M. I. (2018). Budging beliefs, nudging behaviour. *Mind & Society*, 17:15–26. 36

He, X., Liu, Q., and Jung, S. (2024). The impact of recommendation system on user satisfaction: A moderated mediation approach. *Journal of Theoretical and Applied Electronic Commerce Research*, 19(1):448–466. 1, 3

Herlocker, J. L., Konstan, J. A., and Riedl, J. (2000). Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 241–250. 10

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems*, 19(2):4. 24

Hevner, A. R., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS quarterly*, pages 75–105. 24

Hill, W., Stead, L., Rosenstein, M., and Furnas, G. (1995). Recommending and evaluating choices in a virtual community of use. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 194–201. 9

Holte, R. C. and Yan, J. N. Y. (1996). Inferring what a user is not interested in. In *Advances in Artifical Intelligence: 11th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, AI'96 Toronto, Ontario, Canada, May 21–24, 1996 Proceedings 11*, pages 159–171. Springer. 14

Huang, B., Yan, X., and Lin, J. (2016). Collaborative filtering recommendation algorithm based on joint nonnegative matrix factorization. *Pattern Recognition & Artificial Intelligence*, 29(8):725–734. 12

Isinkaye, F. O., Folajimi, Y. O., and Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian informatics journal*, 16(3):261–273. 16

Jannach, D. and Jugovac, M. (2019). Measuring the business value of recommender systems. *ACM Transactions on Management Information Systems (TMIS)*, 10(4):1–23. 1

Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. (2010). *Recommender systems: an introduction*. Cambridge university press. 15, 23

Jesse, M. and Jannach, D. (2021). Digital nudging with recommender systems: Survey and future directions. *Computers in Human Behavior Reports*, 3:100052. 4, 6, 21, 31

Jesse, M., Jannach, D., and Gula, B. (2021). Digital nudging for online food choices. *Frontiers in psychology*, 12:729589. 6

Jiang, H., Wang, W., Liu, M., Nie, L., Duan, L.-Y., and Xu, C. (2019). Market2dish: A health-aware food recommendation system. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2188–2190. 3

Karlsen, R. and Andersen, A. (2019). Recommendations with a nudge. *Technologies*, 7(2):45. 4

Kaya, B. (2018). User profile based paper recommendation system. *International Journal of Intelligent Systems and Applications in Engineering*, 2(6):151–157. 14

Knijnenburg, B. P. and Willemsen, M. C. (2009). Understanding the effect of adaptive preference elicitation methods on user satisfaction of a recommender system. In *Proceedings of the third ACM conference on Recommender systems*, pages 381–384. 37

Knijnenburg, B. P. and Willemsen, M. C. (2015). Evaluating recommender systems with user experiments. In *Recommender systems handbook*, pages 309–352. Springer. 17, 23, 26, 28

Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., and Newell, C. (2012). Explaining the user experience of recommender systems. *User modeling and user-adapted interaction*, 22:441–504. 7, 17, 26, 28

Konstan, J. A. and Riedl, J. (2012). Recommender systems: from algorithms to user experience. *User modeling and user-adapted interaction*, 22:101–123. 25, 26

Koroteev, M. V. (2021). Bert: a review of applications in natural language processing and understanding. *arXiv preprint arXiv:2103.11943*. 14

Kuhfuss, L., Préget, R., Thoyer, S., Hanley, N., Le Coent, P., and Désolé, M. (2016). Nudges, social norms, and permanence in agri-environmental schemes. *Land economics*, 92(4):641–655. 20

Li, M. and Chapman, G. B. (2013). Nudge to health: Harnessing decision research to promote health behavior. *Social and Personality Psychology Compass*, 7(3):187–198. 20

Lin, C.-J., Kuo, T.-T., and Lin, S.-D. (2014). A content-based matrix factorization model for recipe recommendation. In *Advances in Knowledge Discovery and Data Mining: 18th Pacific-Asia Conference, PAKDD 2014, Tainan, Taiwan, May 13-16, 2014. Proceedings, Part II 18*, pages 560–571. Springer. 3

Long, X., Sun, J., Dai, H., Zhang, D., Zhang, J., Chen, Y., Hu, H., and Zhao, B. (2025). The choice overload effect in online recommender systems. *Manufacturing & Service Operations Management*, 27(1):249–268. 1

Loo, J. (2023). Behavioural economics - what is system 1 and system 2 thinking. 19

Lops, P., De Gemmis, M., and Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. *Recommender systems handbook*, pages 73–105. 13

Lops, P., Musto, C., Narducci, F., De Gemmis, M., Basile, P., and Semeraro, G. (2010). Cross-language personalization through a semantic content-based recommender system. In *Artificial Intelligence: Methodology, Systems, and Applications: 14th International Conference, AIMSA 2010, Varna, Bulgaria, September 8-10. 2010. Proceedings 14*, pages 52–60. Springer. 14

Ma, C.-C. (2008). A guide to singular value decomposition for collaborative filtering. *Computer (Long Beach, CA)*, 2008:1–14. 12

Maia, R. and Ferreira, J. C. (2018). Context-aware food recommendation system. *Context-aware food recommendation system*, pages 349–356. 18

McNee, S. M., Riedl, J., and Konstan, J. A. (2006). Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems*, pages 1097–1101. 23, 26

Meeusen, R. E., van der Voorn, B., and Berk, K. A. (2023). Nudging strategies to improve food choices of healthcare workers in the workplace cafeteria: A pragmatic field study. *Clinical Nutrition ESPEN*, 53:126–133. 20

Mitra, M. and Chaudhuri, B. (2000). Information retrieval from documents: A survey. *Information retrieval*, 2:141–163. 14

Montanari, M. (2006). *Food is culture*. Columbia University Press. 1

Moseley, A. and Stoker, G. (2013). Nudging citizens? prospects and pitfalls confronting a new heuristic. *Resources, Conservation and Recycling*, 79:4–10. 20

Newell, R. G. and Siikamäki, J. (2014). Nudging energy efficiency behavior: The role of information labels. *Journal of the Association of Environmental and Resource Economists*, 1(4):555–598. 20

Nguyen, T. T. T. Q. (2016). *Enhancing user experience with recommender systems beyond prediction accuracies*. PhD thesis, University of Minnesota. 3

Pazzani, M. J. and Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web: methods and strategies of web personalization*, pages 325–341. Springer. 15

Pecune, F., Callebert, L., and Marsella, S. (2020). A recommender system for healthy and personalized recipes recommendations. In *HealthRecSys@ RecSys*, pages 15–20. 2

Pecune, F., Callebert, L., and Marsella, S. (2022). Designing persuasive food conversational recommender systems with nudging and socially-aware conversational strategies. *Frontiers in Robotics and AI*, 8:733835. 3

Philip, S., Shola, P., and Ovye, A. (2014). Application of content-based approach in research paper recommendation system for a digital library. *International Journal of Advanced Computer Science and Applications*, 5(10). 14

Poelman, M. P. and Steenhuis, I. H. (2019). Food choices in context. In *Context*, pages 143–168. Elsevier. 18

Pommeranz, A., Broekens, J., Wiggers, P., Brinkman, W.-P., and Jonker, C. M. (2012). Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process. *User Modeling and User-Adapted Interaction*, 22:357–397. 7

Pu, P., Chen, L., and Hu, R. (2011). A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 157–164. 7, 17, 26

Qaiser, S. and Ali, R. (2018). Text mining: use of tf-idf to examine the relevance of words to documents. *International journal of computer applications*, 181(1):25–29. 14

Ricci, F., Rokach, L., and Shapira, B. (2010). Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer. 9, 10, 16

Ricci, F., Rokach, L., and Shapira, B. (2021). Recommender systems: Techniques, applications, and challenges. *Recommender systems handbook*, pages 1–35. 23, 26

Rich, E. (1979). User modeling via stereotypes. *Cognitive science*, 3(4):329–354. 9

Roberts (2025). The great choices of strategic leaders. `https://go.roberts.edu/leadingedge/the-great-choices-of-strategic-leaders`. Accessed: 2025-03-04. 19

Rocchio, J. J. (1971). The smart retrieval system: Experiments in automatic document processing. *Relevance feedback in information retrieval*, pages 313–323. 14

Rokicki, M., Trattner, C., and Herder, E. (2018). The impact of recipe features, social cues and demographics on estimating the healthiness of online recipes. In *Proceedings of the international AAAI conference on web and social media*, volume 12. 3

Salton, G. (1989). Automatic text processing: The transformation, analysis, and retrieval of. *Reading: Addison-Wesley*, 169. 9

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. T. (2000). Application of dimensionality reduction in recommender system-a case study. 12

Schafer, J. B., Frankowski, D., Herlocker, J., and Sen, S. (2007). Collaborative filtering recommender systems. In *The adaptive web: methods and strategies of web personalization*, pages 291–324. Springer. 11

Schneider, C., Weinmann, M., and Vom Brocke, J. (2018). Digital nudging: guiding online user choices through interface design. *Communications of the ACM*, 61(7):67–73. 21

Schwind, C. and Buder, J. (2012). Reducing confirmation bias and evaluation bias: When are preference-inconsistent recommendations effective–and when not? *Computers in Human Behavior*, 28(6):2280–2290. 36

Shani, G. and Gunawardana, A. (2011). Evaluating recommendation systems. In *Recommender systems handbook*, pages 257–297. Springer. 17

Simon, H. A. (1988). The science of design: Creating the artificial. *Design Issues*, pages 67–82. 23

Sobolev, M. (2021). *Digital nudging: using technology to nudge for good*. SSRN. 4

Starke, A., Asotic, E., and Trattner, C. (2021a). "serving each user": Supporting different eating goals through a multi-list recommender interface. In *Proceedings of the 15th ACM conference on recommender systems*, pages 124–132. 25

Starke, A., El Majjodi, A., and Trattner, C. (2022a). Boosting health? examining the role of nutrition labels and preference elicitation methods in food recommendation. In *IntRS@ RecSys*, pages 67–84. 21

Starke, A., Willemsen, M., and Snijders, C. (2021b). Promoting energy-efficient behavior by depicting social norms in a recommender interface. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 11(3-4):1–32. 37

Starke, A. D., Asotic, E., Trattner, C., and Van Loo, E. J. (2023). Examining the user evaluation of multi-list recommender interfaces in the context of healthy recipe choices. *ACM Transactions on Recommender Systems*, 1(4):1–31. 3

Starke, A. D., Sedkowska, J., Chouhan, M., and Ferwerda, B. (2022b). Examining choice overload across single-list and multi-list user interfaces. In *9th joint workshop on interfaces and human decision making for recommender systems (IntRS'22)*. 6, 19

Starke, A. D., Willemsen, M. C., and Trattner, C. (2021c). Nudging healthy choices in food search through visual attractiveness. *Frontiers in Artificial Intelligence*, 4:621743. 6, 21, 28

Sugden, R. (2009). On nudging: A review of nudge: Improving decisions about health, wealth and happiness by richard h. thaler and cass r. sunstein. 20

Tavakol, M. and Dennick, R. (2011). Making sense of cronbach's alpha. *International journal of medical education*, 2:53. 28

Teng, C.-Y., Lin, Y.-R., and Adamic, L. A. (2012). Recipe recommendation using ingredient networks. In *Proceedings of the 4th annual ACM web science conference*, pages 298–307. 3

Thaler, R. H. (2018). From cashews to nudges: The evolution of behavioral economics. *American Economic Review*, 108(6):1265–1287. 4

Thaler, R. H. and Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin. 3

Tintarev, N. and Masthoff, J. (2007). A survey of explanations in recommender systems. In *2007 IEEE 23rd international conference on data engineering workshop*, pages 801–810. IEEE. 7

Tintarev, N. and Masthoff, J. (2012). Evaluating the effectiveness of explanations for recommender systems: Methodological issues and empirical studies on the impact of personalization. *User Modeling and User-Adapted Interaction*, 22:399–439. 36

Toledo, R. Y., Alzahrani, A. A., and Martinez, L. (2019). A food recommender system considering nutritional information and user preferences. *IEEE Access*, 7:96695–96711. 2

Tran, T. N. T., Felfernig, A., Trattner, C., and Holzinger, A. (2021). Recommender systems in the healthcare domain: state-of-the-art and research issues. *Journal of Intelligent Information Systems*, 57(1):171–201. 4

Trattner and Elsweiler, D. (2017a). Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proceedings of the 26th international conference on world wide web*, pages 489–498. 3, 21, 25, 31, 36

Trattner, C. and Elsweiler, D. (2017b). Food recommender systems: important contributions, challenges and future research directions. *arXiv preprint arXiv:1711.02760*. 2, 3, 7, 17, 18

Trattner, C. and Elsweiler, D. (2019). An evaluation of recommendation algorithms for online recipe portals. CEUR Workshop Proceedings. 25

Trattner, C., Kusmierczyk, T., and Nørvåg, K. (2019). Investigating and predicting online food recipe upload behavior. *Information Processing & Management*, 56(3):654–673. 3, 31, 36

Trattner, C., Moesslang, D., and Elsweiler, D. (2018). On the predictability of the popularity of online recipes. *EPJ Data Science*, 7(1):1–39. 3, 18

Van Meteren, R. and Van Someren, M. (2000). Using content-based filtering for recommendation. In *Proceedings of the machine learning in the new information age: MLnet/ECML2000 workshop*, volume 30, pages 47–56. Barcelona. 14

Vasan, S., Rao, A. A., and Gupta, N. (2025). Investigating the opportunities and challenges influencing consumer purchase behaviors and media influence in online food ordering: a thematic analysis. *British Food Journal*. 1

Vlaev, I. (2018). Local choices: Rationality and the contextuality of decision-making. *Brain sciences*, 8(1):8. 19

Warensjo, E., Nolan, D., and Tapsell, L. (2010). Dairy food consumption and obesity-related chronic disease. *Advances in food and nutrition research*, 59:1–41. 2

Wason, P. C. and Evans, J. S. B. (1974). Dual processes in reasoning? *Cognition*, 3(2):141–154. 19

Waynforth, D. (2010). Evolution, obesity, and why children so often choose the unhealthy eating option. *Medical hypotheses*, 74(5):934–936. 36

Weinmann, M., Schneider, C., and Brocke, J. v. (2016). Digital nudging. *Business & Information Systems Engineering*, 58:433–436. 3, 20, 21, 31

Wildavsky, B. (2013). Nudge nation: A new way to prod students into and through college. *ES Select*, 10. 20

Wind, Y. J. and Lilien, G. L. (1993). Marketing strategy models. *Handbooks in Operations Research and Management Science*, 5:773–826. 9

Yang, L., Hsieh, C.-K., Yang, H., Pollak, J. P., Dell, N., Belongie, S., Cole, C., and Estrin, D. (2017). Yum-me: a personalized nutrient-based meal recommender system. *ACM Transactions on Information Systems (TOIS)*, 36(1):1–31. 3

Yao, Y. (1995). Measuring retrieval effectiveness based on user preference of documents. *Journal of the American Society for Information science*, 46(2):133–145. 14

Zanardi, V. and Capra, L. (2008). Social ranking: uncovering relevant content using tag-based recommender systems. In *Proceedings of the 2008 ACM conference on Recommender systems*, pages 51–58. 14

Zhao, T. (2024). Performance comparison and analysis of svd and als in recommendation system. *Applied and Computational Engineering*, 49:142–148. 12

Zhou, X., He, J., Huang, G., and Zhang, Y. (2015). Svd-based incremental approaches for recommender systems. *Journal of Computer and System Sciences*, 81(4):717–733. 12

# Manuscripts

# Manuscript I

## Integrating Digital Food Nudges and Recommender Systems: Current Status and Future Directions.

*Ayoub El Majjodi, Alain D. Stare, Christoph Tratter.* **Under review, IEEE access.**

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

*Digital Object Identifier 10.1109/ACCESS.2017.DOI*

# Integrating Digital Food Nudges and Recommender Systems: Current Status and Future Directions

**AYOUB EL MAJJODI[1], ALAIN D. STARKE[1,2], AND CHRISTOPH TRATTNER[1],**

[1] MediaFutures, Department of Information Science and Media Studies, University of Bergen, 5007 Bergen, Norway
[2] Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Amsterdam, Netherlands

Corresponding author: Ayoub El Majjodi (e-mail: ayoub.majjodi@uib.no).

This work was supported by the Industry Partners and the Research Council of Norway through MediaFutures: Research Centre for Responsible Media Technology and Innovation through the Centres for Research-Based Innovation Scheme under Project 309339.

**ABSTRACT** Recommender systems are widely regarded as effective tools for facilitating the discovery of relevant content. In the food domain, they help users find recipes, choose grocery products, and receive meal suggestions. While they address the challenge of choice overload, their direct influence on promoting healthier food choices remains limited. Digital nudges could further assist in guiding users toward healthier decisions, enhancing the accessibility and visibility of healthy options when integrated into a recommender system. This review examines to what extent food recommender systems have so far successfully incorporated digital nudges for healthy food promotion and which challenges still remain. We present a classification and analysis of various digital nudging strategies employed for this purpose, as well as opportunities for future research. We emphasize that various nudging techniques have the potential to support users in making healthier food choices within food recommender systems. Furthermore, user-centric evaluations represent the most effective approach for assessing the performance of these systems.

**INDEX TERMS** Food recommender systems, Personalization, Digital nudging, Online decision-making, Health, Food choices

## I. INTRODUCTION AND MOTIVATION

Recommender systems areregarded as fundamental technology for running major apps to websites and significant digital food platforms [1]–[3]. Recommender systems help users to navigate the problem of choice overload [4], [5], reducing the choice set of relevant options and presenting content to a user they may like or be interested in. The primary filtering mechanism of a recommender system is the analysis of past user-item interactions, assuming that previously liked items will also be liked by a user in the forthcoming interactions [6].

Recommender systems are integrated across various domains. They have been shown to positively impact both users and service providers positively [7]–[9]. Whereas in many domains they focus on hedonic or taste-driven goals, such as finding music to listen or products to buy, food is often more versatile, with users also pursuing goals related to nutrition and sustainability [10], [11].

Recommender systems can play an important role in digital health promotion [12]. The World Health Organization reports that the global number of overweight adults has reached approximately 2.3 billion in recent decades [13], with overweight and obesity being significant contributors to the rising prevalence of chronic diseases [14] due to primarily unhealthy eating. Given these alarming statistics, recommender systems can be designed to support food choices [15]. They mainly simplify the selection process through personalized suggestions—from meal recipes to restaurant recommendations. Despite their potential benefits, implementing those systems faces inherent challenges due to the complexity of food choices, which are shaped by item characteristics, individual preferences, contextual variables, user mood, and social environments.

Human food choices are influenced by a range of determinants [16]. These include more biological factors [17], such as basic tastes (e.g. sweet, salty, bitter, and sour) and olfactory senses, which shape our perception of food flavor. These biological influences are mediated by genetic factors [17], bodily functions, and metabolic processes. Additionally, external factors also significantly impact food choices. This includes social and environmental influences [16], [18], such as attitudes, beliefs, motivations, values, social status, and interpersonal relationships [19], [20]. The knowledge and skills acquired throughout an individual's life further affect dietary decisions [18], [19], [21]. Furthermore, shifts in global food systems have led to changes in eating patterns, contributing to an increase in the consumption of unhealthy foods [20]. Despite the substantial body of research across various disciplines, the combined factors in food choice and behavior remain difficult to predict [22], [23].

Recommender systems have significantly advanced the modeling of food choice behavior. They apply sophisticated statistical and mathematical models in conjunction with information retrieval techniques [24]. However, not all interaction factors can be effectively modeled [25]. Figure 1 illustrates the traditional process of a food recommender system, highlighted in grey, as originally presented by Elsweiler et al. [24]. It demonstrates how various food choice factors can be integrated into the modeling of such systems. These systems typically feed user characteristics and food attributes (cf. Figure 1, Part (A)) as inputs to predefined algorithms and recommendations approaches (cf. Figure 1, Part (B)) to generate personalized food recommendations. The goals of these systems vary, from recommending ingredients similar to those preferred by the user to offering recipe suggestions based on past user ratings [12]. In later research, food recommender systems have started to incorporate nutritional content or generating tailored nutritional advice based on users' health conditions. Despite these advancements, efforts to optimize recommendations for health outcomes have remained limited. For example, Chen et al. [26] focuses on how nutrition data, such as calorie content and nutrient balance, can improve the health value of recommended recipes. Nevertheless, current literature suggests that food recommender systems tend to prioritize popular content, which is often associated with unhealthy dietary standards. [27]–[29].

While recommender algorithms focus on the generation of content, *how* this content presented also significantly impacts user preferences. These preferences are often context dependent [30] and can be influenced by various presentations factors (see Figure 1, Part C). Such factors present an opportunity to guide users towards healthier food choices, thereby by enhancing the overall effectiveness of food recommender systems [11], [31], [32]. Any changes made to an interface (i.e., choice architecture) that lead to predictable choices, e.g., due to cognitive biases, are called 'nudges'. Originally introduced by behavioral economists Richard Thaler and Cass Sunstein, nudges were primarily examined in offline decision-making contexts and have successfully supported informed choices related to health and wealth [33], [34]. The successful application of nudging techniques in physical environments has generated significant interest in exploring their potential within digital contexts. In 2016, Weinmann et al. [35] introduced the concept of "digital nudges," defining them as interventions designed to modify user interfaces in ways that steer user behavior toward desired outcomes. Digital nudging has been increasingly adopted across various online platforms. For example, social norm nudges aimed at promoting daily walking have been shown to produce meaningful changes in user behavior [36]. Similarly, interventions on online grocery platforms have demonstrated considerable efficacy in fostering more sustainable and environmentally conscious purchasing decisions [37]. Furthermore, repositioning nudges within online food ordering systems have been associated with a significant increase in the selection of healthier food options [38].

Recent research on recommender systems has begun exploring the integration of digital nudges to promote behavioral change and deliver personalized content [39]. For instance, incorporating an explanatory nudge into a food recommender system has been shown to positively influence users' selection of healthier recommendations [40]. Similarly, enhancing user awareness of the healthiness of recommended recipes through the use of multi-color coding as a nudge has been found to encourage healthier food choices [41]. In line with these findings, the inclusion of visually appealing images alongside healthy recipes significantly increases the likelihood that users will select these options in a search-based recipe recommender system [42].

This research aims to investigate how digital nudges have been integrated into food recommender systems to promote healthier food choices. More specifically, it seeks to present digital nudging techniques that effectively promote healthier food choices within these systems. Additionally, the study will explore techniques that warrant further investigation for their potential in this area. To guide the research, the following primary research question has been formulated:

- *RQ: How can digital nudges be integrated into recommender technology to effectively support users to make healthier food choices?*

To systematically organize our study around the main research question, we propose the following sub-research questions:

- *How have digital nudges been integrated in recommender systems to support healthier food choices?*
- *What are the potential future applications of nudges in recommender systems for healthier food choices?*

The paper is organized as follows. Next, in Section II, we detail on background relevant for this review paper. Section III discusses the methodology for performing the literature review. Section IV presents the results in light of our main research question. Finally, Section V discusses future opportunities to integrate food recommendations and digital nudges to support healthier food choices.

## II. BACKGROUND

The aim of this paper is to review the current state of the art of digital food nudges integrated into food recommender systems and to reveal future applications of nudges

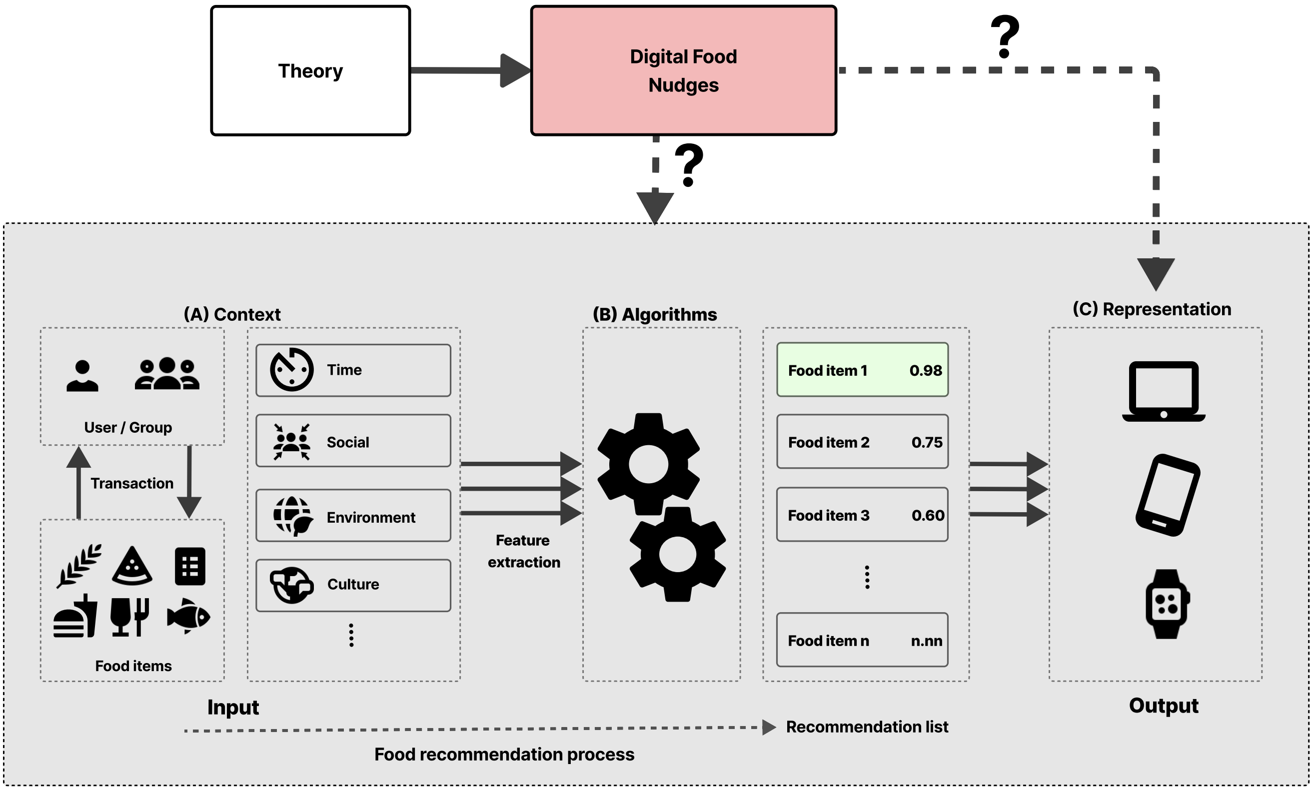



FIGURE 1: Factors that influence user food choices in the recommendation process include (A) context, (B) algorithms and recommendation lists, and (C) interface representations. The gray section, originally developed by Elsweiler et al. [12], illustrates the original food recommendation process, whereas this paper examines the integration of digital nudges (e.g., highlighted in red) into the process.

in food recommender systems. The following sub-sections provide a background on current (i) food choice theory, (ii) food recommender system approaches, and (iii) digital food nudges on platforms that not yet been integrated into food recommender systems.

### A. FOOD CHOICE THEORY

Human food choice is an iterative process governed by dynamic factors that integrate multiple interacting determinants [43]–[45]. The interacting factors can be categorized into those pertaining to the individual, the food item, and the broader environmental or situational context, as articulated in one of the foundational models of food choice [46].

Individual factors influencing food choice are typically categorized into several dimensions. Personal state factors include genetic predispositions, dietary patterns, and physical health conditions. Psychological factors encompass hunger, appetite, and body weight, along with emotional, motivational, and personality characteristics [43]. Additionally, personal habits and past experiences contribute significantly to shaping these personal states [47]. The cognitive aspects of food choice are also critical, going beyond mere knowledge and skills to include preferences, personal identity, and sensory evaluations, thus adding complexity to food-related decision-making processes [48]. Attitudes represent implicit evaluations of food, whereas liking refers to the sensory evaluation of food's appeal. Preferences are based on relative evaluations of different food options. A third cognitive dimension is the anticipated outcomes of food choices, particularly in terms of perceived health risks or benefits [49], [50].

Conversely, food items are shaped by a combination of intrinsic and extrinsic factors. Intrinsic factors pertain to the inherent properties of the food itself, encompassing sensory attributes such as flavor, taste, aroma, and texture, along with perceptual features like color, portion size, nutritional content, health benefits, and overall quality [46], [47]. In contrast, extrinsic factors are bifurcated into informational and environmental categories. Informational factors involve knowledge related to the food, including its origin, nutritional profile, and composition. Additionally, the social environment influences food choices directly and indirectly through cultural norms, culinary practices, and the social context of consumption, such as specific occasions. The physical environment also plays a pivotal role in determining food availability and accessibility, significantly affecting food choice behavior [47], [51].

Comprehensive models have been developed to explain the factors influencing food choice and decision-making. Sobal et al. [44] model suggests that personal, resource, and contextual factors shape personal food systems based on values, situations, and strategies, which in turn influence food behavior in a feedback loop [52]. Connors et al. [53] expanded this model by emphasizing the role of life course experiences, ideals, and social context in shaping food behavior through similar feedback mechanisms. Sensory factors and packaging information can also trigger emotional responses that directly affect the final food choice [54]. Understanding these interrelated factors is key to promoting healthier eating, with critical determinants including food mavenism, knowledge, market availability, and personal norms [55]–[57].

The complexity of explaining final food choices arises from the aforementioned interacting factors and other various determinants [20]. Within the field of recommender systems, the primary focus has been on modeling user food preferences to identify the most appealing recipes, meals, or products. This process is typically framed as a prediction task. Additionally, time and contextual factors have been modeled to a significant extent within food recommendation processes [58]. However, a notable research gap exists between offline food choice theories and the factors that guide user preferences in digital food recommender systems [12].

### B. THE CURRENT STATUS IN FOOD RECOMMENDER SYSTEMS

Recommendation technology is typically divided into different recommendation approach categories: collaborative filtering, content-based, knowledge-based, and hybrid methods. Each of these approaches leverages interactions between the user and the content data [6], but they differ in how this is achieved. In the food domain, these methods are utilized to learn user preferences and retrieve relevant food items. For most content-based and knowledge-based approach, food items (e.g., recipes) tend to be represented in terms of their attributes, such as ingredients, cooking directions, and basic name-based metadata [25], [59].

Collaborative filtering algorithms rely on identifying similarities between user profiles to generate recommendations. These methods operate on the assumption that users with similar preferences and interests are likely to be interested in the same items [6], [60]. By integrating taste factors into user profiles alongside previous ratings, a collaborative-based food recommender system can be developed, offering improved precision compared to traditional linear regression models [61]. Moreover, a relevant and specific diet tailored to individual users in collaborative-based approaches using neighborhood algorithms leads to 85% recommendation accuracy [62].

Over the past decades, content-based food recommendation methods have been the most widely utilized in the literature [59]. The primary distinction of these methods lies in their generation of recommendations through a more in-depth analysis and semantic understanding of item content and user preferences [63]. The item content is represented by features that can range from ingredients to images of prepared dishes. Thus, recommendations are retrieved based on a similarity function that matches the features in a user's profile with those in the system's database.

Knowledge-based recommenders use a different approach. Designers of such systems leverage domain knowledge to score items based on elicited user characteristics and preferences [6], [64]. For example, Cataldo et al. [64] present a knowledge-based food recommender that personalizes recipes by incorporating users' health-related characteristics and personal factors, such as BMI and dietary restrictions. Such approaches would present food content that matches a user's nutritional needs [31]. Knowledge-based approaches are less commonly employed mainly because they limit the food exploration of food options by requiring users to define specific attributes or constraints [3], [25], [65]. The database items are narrowed down according to user criteria and filters, resulting in a selection of predefined food items.

Hybrid recommender systems, as the name suggests, integrate multiple recommendation techniques or components into a single approach [6]. This technique is typically used to reduce limitations present any specific approach, such as the reduced novelty in content-based recommenders [11]. Hybrid systems have shown to enhance both accuracy and performance by drawing on recommendations from diverse sources of information. Pallavi Chavan et al. [66] demonstrated that hybrid food recommender approach outperform collaborative and content-based filtering regarding recall and accuracy, as they benefit from the comprehensive content analysis inherent in traditional methods.

### C. DIGITAL FOOD NUDGES

Beyond recommender algorithms, there is more that can be designed and personalized in a recommender system. The interface is particularly interesting as different presented options can be reorganized to elicit predictable behavior among users. Hence, users have specific cognitive biases that can be leveraged to steer decision-making towards specific options [67], [68].

This is where nudging theory can play a role in the design of such interfaces. the term "Nudge" encompasses various techniques used to support people's behavior and overcome decision biases and heuristics inherent in human thinking systems based on the choice architect in the offline settings. Such techniques have been proven to support and guide informed decision-making across different domains, including the health sector, policy-making, environment, and education [68], [69].

In the context of food choice environments, it is possible to influence individuals to, for example, select healthier options by reducing the perceived effort required during decision-making [70]. This can be achieved by presenting the healthier choice as the more convenient or accessible [71]. The effectiveness of such nudging strategies within physical food environments has been demonstrated in various studies. For instance, research by Bucher et al. [71] has shown that increasing the physical distance of a food option decreases its selection. Cadario et al. [72] introduces a conceptually grounded framework that classifies healthy eating nudges into three distinct, theory-based categories. This framework offers a systematic approach to understanding and differentiating between various nudging interventions to promote healthier eating behaviors. Each category within the framework is anchored in established theoretical principles [33], enabling researchers and practitioners to discern the underlying mechanisms through which each type of nudge exerts its influence. Accordingly, in the present work, we will utilize this framework to categorize the nudging techniques found in literature into three classes, aligning them with the following categorization defined by Cadario et al. [72]:

- *Affectively oriented nudges*: Nudging techniques that encompass interventions designed to enhance the hedonic appeal of healthy food options. This category includes strategies such as "hedonic enhancements", which aim to increase the sensory attractiveness of healthier choices through vivid descriptions, appealing displays, enticing photographs, or aesthetically pleasing containers. Additionally, this category incorporates "healthy eating calls," which are direct appeals that encourage individuals to make healthier choices.
- *Cognitively oriented nudges*: This category includes two types of nutritional labeling: "descriptive nutritional labeling", which offers detailed information such as calorie counts or the content of other nutrients, and "evaluative nutritional labeling", which simplifies the assessment of food items' healthiness through color-coded indicators (e.g., red, yellow, green). A third technique within this category is "visibility enhancement", which aims to inform consumers of the availability of healthier options by increasing their prominence within a food setting.
- *Behavioural oriented nudges*: Interventions that aim to impact people's behaviors without necessarily influencing what they know or how they feel often without people being aware of their existence.

In the contemporary landscape, most human decision-making is made through digital platforms, such as websites and mobile applications [73]. However, individuals are particularly susceptible to making suboptimal decisions within these digital environments. Drawing on the theoretical principles of nudging, Weinmann [35], introduced the concept of "digital nudging" as the use interface design elements to steer individuals' choices or shape users' decisions in online environments. Digital nudges have been explored across various contexts, including privacy, work, and productivity. However, there is growing interest in the application of digital nudges within social contexts, where they are designed to promote more sustainable behaviors, as well as in health-related context, where the objective is to guide users toward actions that are more beneficial to their well-being [74]. In the latter context, digital nudges are highly effective, resulting in a 63% reduction in the proportion of unhealthy food choices and a 30% increase in the selection of healthy food products [75].

### D. DIFFERENCES & CONTRIBUTIONS OF THIS WORK

Several attempts have been made to integrate healthiness into food recommender systems, with some studies focusing on generating healthier food options, while others emphasize providing nutritional advice tailored to the user's goals [24], [76], [77]. However, a systematic review on food recommendation systems over the past decade revealed that less than 20% of the literature incorporates nutritional considerations in the process of making food items recommendations [60]. Furthermore, findings by Trattner et al. [27] indicate that the majority of online recipes are generally unhealthy, highlighting the need for food recommendation systems to prioritize the healthiness of recommended items rather than focusing solely on algorithmic accuracy and user preference optimization.

Digital nudging, on the other hand, is a rapidly a growing field of research that has proven to be an effective solution in the area of behavioral change. The strength of these techniques lies in their ease of implementation and evaluation. In the food domain, digital nudges have demonstrated their ability to guide users toward healthier food choices [78], [79]. Given the potential of these technologies, several efforts have been made to investigate the combination of recommender systems and digital nudges to promote healthier food choices. These attempts recognize the powerful role that such integrated approaches can support and guide individuals toward better dietary decisions, thereby enhancing public health outcomes [39].

To the best of our knowledge, no prior research has systematically reviewed the integration of digital nudges with recommender systems specifically for promoting healthier food choices. However, the latest work of [80] has focused on using digital nudges and recommender systems in the context of obesity prevention. In contrast, our work systematically identifies and classifies digital nudging techniques for food selection. Figure (1) illustrates how digital nudges are integrated into the process of a food recommender system. Specifically, digital nudges are employed in the interface and representation elements rather than in the algorithm generation phase. This distinction highlights that the primary different between a recommender system and a nudge lies in the design elements enhanced by digital nudges towards the desired proposes [39], [81]. Building on this foundation, we present the work that integrates these digital nudges within food recommender systems and propose a mapping framework that outlines potential combinations for facilitating healthier food choices.

## III. METHODOLOGY

This section outlines the search strategy for the literature review and presents the research findings based on the proposed research questions. The paper follows traditional guidelines for conducting a systematic literature review [82], [83].

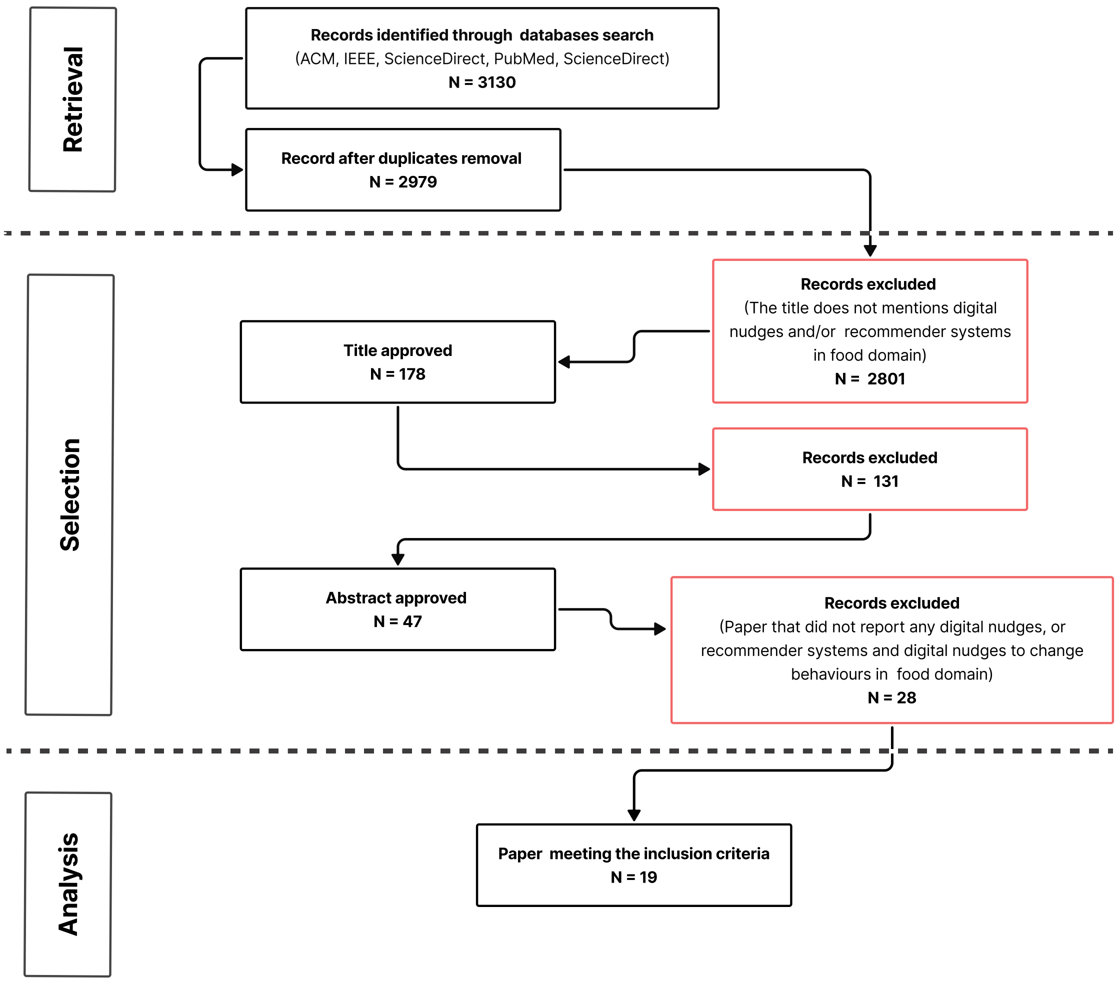



FIGURE 2: Flow diagram revealing the process of how publications were selected or excluded for further analysis.

### A. RETRIEVAL: DATABASE SEARCH

The search strategy started with identifying keywords (e.g, recommender systems, digital nudges, health, food) pertinent to our research questions, which were then employed to construct detailed search queries. These queries were subsequently refined to meet the retrieval criteria of each selected academic database. The search, conducted in June 2024, was filtered to include only publications from 2014 to 2024 to ensure relevance in addition to digital nudging was first introduced in the last decade and more research started to become a hot topic in academia with a ten-fold increase of the literature over the last five years [35], [84]. This process yielded a total of 3,130 papers.

Table 1 summarizes the database sources and the corresponding number of records retrieved per database.

TABLE 1: The total amount of literature retrieved from the different sources.

| Database | Records |
|---|---|
| ACM Digital Library [85] | 1534 |
| IEEEXplore Digital Library [86] | 817 |
| ScienceDirect [87] | 748 |
| SpringerLink [88] | 16 |
| PubMed [89] | 15 |
| **Total** | **3130** |

### B. SELECTION: DEFINE INCLUSION CRITERIA

We applied specific inclusion and exclusion criteria to titles and abstracts during the selection phase, concentrating on references to digital nudging and/or recommender systems to promote healthier food choices. This phase, conducted by the first two authors, identified 47 papers that met the inclusion criteria and were subsequently selected for a full-text review.

The selection process involved detailed discussions to ensure that each abstract met our criteria and review scope. Figure 2 illustrates the methodology employed for article selection and analysis.

### C. ANALYSIS: PAPER EVALUATION AND DATA EXTRACTION

The analysis phase involves a comprehensive review and evaluation of the full texts of papers selected in the previous phase. Out of 47 papers assessed, 19 were included in the final selection. Data is systematically extracted by redefining metadata using a standardized form, which captures details such as the paper title, author, publication venue, nudging techniques, recommendation techniques, datasets, and other relevant information.

The inclusion and exclusion criteria, keywords, and a number of retrieved papers are detailed in [90]. The retrieved paper title, abstract, and the decisions made by the authors is accessible via [91], while the form used for data extraction from full-text readings can be found in [92].

## IV. RESULTS

This section begins with an overview of publication statistics and outlets, followed by an in-depth analysis of key findings related to the integration of digital nudges into food recommender systems. Subsequently, it categorizes the digital nudging techniques investigated for promoting healthier food choices, aligning them with previously validated frameworks. Finally, the section provides a summary of the datasets, measurement methods, and recruited participants used in the user studies, offering insights into the methodologies employed for evaluating user behaviors.

### A. PUBLICATION STATISTICS

A total of 3,130 primary studies were initially retrieved for this literature review, as presented in Table 1. During the selection process based on the inclusion criteria outlined in [90] and after duplicated removal, this number was reduced to 47 studies. In the analysis phase, 19 papers were selected based on the author's abstract and full paper text reading. Most of the studies were published during 2021, while only two contributions were published in the first part of 2024. The distribution of the studies published per year is illustrated in Figure 3.(A). The distribution of relevant papers retrieved from various sources is depicted in Figure 3.(B). The data reveals that the ACM Digital Library [85] contains the largest collection of research on digital nudges and food recommender systems, comprising 70% of the total 19 papers. In contrast, 15% of these papers were published in journals available through ScienceDirect [87]. The IEEE Xplore Digital Library [86] and PubMed [89] contribute the least to the research corpus on this topic, with each source representing 10% and 5% of both conference and journal papers, respectively.

### B. FOOD RECOMMENDER SYSTEMS AND DIGITAL NUDGES

Advancements in technology and computational power have significantly enhanced the accuracy of personalization in recommender systems [59]. However, the multifaceted complexity of the food domain necessitates incorporating behavioral change techniques, particularly through digital nudging, within the representation phase of the recommendation process [12]. This integration aims to support individuals in making healthier food choices by employing several nudging techniques on the front-end interface.

TABLE 2: The integrated digital nudges and recommender systems approaches.

| Recommendation approach | Nudging Technique | Studies |
|---|---|---|
| Content-Based | Nutritional Food Labels | [65], [93], [94] |
| | Explanations | [11], [94] |
| | Positioning | [42] |
| | Re-ranking | [95] |
| | Social Norms | [93] |
| Knowledge-Based | Nutritional Food Labels | [65], [96] |
| | Notifications | [97] |
| Collaborative Filtering | Nutritional Food Labels | [98], [96] |

Approximately 47% of the 19 reviewed papers successfully demonstrate the integration of recommender systems with digital nudges. Content-based recommender systems emerge as the most frequently integrated technique, utilized in 8 studies and combined with 5 distinct digital nudges. Table 2 outlines the studies that employed content-based approaches for generating recommendations and applying nudges. Notably, nutritional food labels were the most extensively examined, integrated with recommender systems in 7 studies, and tested across all recommender approaches. Integrating digital nudges based on various nutritional food labels into a content-based recipe recommender system has been shown to impact the healthiness of selected recipes compared to conditions without nudges. Specifically, using Multiple Traffic Light (MTL) labeling results in a modest improvement in the healthiness of chosen recipes, as measured by the Food Standards Agency (FSA) score. This effect is measured in both web-based recommender systems and chatbot-based mobile applications [65], [94]. However, a study indicates that when food items are annotated with a combination of nutritional and environmental labels, there is a stronger tendency toward sustainability rather than healthier choices [93]. Interestingly, constructing a TF-IDF-based recipe content recommender system, combined with health-related explanations as nudges, leads to healthier food choices when recipe recommendations are presented as a single list rather than multiple lists [11]. Starke et al. [42] suggest that aligning food recommendations with users' dietary goals can overcome the positional nudge effect on user choice behavior. Conversely, another study found that integrating a simple re-ranking mechanism into various content-based recommenders significantly increased the selection of recipes

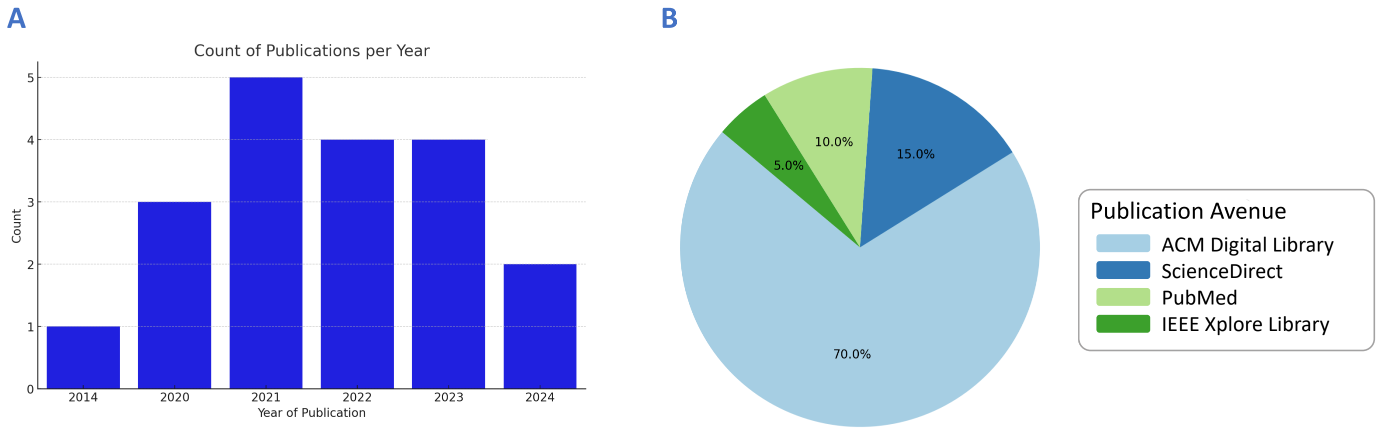



FIGURE 3: Publications statistics: (A) present the number of papers per year. (B) shows the distribution of relevant papers retrieved per avenue.

with lower fat content compared to choices made using a random recommender model [95].

Consistent with previous findings, collaborative filtering and knowledge-based approaches are among the least utilized methods alongside digital nudges in the food domain [12], [59]. However, when nutritional food labels are incorporated as nudges, they have been shown to effectively promote healthier food choices within both collaborative filtering and knowledge-based recommender systems [96], [98]. Additionally, ongoing research has reported the integration of a knowledge-based recommender system with another type of digital nudge, which explicitly sends notifications to users as reminders to make healthy food choices in mobile applications [97].

Integrating digital nudges into personalized content in the food domain has demonstrated significant potential for driving behavioral change. Surprisingly, this direction has received the least attention from both academia and industry [12], [39], as shown in Table 2. The following section reviews and categorizes the primary contributions of digital nudging techniques toward promoting healthier food choices and discusses the impact of each method.

### C. DIGITAL NUDGES FOR HEALTHIER FOOD CHOICE

Our analysis identified 20 studies documenting the use of digital nudges. The three most frequently employed techniques were nutritional food labels, positional nudges, and prompts, while feedback, defaults, and menu labeling were the least utilized among the 11 reported techniques. We categorize those nudging techniques into Cadario et al. digital nudging for a healthier choice framework detailed in Table 3.

Affective-oriented digital nudges have been identified as the most frequently employed techniques, surpassing cognitive and behavioral-oriented nudges, with four distinct methods documented across nine studies, as shown in Table 3. Among these, the combination of prompt and feedback nudges has addressed two key challenges in promoting healthy eating: encouraging mindful, slower eating and

TABLE 3: Implemented digital nudging techniques for healthier food choices, categorized based on the framework explained in (II-C)

| Category | Technique | Studies |
|---|---|---|
| Affective | Prompt | [99], [100] , [101] |
| | Feedback | [99], [100], [102] |
| | Explanations | [102] |
| | Incentives | [103] |
| Cognitive | Nutritional Food Labels | [99], [104], [102], [100] |
| | Visualization | [105], [106] |
| Behavioral | Positioning | [99], [100], [107] |
| | Social Norms | [108], [105] |
| | Defaults | [107] |

fostering healthy eating habits [101]. Incorporating gameful elements into mobile applications, such as providing food explanations, has effectively enhanced users' nutritional knowledge and encouraged healthier food choices [102]. Alternatively, by designing rewards and incentives, users are further motivated to adopt healthy diets that can help prevent the spread of non-communicable diseases. This approach illustrates how a combination of explicit and implicit interventions can promote healthy behaviors in offline settings (i.e., groceries) through digital means [103].

Nutritional food labels are the most prevalent cognitive digital-oriented food nudges, typically presented in the form of Multiple Traffic Light systems [99], [102], [104]. These labels have proven highly effective, leading to an average 59% increase in healthy and sustainable food choices [104]. Interstenly, a visualization cue as a food nudge linked to a self-report mobile application has proven to lead to effectiveness in user self-tracking their eating behaviors and the nutrition content of the food before making the choice [105], [106].

Reordering and setting healthier options as the default in an online food choice environment have been shown to effectively influence users' shopping behaviors [99], [107]. Similarly, the use of social norms, which has proven effective in driving behavioral change in offline settings [109], [110], has also led to a significant impact on calorie intake in digital environments through the food choices made [108].

Nudging has been widely studied in offline settings across various domains; its application in digital environments has only recently begun to gain significant attention due to its potential for influencing behavior. Despite this growing interest, the long-term effects of digital nudges still need to be explored in the literature. However, one study has demonstrated a notable long-term impact, revealing that using multiple food nudges in a digital environment significantly improved the nutritional quality of food choices in an online canteen setting over a 15-month period [100].

### *D. DATASETS, MEASUREMENT, AND PARTICIPANT*

Food recommender system studies focus on both online and offline evaluations. Offline evaluation helps identify the best algorithms based on recommendation accuracy and requires a source dataset for experimentation [111]. To accurately assess the system's impact, selecting appropriate measurement metrics and involving suitable participants in the experiments is essential [112]. A comprehensive evaluation of a food recommender system must ensure both holistic assessment and research reproducibility. Therefore, we evaluate and extract the data source, measurement metrics, and number of participants from each eligible study [12], [113].

Among the 19 studies reviewed, Only six disclosed their data sources for conducting user experiments and analyses. The Allrecipes.com dataset was notably the most frequently utilized in five studies [42], [65], [94]–[96], [98]. A combination of various data sources was reported as the basis for one experiment [97]. The remaining studies opted to keep their datasets private, offering exploration only upon request.

There is a lack of precise details regarding the evaluation metrics for recommendation algorithm performance in the reviewed studies. Only one study provides concrete details on selecting the most accurate recommendation algorithms [98]. While other research suggests using state-of-the-art recommenders or widely adopted approaches appropriate for the context. Concerning health metrics, seven studies reported using international healthiness scores, such as the FSA score, Nutri-Score, or Eco-Score, to assess the healthiness of the datasets utilized and collected [11], [65], [93]–[96], [98].

Among the studies reviewed, all provided comprehensive details on the number of participants and the specific user groups involved. The exception was a study that failed to include such information [101]. Additionally, another study suggested conducting a user experiment as future work but did not specify the number or type of users to be involved [97]. Of the studies that did specify participant demographics, few focused on high school and university students [99], [102], whereas the remaining studies encompassed a range of diverse or general population groups. The sample sizes across these studies varied significantly, ranging from a minimum of 5 participants to a maximum of 1331, with a mean of 398.19 and a standard deviation of 425.60 participants.

## V. FUTURE OPPORTUNITIES

This section critically examines the gaps identified in current research regarding integrating food recommender systems and digital nudges to promote healthier food choices. The analysis is grounded in the reviewed state-of-the-art literature. Subsequently, we highlight emerging opportunities by analyzing the various dimensions of the reviewed studies alongside examples from other domains.

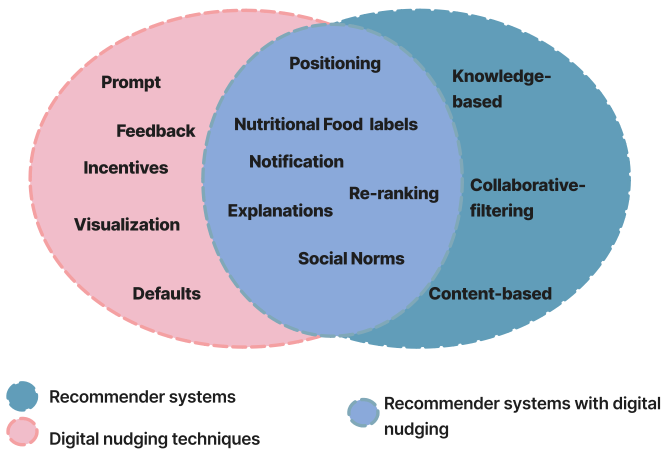



FIGURE 4: Explored digital nudging techniques for healthy food choices. The intersection highlights those implemented within food recommender systems.

Existing literature highlights the integration of various recommendation approaches with digital nudging techniques [39]. This study further indicates that several efforts have been made to incorporate such interventions into food recommender systems. Nudging, which involves shaping the design of choice architecture, should therefore be embedded in the presentation phase of the food recommendation process, irrespective of the recommendation algorithms or approaches applied [32], [35], [114], [115]. Out of the eleven nudging techniques we have identified, six techniques have already been tested in a recommender system context. In contrast, five of them have not yet been implemented in a recommender context. Furthermore, there are various combinations of multiple techniques and approaches that warrant further investigation [39]. Figure 4 offers a comprehensive overview of both the explored and unexplored digital nudging techniques within food recommender systems for healthy food choices.

Implementing digital nudges for behavioral change within recommender systems involves several key steps to ensure successful application and evaluation [115], [116]. The first step is to define the context and objectives of the nudge, with the aim of supporting users in making healthier food choices through the recommender interface. This is followed by a comprehensive diagnosis to understand the decision-making process, how user preferences are elicited, and the techniques used for generating and presenting recommendations [35], [115], [117]. Selecting appropriate nudges requires a thorough understanding of the recommendation process, including the types of interfaces users will interact with, the characteristics of the target audience, and the extent to which nudging strategies should be integrated [11], [65]. The final phase involves evaluating the impact of these nudges on user behavior. However, the evaluation should go beyond measuring behavioral changes, incorporating an analysis of the overall user experience and considering the system's complexity and other factors that may influence food choices [12]. A holistic approach to evaluation will help ensure that nudging strategies not only encourage healthier behaviors but also enhance the usability and effectiveness of the recommender system [112].

Extensive research on designing recommender systems has made selecting approaches for generating recommendations easier in recent years [9], [12]. While, choosing suitable nudging techniques to integrate into the recommendation process remains complex. This review identifies five nudging interventions that have not yet been employed in food recommender systems for healthy food choices. However, we believe that only those outlined below are worth further investigation [39]:

- *Default:* A nudge technique found the most prominent within nudging and behavioral change literature [115], [118]. It consists of a pre-selection of a given choice. Within recommender systems, it has proven to be the most studied technique across various fields [116]. To promote healthy choices within the food recommender system, setting the defaults involves pre-selecting the healthiest options from the generated food recommendations for a given use based on the health metric used to evaluate the recipes. Several studies have explored the impact of default settings on user behavior within recommender systems [119]. Starting with a default option often leads users to select it or a similar choice, while also shaping user preferences, which has the potential to encourage healthier behaviors [116], [120]. For example, one study found that users were satisfied when the most sustainable and environmentally friendly trip routes with lower $CO_2$ emissions were set as the default in a route recommender system [121]. In the food domain, nudging users toward healthier food choices through default settings is worth the exploration regarding the potential of this technique to influence user behaviors in other domains [122].
- *Visualization:* Interventions that increase the salience of items, such as enhancing image quality or emphasizing text, are designed to capture users' attention [39], [123]. In food recommender systems, visual enhancements—such as high-quality images, highlighting the nutritional attributes, or creating graphics for healthier options—can support more informed and health-conscious decision-making in addition to increasing user satisfaction [42], [124]. For example, visual cues have successfully encouraged sustainable fashion consumption by drawing attention to second-hand garments and providing sustainability information [125]. Similarly, in book recommendations, nudging users toward off-profile content has increased the selection of books beyond their preferred genres [126].
- *Prompt:* A prompt is defined as a strategically delivered message or reminder aimed at users in specific contexts to influence behavior [127]. It is frequently employed to address situations where individuals are prone to engage in undesirable or unhealthy behaviors, providing timely guidance toward healthier alternatives. The use of prompts has been shown to facilitate effective behavioral change across various domains, including transportation, healthcare, environmental conservation, and education, both in online and offline settings [128]–[130]. The integration of prompts within recommender systems has not been extensively explored in the literature [39]. Nevertheless, incorporating reminders to encourage healthier food choices within these systems can be approached in various ways, such as prompting users to consider more nutritious alternatives during decision-making processes. In conversational recommender systems, prompts can further enhance the user experience by facilitating reflection on their choices through explanations and cues [131], thereby supporting more informed and health-conscious decision-making.

Feedback interventions are used as nudging strategies in areas like education and nutrition [80], [132] to provide information about past behaviors to guide future actions [133]. However, integrating them into recommender systems is challenging due to the complexity and risk of increasing cognitive load. Research shows that feedback can sometimes have negative effects [134], [135], while incentives, often applied after a decision, may offer only short-term behavior changes, especially monetary ones. Furthermore, incentives can undermine intrinsic motivation, making it harder to sustain desired behaviors [136], [137]. Implementing such interventions in personalized food recommender systems adds complexities that may overwhelm users [138].

Digital nudges are strategically designed to influence user interactions with the interface and ought to be integrated during the final phase of the recommendation process (i.e., Figure 1 Part (C): representation).

The current study elucidates several areas requiring further exploration. Foremost among these is the insufficient evaluation of used datasets and their quality in compliance with established international health standards [139]. This gap in the assessment process undermines the reliability validity, and reproducibility of the proposed research. Addition-

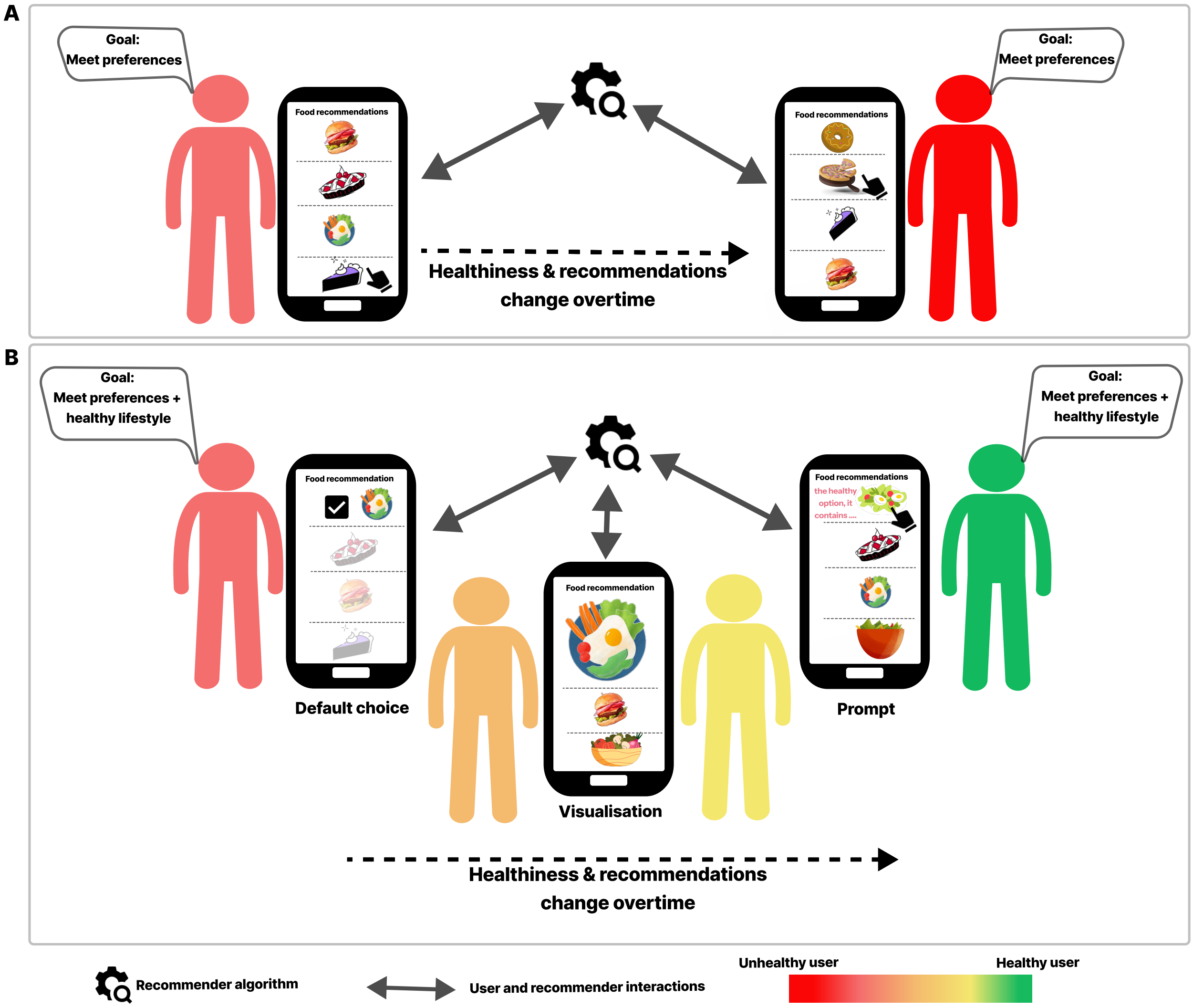



FIGURE 5: The impact of integrating digital nudging techniques into a food recommender system over time. (A) depicts the system without nudging, while (B) illustrates changes in generated recommendations and user choices when nudging is applied.

ally, there is a notable lack of application for user-centered experimental designs, which are crucial for understanding the real-world efficacy of these systems. As demonstrated in [112], [140], incorporating user evaluations remains the most robust method for accurately assessing the performance and effectiveness of recommender systems. Addressing these issues in future research is essential for advancing the field and improving the practical utility of food recommmender systems integrated with digital nudges for healthier choices.

Another crucial area for further research is the long-term impact of nudging techniques on users' lifestyles and behavior [116], [141]. We are confident that these techniques contribute to a deeper understanding of user decision-making, ultimately fostering awareness and supporting informed choices. Additionally, they have been shown to induce positive habit changes, a benefit well-documented in offline settings [142], [143].

Furthermore, integrating nudging techniques into recommender systems can enhance their effectiveness by promoting diverse and healthier recommendations. Figure 5 illustrates this concept with an example: (A) depicts user interaction with a recommender system without nudging, whereas (B) demonstrates how incorporating nudging techniques can guide users toward healthier choices. Over time, this approach not only benefits individual users but also enables the recommender system to prioritize healthier options through continuous learning from user data.

## VI. CONCLUSION

Food recommender systems have demonstrated their utility in helping users find desired food content efficiently and alleviate choice overload. Content-based approaches have integrated several nudging interventions, such as nutritional labels and explanations, whereas collaborative filtering and knowledge-based approaches have primarily been tested with nutritional food labels and notification-based nudges. This integration has shown to enhance the healthiness of food choices within these systems. However, various other interventions could be harnessed to further support users in making healthier food choices across different recommendation approaches. For instance, default interventions, which have yielded promising results in other domains, could be applied here. Additionally, visual techniques offer potential; enhancements in image quality or the use of visualization tools can facilitate easier judgments of food healthiness. Consequently, these strategies could contribute to improved user satisfaction and experience within food recommender systems.

This study revealed that various nudging techniques effectively support users in making healthier and more informed food choices. Additionally, incorporating interventions during the presentation phase of food recommender systems has demonstrated significant potential in achieving this goal. The findings from this study contribute to the existing literature by providing a comprehensive summary and review of food recommender systems that successfully integrate digital nudging interventions to promote healthier food choices and encourage healthy eating behaviors.

Using a pre-validated research framework, we systematically analyzed and categorized various nudging techniques within digital environments. Furthermore, we critically assessed digital nudging strategies that warrant further investigation and explored their potential integration into the food recommendation process to support healthier eating. Lastly, we highlighted the primary limitations identified in existing studies and emphasized the importance of user-centric evaluations for assessing the performance of these systems across various dimensions.

## VII. ACKNOWLEDGEMENTS

The authors acknowledge the use of ChatGPT [144] for checking and correcting the grammar of this article. It is important to note that they did not use ChatGPT to generate totally new text, rather they use it to check/correct grammar of the provided text.

## REFERENCES

[1] Z. Dong, Z. Wang, J. Xu, R. Tang, and J. Wen, "A brief history of recommender systems," arXiv preprint arXiv:2209.01860, 2022.
[2] D. Jannach and G. Adomavicius, "Recommendations with a purpose," in Proceedings of the 10th ACM conference on recommender systems, 2016, pp. 7–10.
[3] P. Mahajan and P. D. Kaur, "A systematic literature review of food recommender systems," SN Computer Science, vol. 5, no. 1, p. 174, 2024.
[4] D. Bollen, B. P. Knijnenburg, M. C. Willemsen, and M. Graus, "Understanding choice overload in recommender systems," in Proceedings of the fourth ACM conference on Recommender systems, 2010, pp. 63–70.
[5] A. Chernev, U. Böckenholt, and J. Goodman, "Choice overload: A conceptual review and meta-analysis," Journal of Consumer Psychology, vol. 25, no. 2, pp. 333–358, 2015.
[6] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in Recommender systems handbook. Springer, 2010, pp. 1–35.
[7] D. Jannach and M. Jugovac, "Measuring the business value of recommender systems," ACM Transactions on Management Information Systems (TMIS), vol. 10, no. 4, pp. 1–23, 2019.
[8] I. A. Christensen and S. Schiaffino, "Entertainment recommender systems for group of users," Expert systems with applications, vol. 38, no. 11, pp. 14 127–14 135, 2011.
[9] F. Ricci, L. Rokach, and B. Shapira, Recommender Systems Handbook, 3rd ed. Springer, 2022.
[10] P. M. Connell and L. F. Mayor, "Activating health goals reduces (increases) hedonic evaluation of food brands for people who harbor highly positive (negative) affect toward them," Appetite, vol. 65, pp. 159–164, 2013.
[11] A. Starke, E. Asotic, and C. Trattner, ""serving each user": Supporting different eating goals through a multi-list recommender interface," in Proceedings of the 15th ACM conference on recommender systems, 2021, pp. 124–132.
[12] D. Elsweiler, H. Hauptmann, and C. Trattner, Food Recommender Food recommenderSystems. New York, NY: Springer US, 2022, pp. 871–925. [Online]. Available: https://doi.org/10.1007/978-1-0716-2197-4_23
[13] A. Robertson, "Food and health in europe: a new basis for action," 2004.
[14] C. Trattner, D. Parra, and D. Elsweiler, "Monitoring obesity prevalence in the united states through bookmarking activities in online food portals," PloS one, vol. 12, no. 6, p. e0179144, 2017.
[15] J. Freyne and S. Berkovsky, "Intelligent food planning: personalized recipe recommendation," in Proceedings of the 15th international conference on Intelligent user interfaces, 2010, pp. 321–324.
[16] B. Scheibehenne, L. Miesler, and P. M. Todd, "Fast and frugal food choices: Uncovering individual decision heuristics," Appetite, vol. 49, no. 3, pp. 578–589, 2007.
[17] D. S. Risso, C. Giuliani, M. Antinucci, G. Morini, P. Garagnani, S. Tofanelli, and D. Luiselli, "A bio-cultural approach to the study of food choice: The contribution of taste genetics, population and culture," Appetite, vol. 114, pp. 240–247, 2017.
[18] R. Shepherd, "Social determinants of food choice," Proceedings of the Nutrition Society, vol. 58, no. 4, pp. 807–812, 1999.
[19] E. C. Monterrosa, E. A. Frongillo, A. Drewnowski, S. de Pee, and S. Vandevijvere, "Sociocultural influences on food choices and implications for sustainable healthy diets," Food and Nutrition Bulletin, vol. 41, no. 2_suppl, pp. 59S–73S, 2020.
[20] P.-J. Chen and M. Antonelli, "Conceptual models of food choice: influential factors related to foods, individual differences, and society," Foods, vol. 9, no. 12, p. 1898, 2020.
[21] S. J. Hardcastle, C. Thøgersen-Ntoumani, and N. L. Chatzisarantis, "Food choice and nutrition: A social psychological perspective," pp. 8712–8715, 2015.
[22] M. T. Conner, "Understanding determinants of food choice: contributions from attitude research," British food journal, vol. 95, no. 9, pp. 27–31, 1993.
[23] E. P. Köster, "Diversity in the determinants of food choice: A psychological perspective," Food quality and preference, vol. 20, no. 2, pp. 70–82, 2009.
[24] D. Elsweiler, H. Hauptmann, and C. Trattner, "Food recommender systems," in Recommender systems handbook. Springer, 2012, pp. 871–925.
[25] C. Trattner and D. Elsweiler, "Food recommender systems: important contributions, challenges and future research directions," arXiv preprint arXiv:1711.02760, 2017.
[26] M. Chen, X. Jia, E. Gorbonos, C. T. Hoang, X. Yu, and Y. Liu, "Eating healthier: Exploring nutrition information for healthier recipe recommendation," Information Processing & Management, vol. 57, no. 6, p. 102051, 2020.
[27] C. Trattner and D. Elsweiler, "Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems," in Proceedings of the 26th international conference on world wide web, 2017, pp. 489–498.
[28] C. Trattner, D. Moesslang, and D. Elsweiler, "On the predictability of the popularity of online recipes," EPJ Data Science, vol. 7, no. 1, pp. 1–39, 2018.
[29] M. Rokicki, C. Trattner, and E. Herder, "The impact of recipe features, social cues and demographics on estimating the healthiness of online recipes," in Proceedings of the international AAAI conference on web and social media, vol. 12, no. 1, 2018.
[30] J. R. Bettman, M. F. Luce, and J. W. Payne, "Constructive consumer choice processes," Journal of consumer research, vol. 25, no. 3, pp. 187–217, 1998.
[31] A. D. Starke, E. Asotic, C. Trattner, and E. J. Van Loo, "Examining the user evaluation of multi-list recommender interfaces in the context of healthy recipe choices," ACM Transactions on Recommender Systems, vol. 1, no. 4, pp. 1–31, 2023.

[32] A. Starke, M. Willemsen, and C. Snijders, “Effective user interface designs to increase energy-efficient behavior in a rasch-based energy recommender system,” in Proceedings of the eleventh ACM conference on recommender systems, 2017, pp. 65–73.
[33] R. H. Thaler and C. R. Sunstein, Nudge: The final edition. Yale University Press, 2021.
[34] S. Mertens, M. Herberz, U. J. Hahnel, and T. Brosch, “The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains,” Proceedings of the National Academy of Sciences, vol. 119, no. 1, p. e2107346118, 2022.
[35] M. Weinmann, C. Schneider, and J. v. Brocke, “Digital nudging,” Business & Information Systems Engineering, vol. 58, pp. 433–436, 2016.
[36] D. Arigo and J. M. Suls, “Smartphone apps providing social comparison for health behavior change: a need for better tailoring to person and context,” Mhealth, vol. 4, 2018.
[37] W. Fechner and E. Herder, “Digital nudging for more ecological supermarket purchases,” in Adjunct proceedings of the 29th ACM conference on user modeling, adaptation and personalization, 2021, pp. 284–292.
[38] H. Ensaff, M. Homer, P. Sahota, D. Braybrook, S. Coan, and H. McLeod, “Food choice architecture: an intervention in a secondary school and its impact on students’ plant-based food choices,” Nutrients, vol. 7, no. 6, pp. 4426–4437, 2015.
[39] M. Jesse and D. Jannach, “Digital nudging with recommender systems: Survey and future directions,” Computers in Human Behavior Reports, vol. 3, p. 100052, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S245195882030052X
[40] C. Musto, A. D. Starke, C. Trattner, A. Rapp, and G. Semeraro, “Exploring the effects of natural language justifications in food recommender systems,” in Proceedings of the 29th ACM conference on user modeling, adaptation and personalization, 2021, pp. 147–157.
[41] M. A. Khan, K. Muhammad, B. Smyth, and D. Coyle, “Investigating health-aware smart-nudging with machine learning to help people pursue healthier eating-habits,” arXiv preprint arXiv:2110.07045, 2021.
[42] A. D. Starke, E. K. Kløverød Brynestad, S. Hauge, and L. S. Løkeland, “Nudging healthy choices in food search through list re-ranking,” in Adjunct proceedings of the 29th ACM conference on user modeling, adaptation and personalization, 2021, pp. 293–298.
[43] G. Leng, R. A. Adan, M. Belot, J. M. Brunstrom, K. De Graaf, S. L. Dickson, T. Hare, S. Maier, J. Menzies, H. Preissl et al., “The determinants of food choice,” Proceedings of the Nutrition Society, vol. 76, no. 3, pp. 316–327, 2017.
[44] J. Sobal and C. A. Bisogni, “Constructing food choice decisions,” Annals of behavioral medicine, vol. 38, no. suppl_1, pp. s37–s46, 2009.
[45] J. Sobal, C. A. Bisogni, and M. Jastran, “Food choice is multifaceted, contextual, dynamic, multilevel, integrated, and diverse,” Mind, Brain, and Education, vol. 8, no. 1, pp. 6–12, 2014.
[46] E. Randall and D. Sanjur, “Food preferences—their conceptualization and relationship to consumption,” Ecology of food and nutrition, vol. 11, no. 3, pp. 151–161, 1981.
[47] D. Asioli, J. Aschemann-Witzel, V. Caputo, R. Vecchio, A. Annunziata, T. Næs, and P. Varela, “Making sense of the “clean label” trends: A review of consumer food choice behavior and discussion of industry implications,” Food Research International, vol. 99, pp. 58–71, 2017.
[48] A. Hoek, D. Pearson, S. James, M. Lawrence, and S. Friel, “Healthy and environmentally sustainable food choices: Consumer responses to point-of-purchase actions,” Food quality and preference, vol. 58, pp. 94–106, 2017.
[49] B. Kumar, A. K. Manrai, and L. A. Manrai, “Purchasing behaviour for environmentally sustainable products: A conceptual framework and empirical study,” Journal of retailing and consumer services, vol. 34, pp. 1–9, 2017.
[50] S. M. Krebs-Smith and L. S. Kantor, “Choose a variety of fruits and vegetables daily: understanding the complexities,” The Journal of nutrition, vol. 131, no. 2, pp. 487S–501S, 2001.
[51] D. Rose, J. N. Bodor, P. L. Hutchinson, and C. M. Swalm, “The importance of a multi-dimensional approach for studying the links between food access and consumption,” The Journal of nutrition, vol. 140, no. 6, pp. 1170–1174, 2010.
[52] T. Furst, M. Connors, C. A. Bisogni, J. Sobal, and L. W. Falk, “Food choice: a conceptual model of the process,” Appetite, vol. 26, no. 3, pp. 247–266, 1996.
[53] M. Connors, C. A. Bisogni, J. Sobal, and C. M. Devine, “Managing values in personal food systems,” Appetite, vol. 36, no. 3, pp. 189–200, 2001.
[54] S. Gutjar, C. de Graaf, V. Kooijman, R. A. de Wijk, A. Nys, G. J. Ter Horst, and G. Jager, “The role of emotions in food choice and liking,” Food Research International, vol. 76, pp. 216–223, 2015.
[55] V. J. van Buul, C. A. Bolman, F. J. Brouns, and L. Lechner, “Back-of-pack information in substitutive food choices: A process-tracking study in participants intending to eat healthy,” Appetite, vol. 116, pp. 173–183, 2017.
[56] C. Contini, F. Boncinelli, E. Marone, G. Scozzafava, and L. Casini, “Drivers of plant-based convenience foods consumption: Results of a multicomponent extension of the theory of planned behaviour,” Food Quality and Preference, vol. 84, p. 103931, 2020.
[57] S. Stranieri, E. C. Ricci, and A. Banterle, “Convenience food with environmentally-sustainable attributes: A consumer perspective,” Appetite, vol. 116, pp. 11–20, 2017.
[58] M. Harvey, B. Ludwig, and D. Elsweiler, “You are what you eat: Learning user tastes for rating prediction,” in String Processing and Information Retrieval: 20th International Symposium, SPIRE 2013, Jerusalem, Israel, October 7-9, 2013, Proceedings 20. Springer, 2013, pp. 153–164.
[59] J. N. Bondevik, K. E. Bennin, Ö. Babur, and C. Ersch, “A systematic review on food recommender systems,” Expert Systems with Applications, p. 122166, 2023.
[60] J. N. Bondevik, K. E. Bennin, Önder Babur, and C. Ersch, “A systematic review on food recommender systems,” Expert Systems with Applications, vol. 238, p. 122166, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423026684
[61] K. Sharma, K. V. Mandapati, M. C. Pattekar, K. T. Kumar, and G. Srinivasa, “Food recommendation system based on collaborative filtering and taste profiling,” in 2023 3rd International Conference on Intelligent Technologies (CONIT). IEEE, 2023, pp. 1–6.
[62] N. Rajabpour, A. Naserasadi, and M. Estilayee, “Tfr: a tourist food recommender system based on collaborative filtering,” Int J Comput Appl, vol. 181, no. 11, pp. 30–39, 2018.
[63] M. J. Pazzani and D. Billsus, “Content-based recommendation systems,” in The adaptive web: methods and strategies of web personalization. Springer, 2007, pp. 325–341.
[64] C. Musto, C. Trattner, A. Starke, and G. Semeraro, “Towards a knowledge-aware food recommender system exploiting holistic user models,” in Proceedings of the 28th ACM conference on user modeling, adaptation and personalization, 2020, pp. 333–337.
[65] A. El Majjodi, A. D. Starke, M. Elahi, C. Trattner et al., “The interplay between food knowledge, nudges, and preference elicitation methods determines the evaluation of a recipe recommender system.” in IntRS@ RecSys, 2023, pp. 1–18.
[66] P. Chavan, B. Thoms, and J. Isaacs, “A recommender system for healthy food choices: building a hybrid model for recipe recommendations using big data sets,” 2021.
[67] R. Sugden, “On nudging: A review of nudge: Improving decisions about health, wealth and happiness by richard h. thaler and cass r. sunstein,” 2009.
[68] B. Wildavsky, “Nudge nation: A new way to prod students into and through college,” ES Select, vol. 10, 2013.
[69] M. Li and G. B. Chapman, “Nudge to health: Harnessing decision research to promote health behavior,” Social and Personality Psychology Compass, vol. 7, no. 3, pp. 187–198, 2013.
[70] R. Vecchio and C. Cavallo, “Increasing healthy food choices through nudges: A systematic review,” Food Quality and Preference, vol. 78, p. 103714, 2019.
[71] T. Bucher, C. Collins, M. E. Rollo, T. A. McCaffrey, N. De Vlieger, D. Van der Bend, H. Truby, and F. J. Perez-Cueto, “Nudging consumers towards healthier choices: a systematic review of positional influences on food choice,” British Journal of Nutrition, vol. 115, no. 12, pp. 2252–2263, 2016.
[72] R. Cadario and P. Chandon, “Which healthy eating nudges work best? a meta-analysis of field experiments,” Marketing Science, vol. 39, no. 3, pp. 465–486, 2020.
[73] S. Benartzi, The smarter screen: Surprising ways to influence and improve online behavior. Penguin, 2017.
[74] M. Valta, J. Menzel, C. Maier, K. Pflügner, M. Meier, and T. Weitzel, “Digital nudging: A systematic literature review and future research directions,” in Proceedings of the 2022 Computers and People Research Conference, 2022, pp. 1–10.
[75] L. Michels, J. Ochmann, K. Schmitt, S. Laumer, and V. Tiefenbeck, “Salience, transparency, and self-nudging: a digital nudge to promote

healthier food product choices," European Journal of Information Systems, pp. 1–31, 2023.

[76] R. Y. Toledo, A. A. Alzahrani, and L. Martinez, "A food recommender system considering nutritional information and user preferences," IEEE Access, vol. 7, pp. 96 695–96 711, 2019.

[77] A. D. Starke, C. Musto, A. Rapp, G. Semeraro, and C. Trattner, ""tell me why": using natural language justifications in a recipe recommender system to support healthier food choices," User Modeling and User-Adapted Interaction, vol. 34, no. 2, pp. 407–440, 2024.

[78] J. Piper, M. T. Adam, N. De Vlieger, C. Collins, and T. Bucher, "A bibliometric review of digital nudging within digital food choice environments," 2021.

[79] E. Valenčič, E. Beckett, C. E. Collins, B. K. Seljak, and T. Bucher, "Digital nudging in online grocery stores: A scoping review on current practices and gaps," Trends in Food Science & Technology, vol. 131, pp. 151–163, 2023.

[80] S. Forberger, L. A. Reisch, P. van Gorp, C. Stahl, L. Christianson, J. Halimi, K. K. De Santis, L. Malisoux, T. de Magistris, and T. Bohn, "'let me recommend…': use of digital nudges or recommender systems for overweight and obesity prevention—a scoping review protocol," BMJ open, vol. 14, no. 7, p. e080644, 2024.

[81] M. Bartmann, "Reasoning with recommender systems? practical reasoning, digital nudging, and autonomy," in Recommender Systems: Legal and Ethical Issues. Springer International Publishing Cham, 2023, pp. 129–145.

[82] N. B. Ali and K. Petersen, "Evaluating strategies for study selection in systematic literature studies," in Proceedings of the 8th ACM/IEEE international symposium on empirical software engineering and measurement, 2014, pp. 1–4.

[83] S. Keele et al., "Guidelines for performing systematic literature reviews in software engineering," Technical report, ver. 2.3 ebse technical report. ebse, Tech. Rep., 2007.

[84] K. Bergram, M. Djokovic, V. Bezençon, and A. Holzer, "The digital landscape of nudging: A systematic literature review of empirical research on digital nudges," in Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 2022, pp. 1–16.

[85] "Acm digital library," 2024, accessed: May 30, 2024. [Online]. Available: https://dl.acm.org/

[86] "Ieee xplore libraryh," 2024, accessed: May 28, 2024. [Online]. Available: https://ieeexplore.ieee.org/Xplore/home.jsp

[87] "Sciencedirect," 2024, accessed: June 6, 2024. [Online]. Available: https://www.sciencedirect.com/

[88] "Springerlink," 2024, accessed: May 28, 2024. [Online]. Available: SpringerLink

[89] "Pubmed," 2024, accessed: June 6, 2024. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/

[90] C. T. Ayoub EL Majjodi, Alain D. Strake, "Digital nudging and recommender system for healthier food choices: Srl process," 2024. [Online]. Available: https://bit.ly/3X15kXL

[91] ——. (2024) Digital nudging and recommender system for healthier food choices: Srl extracte papers. [Online]. Available: https://docs.google.com/spreadsheets/d/1RYoZj1FRWFiSSrKAqnx0W4Hwcs7E328sFN0QcOaOqXo/edit?usp=sharing

[92] ——, "Digital nudging and recommender system for healthier food choices: Srl data extraction from," 2024. [Online]. Available: https://forms.gle/AMcVT72G7Q3WaAy9A

[93] M. De Bauw, L. S. De La Revilla, V. Poppe, C. Matthys, and L. Vranken, "Digital nudges to stimulate healthy and pro-environmental food choices in e-groceries," Appetite, vol. 172, p. 105971, 2022.

[94] G. Castiglia, A. El Majjodi, A. D. Starke, F. Narducci, Y. Deldjoo, and F. Calò, "Nudging towards health in a conversational food recommender system using multi-modal interactions and nutrition labels," in Fourth Knowledge-aware and Conversational Recommender Systems Workshop (KaRS), 2022.

[95] D. Elsweiler, C. Trattner, and M. Harvey, "Exploiting food choice biases for healthier recipe recommendation," in Proceedings of the 40th international acm sigir conference on research and development in information retrieval, 2017, pp. 575–584.

[96] A. Starke, A. El Majjodi, and C. Trattner, "Boosting health? examining the role of nutrition labels and preference elicitation methods in food recommendation," in Interfaces and Human Decision Making for Recommender Systems 2022, 2022, pp. 67–84.

[97] P. K. Prasetyo, P. Achananuparp, and E.-P. Lim, "Foodbot: A goal-oriented just-in-time healthy eating interventions chatbot," in Proceedings of the 14th EAI International Conference on Pervasive Computing Technologies for Healthcare, 2020, pp. 436–439.

[98] A. El Majjodi, A. D. Starke, and C. Trattner, "Nudging towards health? examining the merits of nutrition labels and personalization in a recipe recommender system," in Proceedings of the 30th ACM conference on user modeling, adaptation and personalization, 2022, pp. 48–56.

[99] T. Delaney, S. L. Yoong, H. Lamont, C. Lecathelinais, L. Wolfenden, T. Clinton-McHarg, R. Sutherland, and R. Wyse, "The efficacy of a multi-strategy choice architecture intervention on improving the nutritional quality of high school students' lunch purchases from online canteens (click & crunch high schools): a cluster randomized controlled trial," International Journal of Behavioral Nutrition and Physical Activity, vol. 19, no. 1, p. 120, 2022.

[100] T. Delaney, J. Jackson, C. Lecathelinais, T. Clinton-McHarg, H. Lamont, S. L. Yoong, L. Wolfenden, R. Sutherland, and R. Wyse, "Long-term effectiveness of a multi-strategy choice architecture intervention in increasing healthy food choices of high-school students from online canteens (click & crunch high schools): Cluster randomized controlled trial," Journal of medical Internet research, vol. 26, p. e51108, 2024.

[101] R. Nakaoka, Y. Nakamura, Y. Matsuda, S. Misaki, and K. Yasumoto, "eat2pic: Food-tech design as a healthy nudge with smart chopsticks and canvas," in 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops). IEEE, 2021, pp. 389–391.

[102] M. C. Bomfim, S. I. Kirkpatrick, L. E. Nacke, and J. R. Wallace, "Food literacy while shopping: Motivating informed food purchasing behaviour with a situated gameful app," in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, pp. 1–13.

[103] C. Aigner, G. Hofmann, S. Winkler, R. Baranyi, and T. Grechenig, "Nutrition garden-a gamified mobile app for motivating people to eat specific food to prevent non-communicable diseases," in Proceedings of the 2023 7th International Conference on Medical and Health Informatics, 2023, pp. 203–207.

[104] P. Agyemang, E. M. Kwofie, J. I. Baum, and D. Wang, "The design and development of a dashboard for improving sustainable healthy food choices," Science of The Total Environment, vol. 930, p. 172726, 2024.

[105] K. S.-P. Chang, C. M. Danis, and R. G. Farrell, "Lunch line: using public displays and mobile devices to encourage healthy eating in an organization," in Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2014, pp. 823–834.

[106] Y. Nakamura, "Iot nudge: Iot data-driven nudging for health behavior change," in Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers, 2021, pp. 51–53.

[107] E. Valenčič, E. Beckett, C. E. Collins, B. K. Seljak, and T. Bucher, "Changing the default order of food items in an online grocery store may nudge healthier food choices," Appetite, vol. 192, p. 107072, 2024.

[108] B. M. DiCosola III and G. Neff, "Nudging behavior change: Using in-group and out-group social comparisons to encourage healthier choices," in Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 2022, pp. 1–14.

[109] S. Higgs, "Social norms and their influence on eating behaviours," Appetite, vol. 86, pp. 38–44, 2015.

[110] A. S. Ivanic, "To choose (not) to eat healthy: Social norms, self-affirmation, and food choice," Psychology & Marketing, vol. 33, no. 8, pp. 595–607, 2016.

[111] G. Shani and A. Gunawardana, "Evaluating recommendation systems," Recommender systems handbook, pp. 257–297, 2011.

[112] B. P. Knijnenburg and M. C. Willemsen, "Evaluating recommender systems with user experiments," in Recommender systems handbook. Springer, 2015, pp. 309–352.

[113] J. Beel, C. Breitinger, S. Langer, A. Lommatzsch, and B. Gipp, "Towards reproducibility in recommender-systems research," User modeling and user-adapted interaction, vol. 26, pp. 69–101, 2016.

[114] T. Mirsch, C. Lehrer, and R. Jung, "Making digital nudging applicable: The digital nudge design method," 2018.

[115] ——, "Digital nudging: Altering user behavior in digital environments," 2017.

[116] M. Jesse and D. Jannach, "Digital nudging with recommender systems: Survey and future directions," Computers in Human Behavior Reports, vol. 3, p. 100052, 2021.

[117] E. Murphy-Hill and G. C. Murphy, "Recommendation delivery: Getting the user interface just right," in Recommendation systems in software engineering. Springer, 2013, pp. 223–242.

[118] H. Cronqvist and R. H. Thaler, "Design choices in privatized social-security systems: Learning from the swedish experience," American Economic Review, vol. 94, no. 2, pp. 424–428, 2004.

[119] Y. Liang and M. C. Willemsen, "Promoting music exploration through personalized nudging in a genre exploration recommender," International Journal of Human–Computer Interaction, vol. 39, no. 7, pp. 1495–1518, 2023.

[120] N. Dhingra, Z. Gorn, A. Kener, and J. Dana, "The default pull: An experimental demonstration of subtle default effects on preferences," Judgment and Decision Making, vol. 7, no. 1, pp. 69–76, 2012.

[121] E. Bothos, D. Apostolou, and G. Mentzas, "Recommender systems for nudging commuters towards eco-friendly decisions," Intelligent Decision Technologies, vol. 9, no. 3, pp. 295–306, 2015.

[122] J. M. Jachimowicz, S. Duncan, E. U. Weber, and E. J. Johnson, "When and why defaults influence decisions: A meta-analysis of default effects," Behavioural Public Policy, vol. 3, no. 2, pp. 159–186, 2019.

[123] J. Kumar and N. Tintarev, "Using visualizations to encourage blind-spot exploration." in IntRS@ RecSys, 2018, pp. 53–60.

[124] S. Bostandjiev, J. O'Donovan, and T. Höllerer, "Tasteweights: a visual interactive hybrid recommender system," in Proceedings of the sixth ACM conference on Recommender systems, 2012, pp. 35–42.

[125] A. Geninatti Cossatin, "Promoting green fashion consumption in recommender systems," in Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, 2024, pp. 50–54.

[126] G. Alves, D. Jannach, R. F. de Souza, D. Damian, and M. G. Manzato, "Digitally nudging users to explore off-profile recommendations: here be dragons," User Modeling and User-Adapted Interaction, vol. 34, no. 2, pp. 441–481, 2024.

[127] H. Thillmann, J. Künsting, J. Wirth, and D. Leutner, "Is it merely a question of "what" to prompt or also "when" to prompt? the role of point of presentation time of prompts in self-regulated learning," Zeitschrift für Pädagogische Psychologie, vol. 23, no. 2, pp. 105–115, 2009.

[128] W. Hardeman, J. Houghton, K. Lane, A. Jones, and F. Naughton, "A systematic review of just-in-time adaptive interventions (jitais) to promote physical activity," International Journal of Behavioral Nutrition and Physical Activity, vol. 16, pp. 1–21, 2019.

[129] B. Hedin, C. Katzeff, E. Eriksson, and D. Pargman, "A systematic review of digital behaviour change interventions for more sustainable food consumption," Sustainability, vol. 11, no. 9, p. 2638, 2019.

[130] C. A. B. Andrews, Always Something There to Remind Me: The Role of Nudging in Assignment Submission. Brigham Young University, 2020.

[131] Y. Ma and J. Ziegler, "The effect of proactive cues on the use of decision aids in conversational recommender systems," in Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, 2024, pp. 305–315.

[132] L. Zamprogno, R. Holmes, and E. Baniassad, "Nudging student learning strategies using formative feedback in automatically graded assessments," in Proceedings of the 2020 ACM SIGPLAN Symposium on Splash-E, 2020, pp. 1–11.

[133] L. M. Y. Chung, S. S. Fong, A. Ma, and S.-T. Cheng, "Integration of behavioural feedback in web-based systems nutrition learning among university students," Journal of Computer Assisted Learning, vol. 34, no. 4, pp. 450–457, 2018.

[134] D. Dowding, J. Merrill, and D. Russell, "Using feedback intervention theory to guide clinical dashboard design," in AMIA Annual Symposium Proceedings, vol. 2018. American Medical Informatics Association, 2018, p. 395.

[135] Y. Jin, B. D. L. R. P. Cardoso, and K. Verbert, "How do different levels of user control affect cognitive load and acceptance of recommendations?" in IntRS@ RecSys, 2017, pp. 35–42.

[136] U. Gneezy, A. Kajackaite, S. Meier et al., "Incentive-based interventions," in The handbook of behavior change. Cambridge University Press, 2020, pp. 523–536.

[137] U. Gneezy, S. Meier, and P. Rey-Biel, "When and why incentives (don't) work to modify behavior," Journal of economic perspectives, vol. 25, no. 4, pp. 191–210, 2011.

[138] F. Yao, C. Li, K. A. Sankararaman, Y. Liao, Y. Zhu, Q. Wang, H. Wang, and H. Xu, "Rethinking incentives in recommender systems: are monotone rewards always beneficial?" Advances in Neural Information Processing Systems, vol. 36, 2024.

[139] C. Trattner, D. Elsweiler, and S. Howard, "Estimating the healthiness of internet recipes: a cross-sectional study," Frontiers in public health, vol. 5, p. 16, 2017.

[140] L. Chen and P. Pu, "User evaluation framework of recommender systems," in Workshop SRS, vol. 10, 2010.

[141] A. Caraban, E. Karapanos, D. Gonçalves, and P. Campos, "23 ways to nudge: A review of technology-mediated nudging in human-computer interaction," in Proceedings of the 2019 CHI conference on human factors in computing systems, 2019, pp. 1–15.

[142] H. Allcott and T. Rogers, "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation," American Economic Review, vol. 104, no. 10, pp. 3003–3037, 2014.

[143] J. Beshears and H. Kosowsky, "Nudging: Progress to date and future directions," Organizational behavior and human decision processes, vol. 161, pp. 3–19, 2020.

[144] OpenAI, "Chatgpt," 2024, accessed: 2024-09-16. [Online]. Available: https://openai.com/chatgpt/



AYOUB EL MAJJODI is currently a PhD student in the DARS research group at the University of Bergen. He obtained his MSc in Data Science and Big Data from The National Advanced School of Computer Science and System Analysis, Morocco. The main focus of his PhD research project lies within Recommender Systems for the food domain, where he investigates the utility of Recommender Systems and Digital nudging to change people's eating behaviors and help them to attain healthier eating habits.

ASSOC. PROF. DR. ALAIN D. STARKE graduated as an engineer in Innovation Sciences and obtained a PhD in Human-Technology Interaction at TU Eindhoven, NL. His research focuses on how people can be presented effective personalized advice, through both algorithms and interfaces, focusing on domains that involve behavioral change, such as energy conservation and healthy eating. He is currently an assistant professor at the Amsterdam School for Communication Research, at the University of Amsterdam, NL.

PROF. DR. CHRISTOPH TRATTNER received the B.Sc. degree in computer science and telematics and the M.Sc. and Ph.D. degrees (Hons.) from the Graz University of Technology, Austria. He is a Lead Professor in Data Science and Recommender Systems at the University of Bergen and the Director of the Research Centre for Responsible Media Technology & Innovation — MediaFutures worth around 26 million EUR. His research lies in the intersection of "Behavioral Data Analytics" and "Recommender Systems." He published over 150s paper in reputable journals including NATURE Food and is a ACM Distinguished Speaker and board member of several leading journals in computer science.


• • •

# Manuscript II

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

*Ayoub El Majjodi, Alain D. Starke, Christoph Tratter. UMAP2022*

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

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

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

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

## ABSTRACT

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

## CCS CONCEPTS

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

## KEYWORDS

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

**ACM Reference Format:**
Ayoub El Majjodi, Alain Starke, and Christoph Trattner. 2022. Nudging Towards Health? Examining the Merits of Nutrition Labels and Personalization in a Recipe Recommender System. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '22), July 4–7, 2022, Barcelona, Spain.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3503252.3531312



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

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

## 1 INTRODUCTION

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

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

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

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

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

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

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

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

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

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

## 2 RELATED WORK

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

### 2.1 Recommender Systems

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

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

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

### 2.2 Digital Nudges

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

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

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

### 2.3 Nutrition Labels

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

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

### 2.4 Contribution

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

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

## 3 METHODOLOGY

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

### 3.1 Dataset

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

**Table 1: Allrecipes.com dataset used for algorithm training and the user study.**

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

### 3.2 Offline Evaluation

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

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

### 3.3 System Design and Procedure

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

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

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

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

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

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

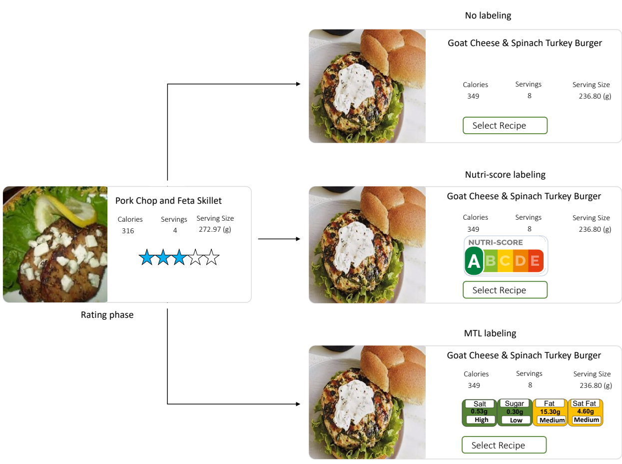



**Figure 1: Examples of how individual recipes were presented to the active user across different labeling conditions.**

## 3.4 Research Design

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

## 3.5 Study Participants

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

## 3.6 Measures

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

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

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

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

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

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

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

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

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

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

## 4 RESULTS

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

### 4.1 RQ1: Healthiness of Chosen Recipes

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

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

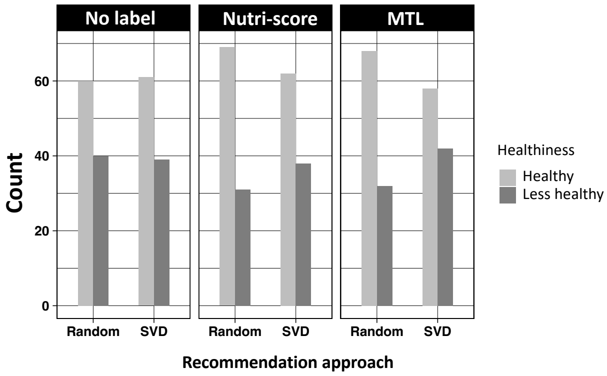



**Figure 2: Distribution of healthy and less healthy recipe choices across our different recommender and labelling conditions.**

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

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

### 4.2 RQ2: User Evaluation

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

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

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

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

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

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

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

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

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

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

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

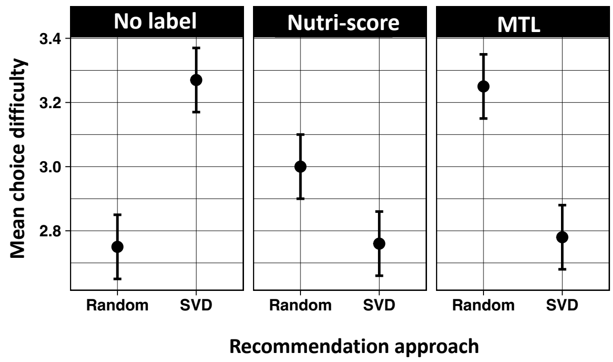



**Figure 3: Means and standard errors of choice difficulty levels, reported by users across conditions.**

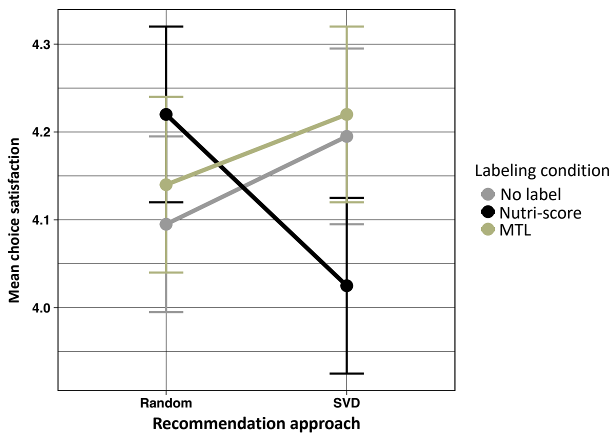



**Figure 4: Mean levels of choice satisfaction across all personalization (random vs SVD) and labeling conditions.**

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

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

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

### 4.3 RQ3: Predicting Recipe Healthiness

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

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

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

## 5 DISCUSSION

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

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

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

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

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

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

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

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

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

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

## ACKNOWLEDGMENTS

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

## REFERENCES

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

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

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

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

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

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

[7] Health Canada. 2014. *The Development and Use of a Surveillance Tool: The Classification of Foods in the Canadian Nutrient File According to Eating Well with Canada's Food Guide.* http://publications.gc.ca/collections/collection_2014/sc-hc/H164-158-2-2014-eng.pdf

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

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

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

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

[12] Michael D Ekstrand and Martijn C Willemsen. 2016. Behaviorism is not enough: better recommendations through listening to users. In *Proceedings of the 10th ACM conference on recommender systems*. ACM, New York, NY, USA, 221–224.

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

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

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

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

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

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

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

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

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

*(TMIS)* 10, 4 (2019), 1–23.
[22] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. 2011. An introduction to recommender systems. *New York: Cambridge* 10 (2011), 1941904.
[23] Mathias Jesse and Dietmar Jannach. 2021. Digital nudging with recommender systems: Survey and future directions. *Computers in Human Behavior Reports* 3 (2021), 100052.
[24] Mathias Jesse, Dietmar Jannach, and Bartosz Gula. 2021. Digital Nudging for Online Food Choices. *Frontiers in Psychology* 12 (2021).
[25] Eric J Johnson, Suzanne B Shu, Benedict GC Dellaert, Craig Fox, Daniel G Goldstein, Gerald Häubl, Richard P Larrick, John W Payne, Ellen Peters, David Schkade, et al. 2012. Beyond nudges: Tools of a choice architecture. *Marketing Letters* 23, 2 (2012), 487–504.
[26] Daniel Kahneman. 2011. *Thinking, fast and slow.* Macmillan.
[27] Florian G Kaiser, Oliver Arnold, and Siegmar Otto. 2014. Attitudes and defaults save lives and protect the environment jointly and compensatorily: Understanding the behavioral efficacy of nudges and other structural interventions. *Behavioral sciences* 4, 3 (2014), 202–212.
[28] Mansura A Khan, Ellen Rushe, Barry Smyth, and David Coyle. 2019. Personalized, health-aware recipe recommendation: an ensemble topic modeling based approach. In *HEALTHI'21: Joint Proceedings of ACM IUI 2021 Workshops.* CEUR, 4–10.
[29] Bart P Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. 2016. Recommender systems for self-actualization. In *Proceedings of the 10th acm conference on recommender systems.* ACM, New York, NY, USA, 11–14.
[30] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User modeling and user-adapted interaction* 22, 4 (2012), 441–504.
[31] Thomas C Leonard. 2008. Richard H. Thaler, Cass R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness.
[32] Christophe Leys, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology* 49, 4 (2013), 764–766.
[33] Jennifer Mankoff, Gary Hsieh, Ho Chak Hung, Sharon Lee, and Elizabeth Nitao. 2002. Using low-cost sensing to support nutritional awareness. In *International conference on ubiquitous computing.* Springer, 371–378.
[34] Makoto Matsumoto and Takuji Nishimura. 1998. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 8, 1 (1998), 3–30.
[35] Cliona Ni Mhurchu, Helen Eyles, Yannan Jiang, and Tony Blakely. 2018. Do nutrition labels influence healthier food choices? Analysis of label viewing behaviour and subsequent food purchases in a labelling intervention trial. *Appetite* 121 (2018), 360–365.
[36] Stefanie Mika. 2011. Challenges for nutrition recommender systems. In *Proceedings of the 2nd Workshop on Context Aware Intel. Assistance, Berlin, Germany.* CEUR, 25–33.
[37] Cataldo Musto, Christoph Trattner, Alain Starke, and Giovanni Semeraro. 2020. Towards a knowledge-aware food recommender system exploiting holistic user models. In *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization.* ACM, New York, NY, USA, 333–337.
[38] Pan American Health Organization. 2020/09/25. *Launching of Front-of-Package Labeling as a Policy Tool for the Prevention of Noncommunicable Diseases in the Americas.* PAHO. https://www.paho.org/en/documents/front-package-labeling-policy-tool-prevention-noncommunicable-diseases-americas
[39] World Health Organization. 2003. *Diet, nutrition, and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation.* Vol. 916. World Health Organization.
[40] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender systems: introduction and challenges. In *Recommender systems handbook.* Springer, 1–34.
[41] Osvaldo Santos, Violeta Alarcão, Rodrigo Feteira-Santos, João Fernandes, Ana Virgolino, Catarina Sena, Carlota Pacheco Vieira, Maria João Gregório, Paulo Nogueira, Pedro Graça, et al. 2020. Impact of different front-of-pack nutrition labels on online food choices. *Appetite* 154 (2020), 104795.
[42] Barthélemy Sarda, Corinne Delamaire, Anne-Juliette Serry, and Pauline Ducrot. 2022. Changes in home cooking and culinary practices among the French population during the COVID-19 lockdown. *Appetite* 168 (2022), 105743.
[43] Hanna Schäfer, Santiago Hors-Fraile, Raghav Pavan Karumur, André Calero Valdez, Alan Said, Helma Torkamaan, Tom Ulmer, and Christoph Trattner. 2017. Towards health (aware) recommender systems. In *Proceedings of the 2017 international conference on digital health.* ACM, New York, NY, USA, 157–161.
[44] Alain Starke. 2019. RecSys Challenges in achieving sustainable eating habits. In *4th International Workshop on Health Recommender Systems, HealthRecSys 2019.* CEUR, 29–30.
[45] Alain Starke, Edis Asotic, and Christoph Trattner. 2021. "Serving Each User": Supporting Different Eating Goals Through a Multi-List Recommender Interface. In *Fifteenth ACM Conference on Recommender Systems.* ACM, New York, NY, USA, 124–132.
[46] Alain Starke, Christoph Trattner, Hedda Bakken, Martin Johannessen, and Vegard Solberg. 2021. The cholesterol factor: Balancing accuracy and health in recipe recommendation through a nutrient-specific metric. In *Proceedings of the 1st Workshop on Multi-Objective Recommender Systems (MORS 2021).* CEUR, 11 pages.
[47] Alain Starke, Martijn Willemsen, and Chris Snijders. 2021. Promoting Energy-Efficient Behavior by Depicting Social Norms in a Recommender Interface. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 11, 3-4 (2021), 1–32.
[48] Alain D Starke and Christoph Trattner. 2021. Promoting Healthy Food Choices Online: A Case for Multi-List Recommender Systems. In *Proceedings of the ACM IUI 2021 Workshops.* CEUR, Aachen, DE, 3 pages.
[49] Alain D Starke, Martijn C Willemsen, and Christoph Trattner. 2021. Nudging healthy choices in food search through visual attractiveness. *Frontiers in Artificial Intelligence* 4 (2021), 20.
[50] Norman J Temple and Joy Fraser. 2014. Food labels: a critical assessment. *Nutrition* 30, 3 (2014), 257–260.
[51] Richard H Thaler and Cass R Sunstein. 2009. *Nudge: Improving decisions about health, wealth, and happiness.* Penguin.
[52] Raciel Yera Toledo, Ahmad A Alzahrani, and Luis Martinez. 2019. A food recommender system considering nutritional information and user preferences. *IEEE Access* 7 (2019), 96695–96711.
[53] Thi Ngoc Trang Tran, Müslüm Atas, Alexander Felfernig, and Martin Stettinger. 2018. An overview of recommender systems in the healthy food domain. *Journal of Intelligent Information Systems* 50, 3 (2018), 501–526.
[54] Thi Ngoc Trang Tran, Alexander Felfernig, Christoph Trattner, and Andreas Holzinger. 2021. Recommender systems in the healthcare domain: state-of-the-art and research issues. *Journal of Intelligent Information Systems* 57, 1 (2021), 171–201.
[55] Christoph Trattner and David Elsweiler. 2017. Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proceedings of the 26th international conference on world wide web.* ACM, New York, NY, USA, 489–498.
[56] Christoph Trattner and David Elsweiler. 2019. An evaluation of recommendation algorithms for online recipe portals. In *Proceedings of the 4th International Workshop on Health Recommender Systems.* CEUR, Aachen, DE.
[57] Christoph Trattner and David Elsweiler. 2019. Food Recommendations. In *Collaborative recommendations: Algorithms, practical challenges and applications.* World Scientific, 653–685.
[58] Christoph Trattner, Dominik Moesslang, and David Elsweiler. 2018. On the predictability of the popularity of online recipes. *EPJ Data Science* 7, 1 (2018), 1–39.
[59] Tsuguya Ueta, Masashi Iwakami, and Takayuki Ito. 2011. A recipe recommendation system based on automatic nutrition information extraction. In *International Conference on Knowledge Science, Engineering and Management.* Springer, 79–90.
[60] L Nynke van der Laan and Oliwia Orcholska. 2022. Effects of Digital Just-In-Time Nudges on Healthy Food Choice–a Field Experiment. *Food Quality and Preference* 98 (2022), 104535 pages.
[61] Ellen J Van Loo, Vincenzina Caputo, Rodolfo M Nayga Jr, and Wim Verbeke. 2014. Consumers' valuation of sustainability labels on meat. *Food Policy* 49 (2014), 137–150.
[62] Martijn C Willemsen, Mark P Graus, and Bart P Knijnenburg. 2016. Understanding the role of latent feature diversification on choice difficulty and satisfaction. *User Modeling and User-Adapted Interaction* 26, 4 (2016), 347–389.
[63] Martijn C Willemsen, Bart P Knijnenburg, Mark P Graus, LC Velter-Bremmers, and Kai Fu. 2011. Using latent features diversification to reduce choice difficulty in recommendation lists. *RecSys* 11, 2011 (2011), 14–20.
[64] Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. A dynamic recurrent model for next basket recommendation. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval.* 729–732.

# Manuscript III

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

*Alain Starke, Ayoub El Majjodi, Christoph Trattner. IntRS@ RecSys 2023.*

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

Alain Starke[1,2], Ayoub El Majjodi[2] and Christoph Trattner[2]

[1]Marketing and Consumer Behaviour Group, Wageningen University & Research, Wageningen, The Netherlands
[2]Department of Information Science and Media Studies, University of Bergen, Bergen, Norway

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

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

## 1. Introduction

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

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

*IntRS'22: Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, September 22, 2022, Seattle, US (hybrid event)*
alain.starke@wur.nl (A. Starke); ayoub.majjodiu@uib.no (A. E. Majjodi); christoph.trattner@uib.no (C. Trattner)
https://www.christophtrattner.info/ (C. Trattner)
0000-0002-9873-8016 (A. Starke); 0000-0002-7478-5811 (A. E. Majjodi); 0000-0002-1193-0508 (C. Trattner)

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

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

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

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

We argue that the extent to which users are interested in recipe healthiness can affect how they evaluate a recommender system and its preference elicitation method. On the one hand, in a collaborative filtering context, users can only indicate their interest in healthier recipes by rating specific recipes, while this might be easier in a recommender system that inquires on specific nutrition-related features, such as knowledge-based or constraint-based recommenders. On the other hand, users who aware of health and nutrition might be able to pick specific recipes that fit their needs and would experience feature-based elicitation as not fully meeting their needs. To this end, we consider a user's level of health consciousness, which measures one's perception of one's diet and the relation between nutrient intake and health [29]. This aspect is adapted from pre-validated scales, used in nutritional studies [30, 29].

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

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

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

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

## 2. Methodology

### 2.1. Dataset

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

**Table 1**
Allrecipes.com dataset used for algorithm training and the user study.

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

## 2.2. Recommender Approaches

To address our research questions, and specifically RQ2, we compared two recommender approaches that were distinct in terms of their preference elicitation (PE) methods[1]. Collaborative Filtering (CF) relies on rating-based PE, asking users to indicate preferences for individual items (i.e., recipes). Such approaches tend to outperform other item-based PE methods, such as content-based recommendation [31]. In contrast, Constraint-based (CB) recommendation exploits user preferences for recipe features, retrieving content based on the relation between user characteristics and recipe features. Both of the selected approaches involve explicit preference elicitation, as this was found to be the best representation of user preferences in food domain [8].

### 2.2.1. Collaborative Filtering (CF)

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

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

### 2.2.2. Constraint-based (CB)

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

## 2.3. Research Design and System Procedure

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

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

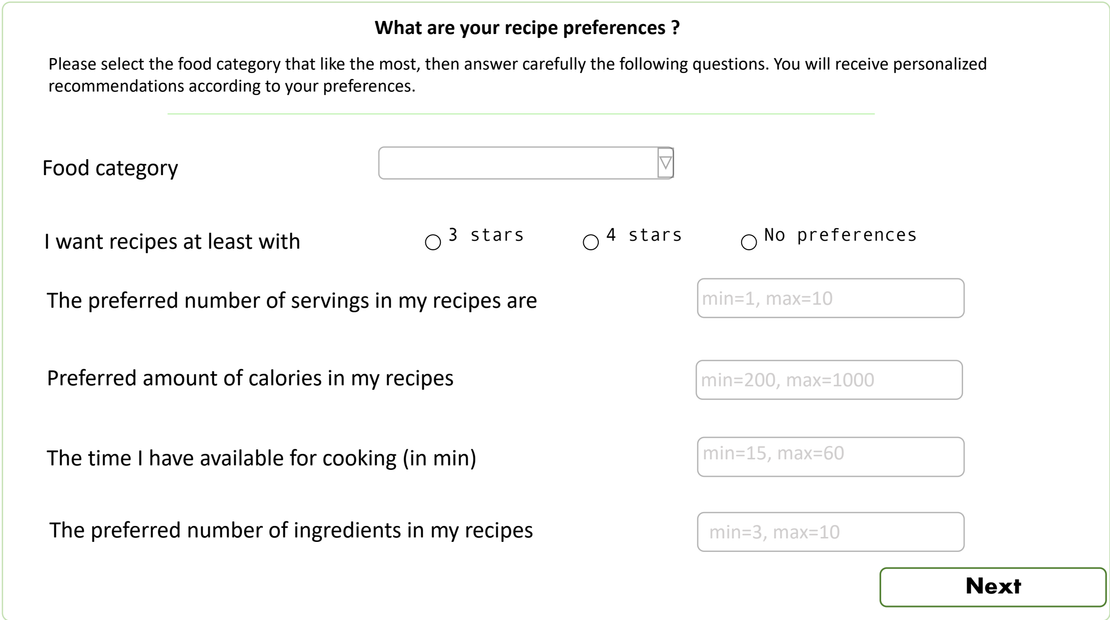



**Figure 1:** Features extracted for the Constraint-based preference elicitation and recommender approach.

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

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

## 2.4. Participants

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

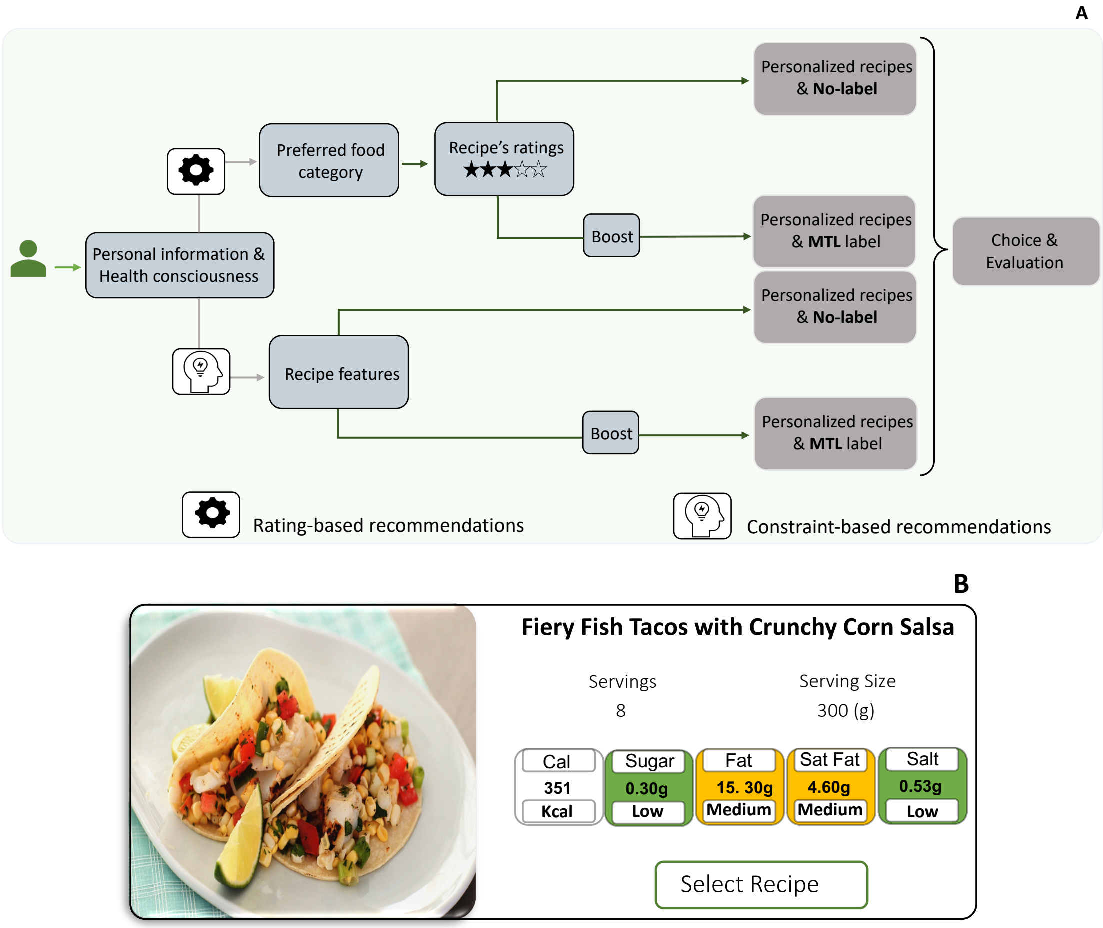



**Figure 2:** A (on the top): User flow of the online evaluation. B (on the bottom): An example of a recipe that was annotated with a Multiple Traffic Light label.

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

## 2.5. Measures

### 2.5.1. Recipe healthiness

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

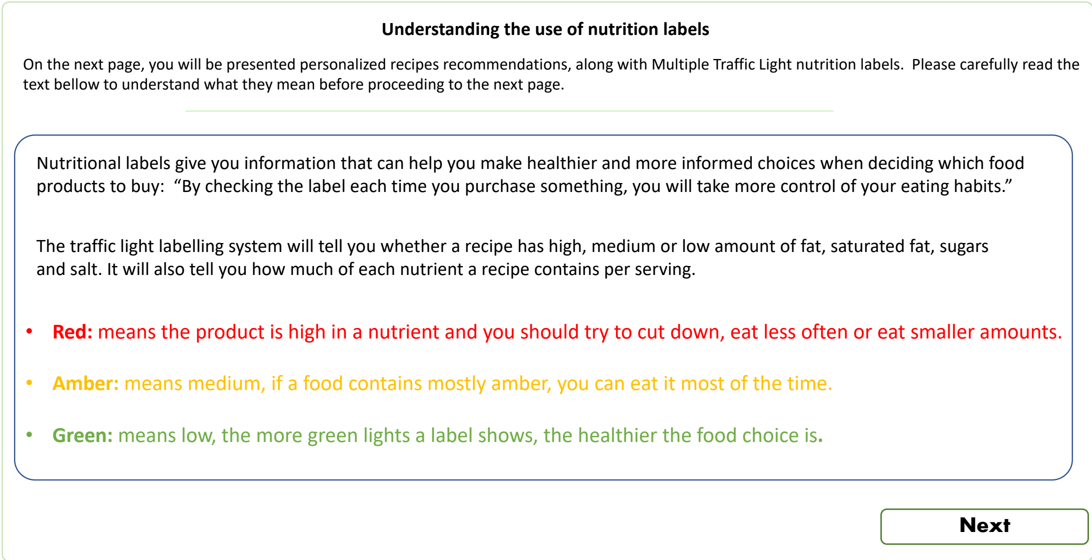



**Figure 3:** Prompt to Boost a user's understanding of the Multiple Traffic Light nutrition labels.

#### 2.5.2. Nutrition Labels

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

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

#### 2.5.3. User Characteristics, Perception, and Experience

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

Whether the items formed the expected aspects was examined using a principal component factor analysis. Table 2 outlines the results, describing the factor loadings for the used aspects and items. Items with low loadings or too many cross-loadings were removed from analysis. Whereas choice difficulty, choice satisfaction and perceived effort could be inferred reliably, there were doubts about the reliability of health consciousness. We observed a low value of Cronbach's Alpha (0.37), even after dropping unreliable items. Since the used items were part of a pre-validated scale and the factor loadings with the retained items were good, we decided to proceed with our analyses including health consciousness.

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

**Table 2**
Results of the principal component factor analysis across different user characteristics and experience aspects. Items were measured on 5-point Likert scales. Cronbach's Alpha is denoted by $\alpha$, items in gray were omitted from analysis.

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

# 3. Results

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

## 3.1. RQ1: Boosting Towards Healthier Choices

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

**Table 3**
Results of a two-way ANCOVA, predicting the healthiness of chosen recipes across different recommendation approaches and interventions conditions. A user's health consciousness was included as a covariate. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

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

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

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

#### 3.1.1. Conclusion

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

### 3.2. RQ2: User Evaluation of Preference Elicitation Methods

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

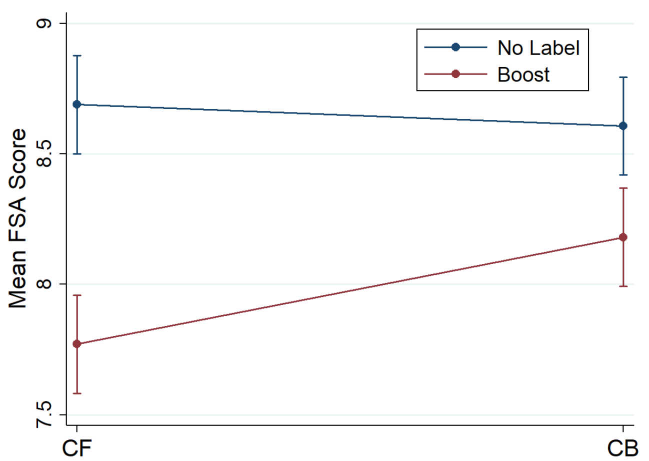



**Figure 4:** Mean FSA score of chosen recipes across different conditions for preference elicitation (CF vs CB) and labelling (No Label vs Boost). Lower scores indicate healthier recipe choices.

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

### 3.2.1. Perceived Effort

We performed a one-way ANCOVA on a user's perceived effort of using our recommender system, including an interaction effect between health consciousness and the used PE method[2]. The results are outlined in Table 4. We found no main effects for the used elicitation method, as the perceived effort of using the CB recommender ($M = -.014$, $SD = 1.03$), which relied on disclosing preferences for recipe features, was only somewhat lower than that of the CF-based recommender ($M = .014$, $SD = .97$), which relied on rating-based PE. In addition, we neither observed a main effect of a user's health consciousness on effort: $F(1, 240) = 1.16$, $p = 0.28$.

**Table 4**
Results of a one-way ANCOVA with interaction effect on the user's level of perceived effort. It was examined across different PE conditions (CF, CB), while controlling for a user's level of health consciousness. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

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

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

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

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

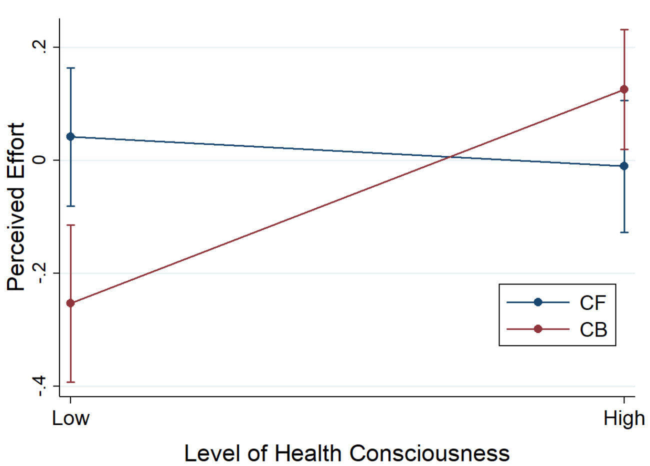



**Figure 5:** Users' perceived effort of using the different Preference Elicitation (PE) methods, based on a user's level of health consciousness. The Collaborative Filtering (CF) PE was rating-based, while the Constraint-Based (CB) PE was feature-based. To interpret the interaction effect in Table 4,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

### 3.2.2. Choice Difficulty and Choice Satisfaction

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

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

With regard to other aspects, we found that users who perceived a recommender as effortful to use, also reported higher levels of choice difficulty: $\beta = .26$, $p < 0.001$. This suggested a possible indirect effect of the interplay between a user's level of health consciousness and the PE method on choice difficulty, via perceived effort. Hence, an earlier analysis revealed that users in the CB condition reported higher levels of perceived effort if they had a higher level of health consciousness, and vice versa. In contrast, the healthiness of the chosen recipe (i.e., FSA) was not related to choice difficulty.

**Table 5**
Linear regression analyses, with models to predict the user's experienced choice difficulty and choice satisfaction, based on the experimental conditions, user characteristics, perception aspects and choices. 'Boost' and 'CB' were coded as 0.5, 'No label' and 'CF' as -0.5. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

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

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

### 3.2.3. Conclusion

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

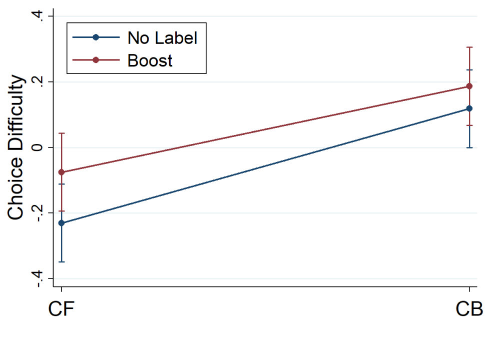



**Figure 6:** Standardized scores for the choice difficulty experience aspect across conditions.
Errors bars represent 1 S.E.

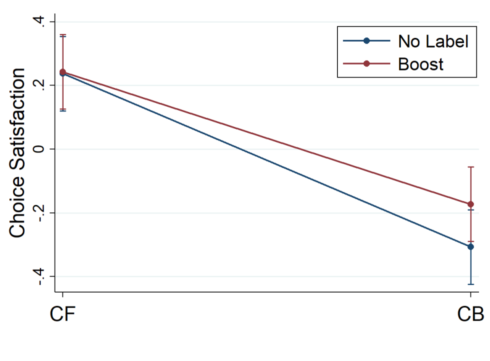



**Figure 7:** Standardized scores for the choice satisfaction experience aspect across conditions.
Errors bars represent 1 S.E.

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

## 4. Discussion

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

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

Second, we have found that the user's evaluation of a food recommender system is significantly associated with the used preference elicitation method, based on that user's level of health consciousness (RQ2). We have compared two distinct recommender approaches, collaborative filtering and constraint-based, that also involve different methods of preference. Across all types of users, we find that constraint-based recipe recommendation lists are more difficult to choose from and, in turn, lead to lower levels of satisfaction. With regard to choice satisfaction, it could be argued that this outcome is unsurprising, since a constraint-based recommender system tends to be less effective than a CF-based recommender as it makes much simpler assumptions about the user's preferences [51].

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

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

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

A limitation regarding the recommended items is the arguably smaller dataset of recipes. While some other studies with internet-sourced recipes have been able to leverage datasets with more recipes [11], our dataset is smaller due to the focus on high-quality ratings for our collaborative filtering recommender. We have only used ratings from users that have rated at least 20 recipes, to make sure that the collected ratings are of rather active and experienced users. In addition,

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

## Acknowledgments

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

## References

[1] W. Hill, L. Stead, M. Rosenstein, G. Furnas, Recommending and evaluating choices in a virtual community of use, in: Proceedings of the SIGCHI conference on Human factors in computing systems, 1995, pp. 194–201.

[2] J. Konstan, J. Riedl, Recommended for you, IEEE Spectrum 49 (2012) 54–61. doi:10.1109/MSPEC.2012.6309257.

[3] M. D. Ekstrand, M. C. Willemsen, Behaviorism is not enough: better recommendations through listening to users, in: Proceedings of the 10th ACM conference on recommender systems, ACM, New York, NY, USA, 2016, pp. 221–224.

[4] B. Knijnenburg, M. Willemsen, R. Broeders, Smart sustainability through system satisfaction: tailored preference elicitation for energy-saving recommenders, in: 20th Americas Conference on Information Systems (AMCIS 2014), August 7-9, 2014, Savannah, Georgia, United States, AIS/ICIS, 2014, pp. 1–15.

[5] H. Schäfer, M. C. Willemsen, Rasch-based tailored goals for nutrition assistance systems, in: Proceedings of the 24th International Conference on Intelligent User Interfaces, 2019, pp. 18–29.

[6] A. Starke, M. Willemsen, C. Snijders, Effective user interface designs to increase energy-efficient behavior in a rasch-based energy recommender system, in: Proceedings of the eleventh ACM conference on recommender systems, 2017, pp. 65–73.

[7] C. Trattner, D. Elsweiler, Food recommendations, in: Collaborative recommendations: Algorithms, practical challenges and applications, World Scientific, 2019, pp. 653–685.

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

[9] D. Elsweiler, H. Hauptmann, C. Trattner, Food recommender systems, in: F. Ricci, L. Rokach, B. Shapira (Eds.), Recommender Systems Handbook, Springer US, New York, NY, 2022, pp. 871–925. URL: https://doi.org/10.1007/978-1-0716-2197-4_23. doi:10.1007/978-1-0716-2197-4_23.
[10] S. Mika, Challenges for nutrition recommender systems, in: Proceedings of the 2nd Workshop on Context Aware Intel. Assistance, Berlin, Germany, CEUR, 2011, pp. 25–33.
[11] C. Trattner, D. Elsweiler, Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems, in: Proceedings of the 26th international conference on world wide web, ACM, New York, NY, USA, 2017, pp. 489–498.
[12] C. Musto, C. Trattner, A. Starke, G. Semeraro, Towards a knowledge-aware food recommender system exploiting holistic user models, in: Proceedings of the 28th ACM conference on user modeling, adaptation and personalization, ACM, New York, NY, USA, 2020, pp. 333–337.
[13] C. Musto, A. D. Starke, C. Trattner, A. Rapp, G. Semeraro, Exploring the effects of natural language justifications in food recommender systems, in: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, 2021, pp. 147–157.
[14] A. Starke, E. Asotic, C. Trattner, "serving each user": Supporting different eating goals through a multi-list recommender interface, in: Fifteenth ACM Conference on Recommender Systems, ACM, New York, NY, USA, 2021, pp. 124–132.
[15] E. J. Johnson, S. B. Shu, B. G. Dellaert, C. Fox, D. G. Goldstein, G. Häubl, R. P. Larrick, J. W. Payne, E. Peters, D. Schkade, et al., Beyond nudges: Tools of a choice architecture, Marketing Letters 23 (2012) 487–504.
[16] R. H. Thaler, C. R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness, Penguin, 2009.
[17] R. Hertwig, T. Grüne-Yanoff, Nudging and boosting: Steering or empowering good decisions, Perspectives on Psychological Science 12 (2017) 973–986.
[18] G. Loewenstein, C. Bryce, D. Hagmann, S. Rajpal, Warning: You are about to be nudged, Behavioral Science & Policy 1 (2015) 35–42.
[19] L. M. König, B. Renner, Boosting healthy food choices by meal colour variety: results from two experiments and a just-in-time ecological momentary intervention, BMC Public Health 19 (2019) 1–15.
[20] T. Rouyard, B. Engelen, A. Papanikitas, R. Nakamura, Boosting healthier choices, bmj 376 (2022).
[21] M. Egnell, Z. Talati, S. Hercberg, S. Pettigrew, C. Julia, Objective understanding of front-of-package nutrition labels: an international comparative experimental study across 12 countries, Nutrients 10 (2018) 1542.
[22] E. J. Van Loo, V. Caputo, R. M. Nayga Jr, W. Verbeke, Consumers' valuation of sustainability labels on meat, Food Policy 49 (2014) 137–150.
[23] N. J. Temple, J. Fraser, Food labels: a critical assessment, Nutrition 30 (2014) 257–260.
[24] P. Ducrot, C. Julia, C. Méjean, E. Kesse-Guyot, M. Touvier, L. K. Fezeu, S. Hercberg, S. Péneau, Impact of different front-of-pack nutrition labels on consumer purchasing intentions: a randomized controlled trial, American journal of preventive medicine 50 (2016) 627–636.
[25] D. Jannach, M. Zanker, A. Felfernig, G. Friedrich, An introduction to recommender systems,

New York: Cambridge 10 (2011) 1941904.
[26] F. Ricci, L. Rokach, B. Shapira, Recommender systems: introduction and challenges, in: Recommender systems handbook, Springer, 2015, pp. 1–34.
[27] C. C. Aggarwal, Knowledge-based recommender systems, in: Recommender systems, Springer, 2016, pp. 167–197.
[28] B. P. Knijnenburg, M. C. Willemsen, Understanding the effect of adaptive preference elicitation methods on user satisfaction of a recommender system, in: Proceedings of the third ACM conference on Recommender systems, 2009, pp. 381–384.
[29] S. J. Gould, Health consciousness and health behavior: the application of a new health consciousness scale, American Journal of Preventive Medicine 6 (1990) 228–237.
[30] A. Gámbaro, A. C. Ellis, V. Prieto, Influence of subjective knowledge, objective knowledge and health consciousness on olive oil consumption—a case study, Food and Nutrition 4 (2013) 445–453.
[31] C. Trattner, D. Elsweiler, An evaluation of recommendation algorithms for online recipe portals, in: Proceedings of the 4th International Workshop on Health Recommender Systems, CEUR, Aachen, DE, 2019.
[32] A. El Majjodi, A. D. Starke, C. Trattner, Nudging towards health? examining the merits of nutrition labels and personalization in a recipe recommender system, in: Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, 2022, pp. 48–56.
[33] S. Funk, Netflix update: Try this at home (december 2006), 1999. URL: http://sifter.org/simon/journal/20061211.html.
[34] W. H. Organization, Diet, nutrition, and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation, volume 916, World Health Organization, 2003.
[35] H. Canada., The development and use of a surveillance tool: The classification of foods in the canadian nutrient file according to eating well with canada's food guide, 2014. URL: http://publications.gc.ca/collections/collection_2014/sc-hc/H164-158-2-2014-eng.pdf.
[36] A. D. Starke, M. C. Willemsen, C. Trattner, Nudging healthy choices in food search through visual attractiveness, Frontiers in Artificial Intelligence 4 (2021) 20.
[37] Department of Health and Social Care UK, Front of Pack nutrition labelling guidance, 2016. URL: https://www.gov.uk/government/publications/front-of-pack-nutrition-labelling-guidance.
[38] K. Anastasiou, M. Miller, K. Dickinson, The relationship between food label use and dietary intake in adults: A systematic review, Appetite 138 (2019) 280–291.
[39] M. L. Neuhouser, A. R. Kristal, R. E. Patterson, Use of food nutrition labels is associated with lower fat intake, Journal of the American dietetic Association 99 (1999) 45–53.
[40] M. Cecchini, L. Warin, Impact of food labelling systems on food choices and eating behaviours: a systematic review and meta-analysis of randomized studies, Obesity reviews 17 (2016) 201–210.
[41] K. G. Grunert, J. M. Wills, L. Fernández-Celemín, Nutrition knowledge, and use and understanding of nutrition information on food labels among consumers in the uk, Appetite 55 (2010) 177–189.
[42] K. L. Hawley, C. A. Roberto, M. A. Bragg, P. J. Liu, M. B. Schwartz, K. D. Brownell, The science on front-of-package food labels, Public health nutrition 16 (2013) 430–439.

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

# Manuscript IV

## The Interplay between Food Knowledge, Nudges, and Preference Elicitation Methods Determines the Evaluation of a Recipe Recommender System.

*Ayoub El Majjodi, Alain D. Starke, Mehdi Elahi, Christoph Trattner. IntRS@ RecSys 2023.*

# The Interplay between Food Knowledge, Nudges, and Preference Elicitation Methods Determines the Evaluation of a Recipe Recommender System

Ayoub El Majjodi[1], Alain D. Starke[2,1], Mehdi Elahi[1] and Christoph Trattner[1]

[1]*MediaFutures, University of Bergen, 5007 Bergen, Norway*
[2]*ASCoR, University of Amsterdam, Nieuwe Achtergracht 166, 1001 NG Amsterdam, Netherlands*

Abstract
Domain knowledge can affect how a user evaluates different aspects of a recommender system. Recipe recommendations might be difficult to understand, as some health aspects are implicit. The appropriateness of a recommender's preference elicitation (PE) method, whether users rate individual items or item attributes, may depend on the user's knowledge level. We present an online recipe recommender experiment. Users ($N = 360$) with varying levels of subjective food knowledge faced different cognitive digital nudges (i.e., food labels) and PE methods. In a 3 (recipes annotated with no labels, Multiple Traffic Light (MTL) labels, or full nutrition labels) x 2 (PE method: content-based PE or knowledge-based) between-subjects design. We observed a main effect of knowledge-based PE on the healthiness of chosen recipes, while MTL label only helped marginally. A Structural Equation Model analysis revealed that the interplay between user knowledge and the PE method reduced the perceived effort of using the system and in turn, affected choice difficulty and satisfaction. Moreover, the evaluation of health labels depends on a user's level of food knowledge. Our findings emphasize the importance of user characteristics in the evaluation of food recommenders and the merit of interface and interaction aspects.

**Keywords**
Recommender Systems, Food, Digital nudges, Nutrition labels, Preference Elicitation

## 1. Introduction

An increasing number of food decisions are made digitally [1]. In addition to online grocery stores, recipe websites play a pivotal role in supporting home cooking [2]. At the same time, users of such websites may find it difficult to navigate a large number of recipes. Some recipes are more challenging to understand or cook than others [3], where users may lack sufficient food knowledge or skills to engage with them [4]. For example, while some users have clear preferences regarding specific recipe features, such as cooking time and the number of ingredients, others may make food choices based on past positive experiences with a recipe and seek out 'more like this' [5].

The large availability of recipes requires the use of information-filtering systems. Food recommender systems have played an instrumental role in the popularity of recipe website [6],

*IntRS '23: 10th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems at the 17th ACM Conference on Recommender Systems (RecSys '23), September 18–22, 2023, Singapore*
ayoub.majjodi@uib.no (A. E. Majjodi); a.d.starkeg@uva.nl (A. D. Starke); mehdi.elahi@uib.no (M. Elahi); christoph.trattner@uib.no (C. Trattner)

largely focusing on predictive accuracy [7, 2]. However, eating is a particularly challenging domain to achieve robust improvements in accuracy when aiming to go beyond popularity-based approaches. Food preferences strongly depend on various factors, such as context [8], group size [9], time of the day, and the day of the week [2].

Improvements may therefore be sought in other recommender aspects. User preferences are also formed by how information is presented, combined with the user's level of understanding [10, 9]. While users may have a specific eating goal (e.g., health or sustainability [5]), the information presented by recommenders may be too limited to make an informed decision [11]. To this end, summary labels, as also used on supermarket products, may help to grasp the healthiness or nutritional content of a recipe, particularly for novice users.

The user's knowledge may also affect how a recommender's preference elicitation method is evaluated. In the energy conservation domain, a domain subject to behavioral change [12], a user's evaluation of a preference elicitation (PE) method seems to depend on the user's level of domain knowledge [10, 13]. Users that do not know much about energy conservation tend to be more satisfied when using a case-based PE method, in which individual items are (dis)liked. In contrast, experienced users are found to be more satisfied when interacting with attributes of energy-saving measures, such as effort and investment costs.

We argue that a similar dynamic applies to the food recommender domain. We expect novice users of recipe recommenders and websites to incrementally explore recipes based on what they liked previously, as one would in a content-based method [7]. In contrast, experienced users would seek out new recipes more easily based on preferred attributes, as one would through a knowledge-based method [14].

This paper examines the role of user knowledge on decisions in a recipe recommender system. Instead of following an algorithmic optimization approach that only focuses on user preferences [7], we examine the influence of two other recommender aspects: food labels and preference elicitation methods. A user's domain knowledge is critical here, regarding the most optimal interface representation or interaction method.

First, we apply digital, informational nudges to recommendation lists in the form of nutrition and health labels. In the food domain, these have been used primarily in supermarkets to communicate nutrition information to consumers in a simplified manner [15, 16]. Most notable are summary labels, such as the Nutriscore to classify foods between A and E. Most popular is the Multiple Traffic Light (MTL) label [17], which uses colors to indicate (un)healthy intakes of four nutrients: fat, saturated fat, sugar, and salt. This study considers two labels, with varying degrees of difficulty: The MTL label and the back-of-pack nutritional facts [17].

Second, we examine the role of two different PE methods. First, we develop a content-based recommender in which a user picks a favorite recipe from a randomly generated list of recipes. Second, we develop a knowledge-based recommender system in which users indicate their preferences for recipes based on a set of attributes, such as cooking time and difficulty. We expect that user choices might be affected by the use of labels, and the evaluation depends on the interplay between user knowledge levels and either the labels or PE methods used. We formulate the following research questions:

- **RQ1**: To what extent do different nutrition labels support healthier recipe choices?

- **RQ2**: Does the user evaluation of a recipe recommender system depend on the interplay between food knowledge and different preference elicitation methods?

## 2. Related Work

### 2.1. Food Recommender Systems

The field of recommender systems has received considerable research attention due to the complex and fundamental nature of food [6, 2]. However, most studies have focused on improving prediction accuracy [9]. To this end, various techniques have been explored. While content-based approaches are initially found to outperform other methods (e.g., collaborative, knowledge-based) [7], collaborative filtering has also yielded better results in other studies [2].

A recurring problem is to support healthier food and recipe choices. There is an apparent tradeoff between 'user preferences' and health in recipe recommendation [6], particularly due to the popularity of unhealthy recipes [1]. Rather than restricting content based on health, various studies have examined hybrid solutions (e.g., through post ranking on health [18]) and interface solutions. For example, multi-list interfaces have been developed to support healthier eating goals, where multi-list recommenders are evaluated more favorably than single-list interfaces [5].

Nonetheless, very few studies have explored the impact of preference elicitation methods on user evaluation, particularly not in relation to a user's knowledge level. Research that presents novel preference elicitation (PE) methods typically involve conversational recipe recommenders [19]. For example, two studies have compared the effect of using different modalities on user choice or evaluation [20, 21]. While these modalities might affect how a user experiences an interaction or even what item is chosen, these modality types seem to bear no relation to a user's knowledge level. Thus, the current study aims to fill this gap by empirically examining the effects of different preference elicitation methods and subjective food knowledge on the user experience.

### 2.2. Adaptive Preference Elicitation Methods

Across all recommender domains, several preference elicitation (PE) approaches have been proposed and evaluated [2]. One of the earlier works in the food domain was [7], following a rather simplistic procedure where recipe ratings are obtained from users and transferred into ingredient ratings. While the user could still interact with individual recipes (as also in [22]), the system then aggregated the ratings of the ingredients to generate rating predictions.

More extensive approaches would include user preferences for individual recipes and attributes. Elahi et al. [23] consider additional factors when building a recommendation model, by including user food preferences, nutritional indicators, and ingredient costs. This results in a model that combines the predicted value of a recipe along with the above-noted factors to generate recommendations.

Food recommender studies do not explicitly discern between PE methods. In the end, most methods are simply a requirement for following a specific recommender model. Knowledge-based food recommender studies elicit extensive recipe attribute preferences [14, 24], while content-based and collaborative approaches can deal with interactions at the individual recipe level. To date, besides a preliminary study [22], it has not been considered that specific users, based on their knowledge or capabilities, may prefer specific PE methods or require specific information to be presented in a food recommender interface.

The interplay between user characteristics and PE methods is investigated in the energy conservation domain. Knijnenburg and Willemsen [10] compare two preference elicitation methods for a household energy recommender system. Users of a case-based PE method could indicate to (dis)like individual energy-saving measures (e.g., 'turn off the lights after leaving a room'). In contrast, users of an attribute-based PE method could either decrease or increase the weights of different energy-saving attributes, such as 'investment costs' or 'effort'. They find, also in follow-up studies [13], that users with high domain knowledge are more satisfied when using an attribute-based PE method, while users with lower domain knowledge prefer case-based PE. As food PE methods can also be differentiated in terms of interacting with individual recipes (i.e., case-based, such as in content-based recommendation) and recipe features (i.e., feature-based, such as in knowledge-based recommendation), we expect to observe an interaction effect between food knowledge and the PE method on how a recommender is evaluated.

## 2.3. Nutrition Labels and Digital Nudging

Nutritional information about food and recipes might not always be apparent to users. This is another area where the user's domain knowledge may affect their preferences, particularly in cases where it is either emphasized or not.

One way to communicate the healthiness of foods and recipes is through nutrition labels [17]. One category is back-of-package labels that outline the nutritional contents of a food product in detail [25]. Another category is front-of-package (FoP) labels that are used for both products and recipes, which typically summarize the product's or recipe's nutritional content. For example, this could be by highlighting specific nutrients of aggregating nutrition information towards a score [17, 25].

The health benefits of such FoP labels have been shown longitudinally, supporting healthy food intake [26]. While a growing body of evidence supports the effectiveness of FoP nutrition labels in promoting healthy food choices in physical settings, the impact of nutrition labels in digital contexts has been relatively understudied [16]. In particular, little is known about the effectiveness of FoP labels in personalized environments [27].

FoP labels can be regarded as a digital nudge, a change in an interface that leads to predictable choices [28]. More specifically, such a label is a cognitively oriented healthy eating nudge [15], as users are encouraged to re-consider their preferences and choices based on deliberation. Some of these labels, such as the Multiple Traffic Light (MTL) label, is also accompanied by a coloring system, which supports intuitive decision-making [29].

Various digital nudges mainly relate to interface aspects of food recommenders. For example, visual and textual explanations have been shown to shift user preferences towards healthier recipes [30, 24]. This study mainly builds upon earlier work where 'boosting' is examined to first explain FoP labels to users, after which recipes are annotated with them [22]. This has led to a higher proportion of healthier choices in the recommender interface. Instead of applying 'boosts' in this study, we seek to examine the effectiveness of two cognitive, informational nudges. To examine the differentiating effect of user knowledge when combined with nudges, we either annotate recipes with back-of-package labels (i.e., Nutritional Facts label[1]) or Multiple Traffic Light (MTL) labels [31]. We expect that users with higher levels of domain knowledge are better able to understand the full back-of-package label, compared to the MTL label also being appropriate for low-knowledge users.

### 2.4. Objectives

We extend previous work of food recommender systems [6], by examining the role of preference elicitation methods and digital nudges (through nutrition labels) [28, 10, 13]. First, we investigate how different labeling systems can facilitate healthier decision-making when selecting recipes (RQ1). We compare two label-based scenarios (with either an MTL label or a 'full' back-of-pack label) with a no-label baseline, focusing on the interplay between labels and the preference elicitation method and how this affects the user's evaluation. In doing so, we also consider a user's knowledge level and the preference elicitation method.

Second, we examine the impact of the interplay of user knowledge and preference elicitation methods on user choice and evaluation. We differentiate between two methods, content-based and knowledge-based, being 'case-based' or 'attribute-based' PE methods, respectively; in line with Knijnenburg et al. [13]. For the content-based approach, users are asked to indicate whether they like individual recipes, while the knowledge-based approach elicits preferences based on recipe features and personal characteristics, such as cooking time and self-reported weight goals. In line with Knijnenburg and Willemsen [32], we used Structural Equations Modeling (SEM) to construct a path model, in which changes to the recommender were related to perception aspects and, in turn, experience aspects.

## 3. Study Design

### 3.1. Dataset

To address our research questions, we used a dataset from the popular recipe website Allrecipes.com. From the larger corpus of 58,000 recipes, we sampled 5,000 recipes from different food categories for the main dish. In addition to the recipe title, all nutrients required to build the recommender were extracted.

### 3.2. Recommender Approaches

We employed two recommendation approaches, both of which rely on explicit preference elicitation methods [9].

#### 3.2.1. Content-based (CB)

The content-based recommender system generated recommendations based on similarity with recipes liked by the user. It employed the Term Frequency-Inverse Document Frequency (TF-

[1]https://www.fda.gov/food/new-nutrition-facts-label/how-understand-and-use-nutrition-facts-label

IDF) model to generate personalized recommendations based on the recipe's ingredients. The ingredient list was vectorized, operationalizing TF as the weight (per 100g) present in the recipe. To build a user model in our study, we presented a list of 10 recipes to the user that included detailed descriptions of their ingredients, pictures, servings, and calorie information. For computing the final recommendations for the user, we computed similarities between the user and item profile, employing a cosine similarity metric and the ingredient vectors (cf. [14]). This method adhered to the standard methods in food recommender systems and had been shown to generate decent results in the domain of food recommendations where no collaborative filtering is possible [33].

#### 3.2.2. Knowledge-based (KB)

We developed a knowledge-aware recommender system that extended the work of Musto et al. [14]. An overview of elicited features is described in Table 1. Users were asked to disclose personal characteristics and practical and health-related preferences related to recipes. A score-based ranker used encoded knowledge relations between user factors and recipe features to recipes, based on the user's profile. The scores are adjusted based on metadata such as ingredients or nutritional value. Table 1 also presents the rules to score recipes based on the user needs, which is among others based on information found in [34, 35, 36, 24].

**Table 1**
User Factors and rules to score and generate relevant recipes. The recipes were subject to three groups (higher, normal, and lower value) based on the average values of nutritional and practical aspects.

| User factor | Ranges | Recipe attributes affected |
|---|---|---|
| BMI | < 18 | higher calories, higher protein, higher carbohydrates |
| | 18-24 | normal calories |
| | > 24 | lower calories, lower fat |
| Eating Goals | lose weight | higher calories, higher protein, higher carbohydrate |
| | neutral | normal calories |
| | gain weight | lower calories, lower fat |
| Behaviours | > 30min/day activity | higher protein, higher calories |
| Sleep | < 7h hours | higher magnesium, higher vitamin B6, lower fat, higher protein |
| Depressed feelings | yes/no | higher protein, lower carbohydrates |
| Cooking experience | high, medium, low | number of instructions, number of ingredients |
| Cooking time | 30 min, 40-60 min, > 60 min | preparation time, number of ingredients |

### 3.3. Research Design and System Procedure

The participants were assigned to a between-subjects design with a 2 (Preference Elicitation (PE): Content-Based (CB) vs. Knowledge-Based (KB)) x 3 (labeling systems: no label vs. Multiple Traffic Light (MTL) vs. Full label) configuration. In the content-based method, users selected their preferred recipe from randomly generated options, while the knowledge-based condition involved users providing health and food-related information. Personalized recipes were then labeled with: No label, MTL label, or Full labels, as depicted in Figure 1 (A-C).

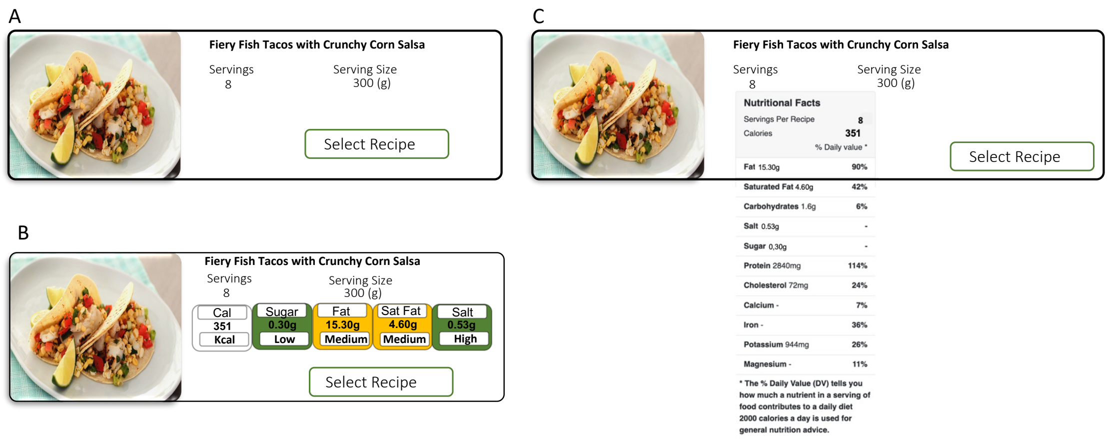



**Figure 1:** An example of personalized recipes annotated with No label (A), Multiple Traffic Light label (B), or Full label (C).

We implemented a user flow that started with obtaining consent from the study participants. The whole online user study flow is illustrated in Figure 2. Once the participants agreed to take part in the study, they provided basic demographic information, including age group, education level, and gender. Information processing was in line with Ethical guidelines at University of Bergen, Norway. In both the content and knowledge-based conditions, the choice task and evaluation questionnaire were similar, with users choosing a single preferred recipe from the list of recommended items (top-10 list). In both conditions, recipes were labeled either with an MTL label, a Full label or no label. Finally, the participants evaluated the system based on the performed choice regarding satisfaction, difficulty, and effort.

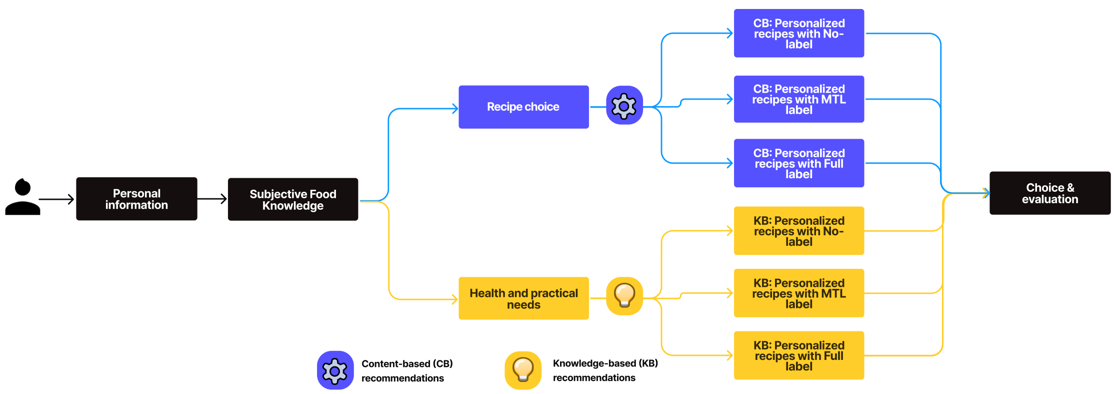



**Figure 2:** The User flow for the online experiment.

## 3.4. Participants

We utilized the Prolific crowdsourcing platform to recruit users for our study, offering 0.80 GBP as compensation. A total of 360 participants took part in the study. However, after pre-screening the data, we had to exclude 54 participants. Grounds for exclusion were based on multiple violations of non-attentiveness: not disclosing realistic knowledge-based criteria (e.g., a weight of 15kg), providing uniform responses in the user evaluation questionnaires, and/or completing the study in under 2 minutes. Our final analysis was performed on a sample of 306 (65% female) participants split equally on study conditions, with an average age of 30.5 years.

#### 3.4.1. Ethical Statement

This research adhered to the ethical guidelines of the University of Bergen and the Norwegian guidelines for scientific research. It was judged to pass without further extensive review, for it contained no misleading information, stress tasks, nor would it elicit extreme emotions.

### 3.5. Measures

#### 3.5.1. Recipe Healthiness

To determine the healthiness of the recipes in our dataset, we utilized the FSA score. It was introduced by the UK Food Standards Agency [31] and is considered a reliable measure to estimate recipes' healthiness. It was successfully used in multiple human-computer interaction and recommender systems studies on recipes [1, 22, 20, 18]. The metric represented an inverse healthiness score and ranged from 4 (very healthy) to 12 (not healthy). It was based on the levels of fat, saturated fat, sugar, and salt per 100g in a recipe, adhering to nutritional intake guidelines. For our algorithmic sampling, we classified recipes as healthy up to a score of 8, while higher scores were designated as unhealthy.

#### 3.5.2. Food knowledge and user evaluation

To measure the users' nutritional knowledge levels, we employed the Subjective Food Knowledge (SFD) questionnaire, which was validated in prior studies [37, 38]. The SFD questionnaire comprised five items, that are rated on a five-point Likert scale.

A user's experience of using our system was assessed through the recommender system evaluation framework Knijnenburg and Willemsen [32]. For this study, we expected changes in terms of how effortful a user perceived the interaction to be, while also inquiring on two different choice outcomes: choice difficulty and choice satisfaction. These two experience aspects are commonly used to evaluate recommender interactions [32, 13]. All questionnaire items b used for the user evaluation were previously validated in relevant domains through earlier studies: for perceived effort [39], choice difficulty [40, 5], and choice satisfaction [12, 5].

All items were submitted to a confirmatory factor analysis. Subjective food knowledge was analyzed separately to allow for interaction effects, while the other aspects were inferred as part of a structural equation model analysis. Table 2 describes the factor loadings and Cronbach's Alpha, showing that items with low loadings (indicated in grey) were excluded from the analysis. All aspects adhered to internal consistency guidelines ($\alpha > .70$), while they also met guidelines for convergent validity based on the average variance explained ($AVE > 0.5$).

**Table 2**
Results of the confirmatory factor analysis across different user characteristics and experience aspects. Items were measured on 5-point Likert scales. To allow for interaction effects in the path model, subjective food knowledge was analyzed separately. Cronbach's Alpha is denoted by $\alpha$, items in gray were omitted from analysis, due to low factor loading .

| Aspect | Item | Loading |
|---|---|---|
| Subj. Food Knowledge $AVE = .744$ $\alpha = .872$ | Compared with an average person, I know a lot about healthy eating. | .836 |
| | I think, I know enough about healthy eating to feel pretty confident when choosing a recipe. | .879 |
| | I know a lot about food to evaluate the healthiness of a recipe. | .847 |
| | I do not feel very knowledgeable about healthy eating. | -.887 |
| Choice Satisfaction $AVE = .616$ $\alpha = .795$ | I like the recipe I have chosen. | .750 |
| | I think I will prepare the recipe I have chosen. | .637 |
| | The chosen recipe fits my preference. | .746 |
| | I know many recipes that I like more than the one I have chosen. | |
| | I would recommend the chosen recipe to others. | .681 |
| Choice Difficulty $AVE = .554$ $\alpha = .659$ | I changed my mind several times before making a decision. | |
| | Making a choice was overwhelming. | .766 |
| | It was easy to make this choice. | -.549 |
| Perceived Effort $AVE = .504$ $\alpha = .632$ | The system takes up a lot of time. | .851 |
| | I quickly understood the functionalities of the system. | -.683 |
| | Many actions were required to use the system. | .557 |

# 4. Results

We examined our research questions through two different analyses. First, we examined the healthiness of user choices (RQ1) through a two-way ANCOVA, predicting the FSA score based on our research design. Second (RQ2), we investigated how users evaluated different labels and preference elicitation (PE) methods through Structural Equation Modelling, also assessing mediated relations. This analysis was performed in line with the recommender system user experience framework [32], relating our research design (objective system aspects) to user perception (effort) and experience (choice difficulty and satisfaction) while considering user characteristics (food knowledge) and behavior (healthiness of recipes chosen).

## 4.1. RQ1: Healthiness of Chosen Recipes

We predicted the FSA score of chosen recipes based on the labels presented and the user preference elicitation method. We also included food knowledge as a covariate and examined possible interaction effects with PE or labels, but did not observe any. Descriptive statistics indicated 64% of recipes were chosen in the MTL condition, compared to the 54% for the full labeling system.

The results of the two-way ANCOVA are presented in Table 3. The FSA score of chosen recipes was found to not significantly depend on the type of nutritional label presented: $F(2, 299) = 2.93$, $p = 0.055$. As the relatively small $p$-value shows, we did observe small differences across conditions, where the healthiest choices were made when facing MTL labels ($M_{FSA} = 7.17, SD_{FSA} = 2.01$), while scores were higher in the baseline ($M_{FSA} = 7.64, SD_{FSA} = 2.01$), and the full label condition ($M_{FSA} = 7.70, SD_{FSA} = 2.05$). However, these were not significant, suggesting that in a personalized choice context, cognitively oriented labelling nudges could not further support healthier recipe choices.

**Table 3**
Results of two-way ANCOVA tests, predicting the healthiness of chosen recipes across different labeling conditions and preference elicitation methods. Food knowledge was added as a covariate. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

| **Factor** (FSA score) | **df** | ***F*** |
|---|---|---|
| Model | 6 | 4.14** |
| Nutrition labels (No Label-MTL-Full label) | 2 | 2.93 |
| Preference Elicitation (CB-KB) | 1 | 8.30** |
| Labels * Preference Elicitation | 2 | .75 |
| Food Knowledge | 1 | 8.61** |

We did observe that the preference elicitation method employed had a significant effect on the healthiness of choices made. Participants using the knowledge-based method ($M_{FSA} = 7.14, SD_{FSA} = 2.03$) made healthier choices than those using a content-based approach ($M_{FSA} = 7.85, SD_{FSA} = 1.98$): $F(1, 299) = 8.30$, $p = 0.004$. This suggested that a knowledge-based PE allowed users to find recipes more easily. Our analysis further revealed that there was no significant interaction effect between the type of labels used and the preference elicitation method employed, which is also depicted in Figure 3. Finally, the two-way ANCOVA also revealed a relation between food knowledge and the FSA score, indicating that users with higher levels of food knowledge made healthier choices ($r(306) = -.17$).

#### 4.1.1. Conclusion

We found that annotating personalized recipes with either a Multiple Traffic Light or a Full label did not significantly lead to healthier recipe choices. Although previous studies suggested the possible merit of labels in recommender systems [27, 22], we only observed a small, non-significant ($p = 0.054$) improvement, particularly when using a front-of-pack Multiple Traffic Light label. In contrast, the full, back-of-pack nutrition facts label led to similar outcomes as in the baseline. As no interactions with food knowledge were found, this suggested that such cognitive digital nudges are less effective in a personalized recommender context, and choice outcomes do not depend on a label's understandability.

We did observe that a knowledge-based preference elicitation (PE) method led to healthier recipe choices. This indicated that a knowledge-based PE method could support users to make healthier choices, particularly by generating healthier recommendations as a translation

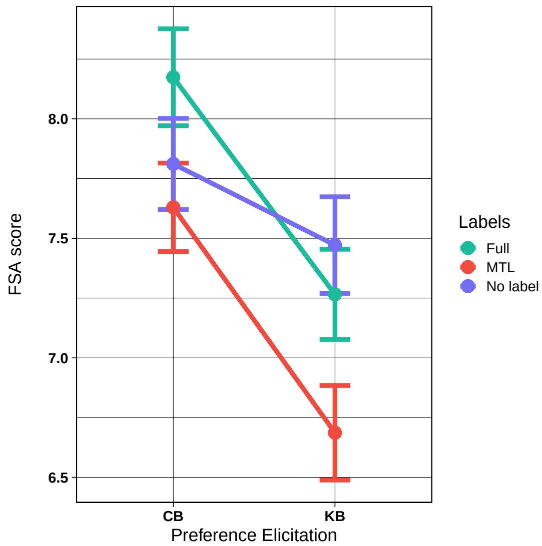



**Figure 3:** Mean FSA scores (i.e., inverse healthiness scores) of chosen recipes, across different conditions. Error bars depict 1 S.E.

to the user greater control in guiding the recommendations process in this recommender approach. Moreover, users with higher food knowledge made healthier choices. Although we did not observe any interaction effects on choice, we explored this relationship again in the next subsection, examining the interplay between knowledge level, PE methods and labelling conditions on perception and experience aspects.

### 4.2. RQ2: User Evaluation and Preference Elicitation Methods

To examine how users evaluated different interaction methods and interface nudges based on their knowledge level, we formed a Structural Equation Modeling (SEM). All objective and subjective aspects, along with user characteristics and interaction metrics, were organized in a path model. Following the guidelines by Knijnenburg and Willemsen [32], we first fitted a fully saturated model, organizing the path from objective aspects (i.e., the conditions) to perception (i.e., effort) and experience aspects (i.e., choice difficulty and satisfaction), after which non-significant relations were omitted.

The resulting model is presented in Figure 4, showing a good fit: $\chi^2(94)$, $p < .01$, $CFI = .963$, $TLI = .955$, $RMSEA = .038$, 90%-CI: $[.022; .052]$. The model met the guidelines for discriminant validity, as the correlations between latent constructs were smaller than the square root of each construct's Average Variance Explained (AVE). Please note that the chosen FSA score was included in our analysis, but it was not related to any mediated path.

Figure 4 depicts two paths towards choice satisfaction, stemming from the objective aspects. First, we observed an interaction effect between the use of an MTL label and a user's subjective food knowledge, in addition to a main effect of MTL. These effects be best explained through the marginal effects plot in Figure 5. We found that people facing MTL labels were on average, less satisfied with the recipe they had chosen than users in our conditions ($coef. = -1.412$,

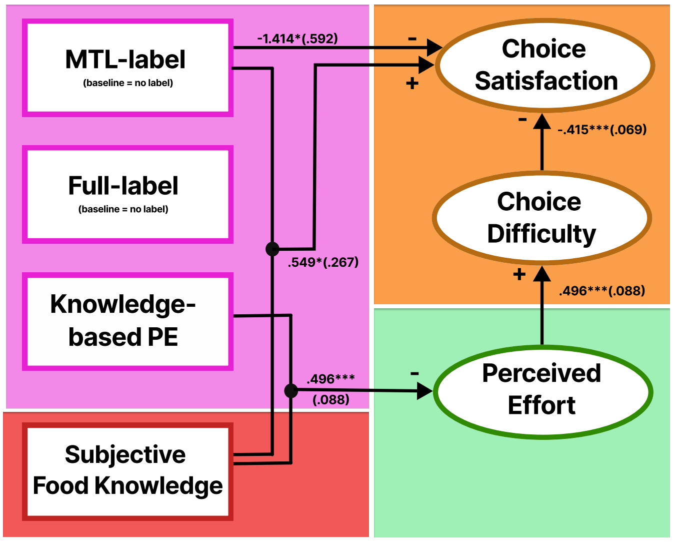



**Figure 4:** Structural Equation Model (SEM). Numbers on the arrows represent the $\beta$-coefficients, and standard errors are denoted between brackets. Effects between the subjective constructs are standardized and can be considered as correlations. Other effects show regression coefficients. Objective system aspects are purple, perception aspects are green, and experience aspects are orange. User characteristics are red. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

$p = 0.017$). However, the interaction effect with food knowledge showed that this particularly applied to users with low knowledge levels, as choice satisfaction significantly increased among users with a higher knowledge level facing MTL labels ($coef. = .547$, $p = 0.041$).

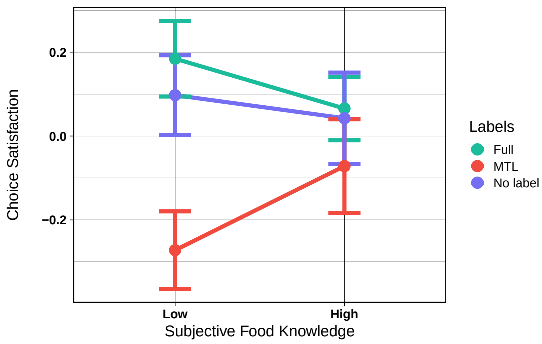
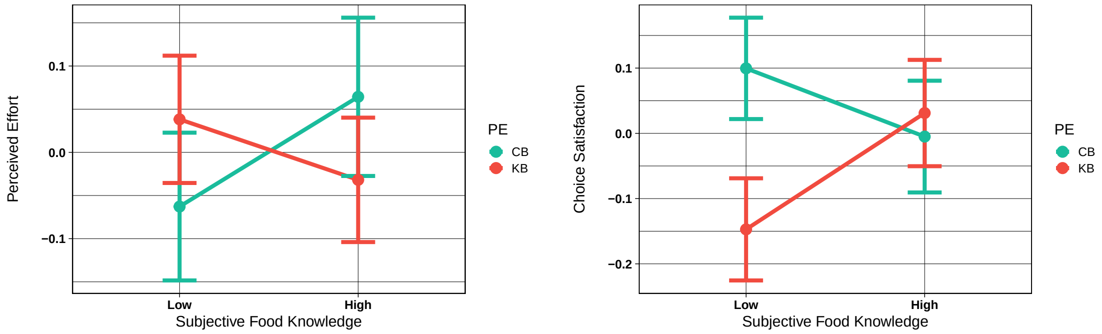



**Figure 5:** Standardized score for choice satisfaction across different labeling conditions and subjective food knowledge levels. Errors bars represent 1 S.E.

The path towards perceived effort stemmed from the interaction between the PE method and food knowledge. Higher levels of food knowledge led to lower levels of perceived effort among those facing a knowledge-based recommender ($coef. = -.436$, $p < .05$). To better understand this effect, please inspect Figure 6a, which depicts a two-sided interaction effect. For a KB recommender, effort was slightly reduced for users with higher knowledge levels. For a CB recommender, in contrast, perceived effort increased among users with higher knowledge levels.




(a) Scores for **Perceived Effort**

(b) Scores for **Choice Satisfaction**

**Figure 6:** Standardized Scores for Perceived Effort (a) and Choice Satisfaction (b), across different preference elicitation (PE) methods and subjective food knowledge levels. Errors bars represent 1 S.E.

This suggested that the perception of recommender use depended on the interplay between knowledge and the PE method, in line with [13].

The perceived effort, in turn, affected the experience aspects. We observed a positive relationship between effort and choice difficulty $(coef. = .496, p < .001)$, suggesting that effortful interactions were also related to harder decision-making processes. The full path towards choice difficulty, as depicted in Figure 4, shows that effort decreased due to the interplay between knowledge-based PE and knowledge, which in turn was related to choice difficulty. However, a test of indirect effects did not reveal significant support for a path towards choice difficulty that was fully mediated by perceived effort: $\beta = -.216$, $p = .053$, even though a lack of statistical power may have undermined this test.

Finally, we observed a negative relation between choice difficulty and choice satisfaction $(coef. = -.415, p < .001)$. This relation was observed in various previous studies (cf. [32]), but here it highlighted that the effects of the interplay between the PE method and user knowledge affected different evaluative aspects. Figure 6b provides a visual representation of these effects, indicating that knowledge-based users were more satisfied if they had higher knowledge levels and vice versa for users of content-based recommenders. We examined whether the path towards choice satisfaction was fully mediated by perceived effort and choice difficulty, but we found no support: $\beta = .090$, $p = .063$.

#### 4.2.1. Conclusion

We examined the evaluative effects of the interplay between the preference elicitation (PE) method (i.e., content-based or knowledge-based) and user knowledge, operationalized as subjective food knowledge. In doing so, we also explored the interaction effects between labels and user knowledge. Our results revealed two ways in which choice satisfaction was affected. First, users facing MTL labels tended to be more satisfied with their choices if they had a higher knowledge level, while satisfaction levels were slightly lower on average for MTL. On the one hand, this showed the importance of considering domain knowledge in the evaluation of a food recommender system, similar to [10]. On the other hand, it showed that although users of MTL labels seemed to make slightly healthier choices (but not significantly so), it might have come at the cost of choice satisfaction.

The second main finding concerned the interplay between subjective food knowledge and the PE method. We found that both the perception and the experience of using the system depended on this interaction effect. Users with higher levels of domain knowledge evaluated a recommender more positively if it allowed them to disclose personal characteristics and feature-based preferences, as was done for the knowledge-based PE. In contrast, users with comparatively low levels of domain knowledge evaluated the recommender more positively when facing a content-based method. Although the fully mediated path method was significant, this 'crossed' interaction effect was observed for all of the evaluative aspects.

Overall, this study showed that effective personalization in food recommender systems goes beyond algorithmic optimization. In fact, the design of the recommender, in terms of nudges and PE methods, seemed to require careful consideration of the user's knowledge level. This support the calls for adaptive preference elicitation methods [10].

## 5. Discussion

For many years, food recommender systems have fallen in line with traditional recommender approaches, focusing on algorithmic optimization. This paper has built upon research in which the interaction methods and interface aspects of a recommender are adapted to support specific user choices [13]. We have singled out the role of a user's level of domain knowledge, for it may affect not only how recommendations are evaluated, but also what types of interface aspects and interaction methods are appropriate. In doing so, we have focused on the one hand on cognitive food nudges [15], in the form of nutrition labels. On the other hand, we have examined the role of the recommender's preference elicitation method [10, 13], regarding it as a possible barrier for some users to due to either its complexity (knowledge-based) or simplicity (content-based).

We have set out with two research goals, examining two dimensions of a food recommender system. Firstly, we sought to investigate whether food nutritional labels support recipe recommender users in making healthy food choices. Secondly, we have examined the interplay between subjective food knowledge and preference elicitation methods on the user's perception and experience in a food recommender system [32], being the first to do so in this domain.

The first main contribution of this paper (RQ1) indicates that annotating personalized recipes with MTL labels or Full labels does not significantly affect the healthiness of recipes chosen by users. This finding falls in line with previous research in food recommender research [27, 22], which found that boosting nutritional food labels is the best way to help users make healthier food choices in a personalized interface, rather than using nudges only. Hence, although digital nudging in recommender systems has gained attention [28], its effectiveness might be limited due to the personalized decision-making context.

Interestingly, our study also reveals that a knowledge-based PE method (and subsequent recommender) leads to healthier outcomes than a more simple content-based recommender. It is possible that simply allowing users to reflect on their own preferences and needs leads to healthier outcomes than relying on recipe ingredients and images only. In the context of the psychological dual-process thinking [29], knowledge-based recommenders might encourage system-2 thinking, involving extensive and conscious deliberation, while content-based recommenders might elicit intuitive choices (based on system-1). This is consistent with the type of nudges related to effect (e.g., adapting images [18]) and cognition (e.g., labels) [15]. This also highlights the importance of considering user preferences and the interaction method, along with nutritional education, when examining food recommenders.

The second main contribution concerns the interplay between the user's nutritional knowledge and the preference elicitation method. This interaction is shown to influence various evaluative aspects of food recommender systems (RQ2). We find that users with higher levels of food knowledge experience additional benefits when using a knowledge-based recommender and vice versa for a content-based recommender. This has been operationalized into a path model that includes perceived effort, choice difficulty, and choice satisfaction.

Our findings are largely consistent with the work done in the energy recommender system domain [13]. That work differentiates between 'case-based PE' and 'attributed-based PE', which we regard to be similar to our content-based and knowledge-based approaches, respectively. Where both case-based PE and content-based PE (dis)like individual recipes, attribute-based PE differs slightly from a knowledge-based recommender. In [10, 13], users had to make tradeoffs between different energy-saving measure features but did not disclose any personal characteristics. In a knowledge-based recommender, the interactions that involve disclosing needs or personal preferences might have been easier to do, compared to some attribute-based tradeoffs. Another difference is that we have not been able to relate the FSA score in our path model to other aspects, while Knijnenburg et al. [13] also observe a relation between choice satisfaction and the interaction metric 'kWhs saved'.

Another striking finding from our structural equation model is that the evaluation of different labels depends on user knowledge. Where we expected this to be strongest for the full label due to its complexity, we have observed a positive interaction between domain knowledge and the use of an MTL label on choice satisfaction. It seems that the comparatively simple front-of-package label [16] still requires a significant level of understanding to be used satisfactorily.

### 5.1. Limitations and Future Work

A few limitations might confound parts of this paper. Unfortunately, we have had to exclude around 15% of our participants due to one or more issues regarding non-attention. The fact that this group is rather large could suggest that more users have not engaged with the recommended content with much deliberation. Intuition-based decisions could have undermined the effectiveness of our cognitive labeling nudge [15]. Nonetheless, since we have observed some differences regarding labels in terms of choice and evaluation, we argue that a sufficient number of participants is still part of this study's analysis sample. This also applies to use of Structural Equation Modelling, for which we had a sufficient number of degrees of freedom [41].

The recommender in this study has focused on dinner recipes. Although this is quite a common approach for recipe recommendation [27, 24, 5, 1], it is challenging to assess the implications or a slightly (un)healthier dinner meal if nothing is known about the daily dietary intake of a user. For example, it could be that a person eats relatively healthy dinner meals but has numerous eating moments a day, thereby exceeding the caloric intake limit. We advocate for extending this recommender approach towards meal plans for the day, possibly mixing recipe recommendations with food product recommendations.

Another limitation, which is shared with many food recommender studies [2], is that we have not checked whether people actually cooked the food. In that sense, the choices made in this study can only be regarded as behavioral intention. Although some commitment mechanisms may take place that may support actual engagement with chosen recipes, we would encourage performing a follow-up study that considers longitudinal aspects as well. An increasing number of health-based personalized advice applications are developed [8], among others using Digital Twins as a means of user profiling to suggest meal plans or exercise behavior. Overall, it is important to examine whether adaptations in a recommender interface can spill over into longer last effects. For instance, a knowledge-based recommender can only be regarded as being effective in supporting healthier choice if these last over the course of a few weeks.

## Acknowledgments

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

## References

[1] C. Trattner, D. Elsweiler, Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems, in: Proceedings of the 26th international conference on world wide web, ACM, New York, NY, USA, 2017, pp. 489–498.

[2] C. Trattner, D. Elsweiler, Food recommendations, in: Collaborative recommendations: Algorithms, practical challenges and applications, World Scientific, 2019, pp. 653–685.

[3] S. Mills, M. White, H. Brown, W. Wrieden, D. Kwasnicka, J. Halligan, S. Robalino, J. Adams, Health and social determinants and outcomes of home cooking: A systematic review of observational studies, Appetite 111 (2017) 116–134.

[4] L. McGowan, M. Caraher, M. Raats, F. Lavelle, L. Hollywood, D. McDowell, M. Spence, A. McCloat, E. Mooney, M. Dean, Domestic cooking and food skills: A review, Critical reviews in food science and nutrition 57 (2017) 2412–2431.

[5] A. D. Starke, E. Asotic, C. Trattner, E. J. Van Loo, Examining the user evaluation of multi-list recommender interfaces in the context of healthy recipe choices, ACM Transactions on Recommender Systems (2023).

[6] D. Elsweiler, H. Hauptmann, C. Trattner, Food recommender systems, in: F. Ricci, L. Rokach, B. Shapira (Eds.), Recommender Systems Handbook, Springer US, New York, NY, 2022, pp. 871–925. URL: https://doi.org/10.1007/978-1-0716-2197-4_23.

[7] J. Freyne, S. Berkovsky, Intelligent food planning: personalized recipe recommendation, in: Proceedings of the 15th international conference on Intelligent user interfaces, ACM, New York, NY, USA, 2010, pp. 321–324.

[8] M. J. Reinders, A. D. Starke, A. R. Fischer, M. C. Verain, E. L. Doets, E. J. Van Loo, Determinants of consumer acceptance and use of personalized dietary advice: A systematic review, Trends in Food Science & Technology (2022).

[9] T. N. T. Tran, M. Atas, A. Felfernig, M. Stettinger, An overview of recommender systems in the healthy food domain, Journal of Intelligent Information Systems 50 (2018) 501–526.
[10] B. P. Knijnenburg, M. C. Willemsen, Understanding the effect of adaptive preference elicitation methods on user satisfaction of a recommender system, in: Proceedings of the third ACM conference on Recommender systems, 2009, pp. 381–384.
[11] D. Elsweiler, C. Trattner, M. Harvey, Exploiting food choice biases for healthier recipe recommendation, in: Proceedings of the 40th international acm sigir conference on research and development in information retrieval, 2017, pp. 575–584.
[12] A. Starke, M. Willemsen, C. Snijders, Promoting energy-efficient behavior by depicting social norms in a recommender interface, ACM Transactions on Interactive Intelligent Systems (TiiS) 11 (2021) 1–32.
[13] B. Knijnenburg, M. Willemsen, R. Broeders, Smart sustainability through system satisfaction: tailored preference elicitation for energy-saving recommenders, in: 20th Americas Conference on Information Systems (AMCIS 2014), August 7-9, 2014, Savannah, Georgia, United States, AIS/ICIS, 2014, pp. 1–15.
[14] C. Musto, C. Trattner, A. Starke, G. Semeraro, Towards a knowledge-aware food recommender system exploiting holistic user models, in: Proceedings of the 28th ACM conference on user modeling, adaptation and personalization, ACM, New York, NY, USA, 2020, pp. 333–337.
[15] R. Cadario, P. Chandon, Which healthy eating nudges work best? a meta-analysis of field experiments, Marketing Science 39 (2020) 465–486.
[16] M. Egnell, I. Boutron, S. Péneau, P. Ducrot, M. Touvier, P. Galan, C. Buscail, R. Porcher, P. Ravaud, S. Hercberg, et al., Randomised controlled trial in an experimental online supermarket testing the effects of front-of-pack nutrition labelling on food purchasing intentions in a low-income population, Bmj Open 11 (2021) e041196.
[17] K. L. Hawley, C. A. Roberto, M. A. Bragg, P. J. Liu, M. B. Schwartz, K. D. Brownell, The science on front-of-package food labels, Public health nutrition 16 (2013) 430–439.
[18] A. D. Starke, M. C. Willemsen, C. Trattner, Nudging healthy choices in food search through visual attractiveness, Frontiers in Artificial Intelligence 4 (2021) 20.
[19] S. Barko-Sherif, D. Elsweiler, M. Harvey, Conversational agents for recipe recommendation, in: Proceedings of the 2020 Conference on Human Information Interaction and Retrieval, 2020, pp. 73–82.
[20] G. Castiglia, A. El Majjodi, F. Calò, Y. Deldjoo, F. Narducci, A. Starke, C. Trattner, Nudging towards health in a conversational food recommender system using multi-modal interactions and nutrition labels (2022).
[21] C. Ischen, T. B. Araujo, H. A. Voorveld, G. Van Noort, E. G. Smit, Is voice really persuasive? the influence of modality in virtual assistant interactions and two alternative explanations, Internet Research (2022).
[22] A. Starke, A. El Majjodi, C. Trattner, Boosting health? examining the role of nutrition labels and preference elicitation methods in food recommendation, in: Interfaces and Human Decision Making for Recommender Systems 2022, 2022, pp. 67–84.
[23] M. Elahi, M. Ge, F. Ricci, D. Massimo, S. Berkovsky, Interactive food recommendation for groups, in: Poster Proceedings of the 8th ACM Conference on Recommender Systems, RecSys 2014, Foster City, Silicon Valley, CA, USA, October 6-10, 2014, 2014.

[24] C. Musto, A. D. Starke, C. Trattner, A. Rapp, G. Semeraro, Exploring the effects of natural language justifications in food recommender systems, in: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, 2021, pp. 147–157.
[25] N. J. Temple, Front-of-package food labels: A narrative review, Appetite 144 (2020) 104485.
[26] M. W. Kreuter, L. K. Brennan, D. P. Scharff, S. N. Lukwago, Do nutrition label readers eat healthier diets? behavioral correlates of adults' use of food labels, American journal of preventive medicine 13 (1997) 277–283.
[27] A. El Majjodi, A. D. Starke, C. Trattner, Nudging towards health? examining the merits of nutrition labels and personalization in a recipe recommender system, in: Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, 2022, pp. 48–56.
[28] M. Jesse, D. Jannach, Digital nudging with recommender systems: Survey and future directions, Computers in Human Behavior Reports 3 (2021) 100052.
[29] D. Kahneman, Thinking, fast and slow, Macmillan, 2011.
[30] A. Berge, V. Velle Sjøen, A. D. Starke, C. Trattner, Changing salty food preferences with visual and textual explanations in a search interface, in: HEALTHI'21: Joint Proceedings of ACM IUI 2021 Workshops, CEUR, 2021.
[31] Department of Health and Social Care UK, Front of Pack nutrition labelling guidance, 2016. URL: https://www.gov.uk/government/publications/front-of-pack-nutrition-labelling-guidance.
[32] B. P. Knijnenburg, M. C. Willemsen, Evaluating recommender systems with user experiments, in: Recommender systems handbook, Springer, 2015, pp. 309–352.
[33] C. Trattner, D. Elsweiler, An evaluation of recommendation algorithms for online recipe portals, in: Proceedings of the 4th International Workshop on Health Recommender Systems, CEUR, Aachen, DE, 2019.
[34] J. Oh, K. Yun, J.-H. Chae, T.-S. Kim, Association between macronutrients intake and depression in the united states and south korea, Frontiers in psychiatry 11 (2020) 207.
[35] C. N. Sutanto, M. X. Wang, D. Tan, J. E. Kim, Association of sleep quality and macronutrient distribution: a systematic review and meta-regression, Nutrients 12 (2020) 126.
[36] S. Moats, C. W. Suitor, A. L. Yaktine, S. P. Murphy, et al., Child and adult care food program: Aligning dietary guidance for all (2011).
[37] L. R. Flynn, R. E. Goldsmith, A short, reliable measure of subjective knowledge, Journal of business research 46 (1999) 57–66.
[38] Z. Pieniak, J. Aertsens, W. Verbeke, Subjective and objective knowledge as determinants of organic vegetables consumption, Food quality and preference 21 (2010) 581–588.
[39] A. Starke, M. Willemsen, C. Snijders, Effective user interface designs to increase energy-efficient behavior in a rasch-based energy recommender system, in: Proceedings of the eleventh ACM conference on recommender systems, 2017, pp. 65–73.
[40] M. C. Willemsen, M. P. Graus, B. P. Knijnenburg, Understanding the role of latent feature diversification on choice difficulty and satisfaction, User Modeling and User-Adapted Interaction 26 (2016) 347–389.
[41] J. C. Westland, Lower bounds on sample size in structural equation modeling, Electronic commerce research and applications 9 (2010) 476–487.

# Manuscript V

## Nudging Towards Health in a Conversational Food Recommender System Using Multi-Modal Interactions and Nutrition Labels.

*Giovanni Castiglia, Ayoub El Majjodi, Federica Calò, Yashar Deldjoo, Fedelucio Narducci1, Alain Starke, Christoph Trattner. KaRS@ RecSys 2022.*

# Nudging Towards Health in a Conversational Food Recommender System Using Multi-Modal Interactions and Nutrition Labels

Giovanni Castiglia[1], Ayoub El Majjodi[2], Federica Calò[1], Yashar Deldjoo[1], Fedelucio Narducci[1], Alain Starke[2,3] and Christoph Trattner[2]

[1]*Polytechnic University of Bari, Bari, Italy*

[2]*Department of information science and media studies, University of Bergen, Bergen, Norway*

[3]*Marketing and Consumer Behaviour Group, Wageningen University & Research, Wageningen, The Netherlands*

**Abstract**

Humans engage with other humans and their surroundings through various modalities, most notably speech, sight, and touch. In a conversation, all these inputs provide an overview of how another person is feeling. When translating these modalities to a digital context, most of them are unfortunately lost. The majority of existing conversational recommender systems (CRSs) rely solely on natural language or basic click-based interactions.

This work is one of the first studies to examine the influence of multi-modal interactions in a conversational food recommender system. In particular, we examined the effect of three distinct interaction modalities: pure textual, multi-modal (text plus visuals), and multi-modal supplemented with nutritional labeling. We conducted a user study ($N$=195) to evaluate the three interaction modalities in terms of how effectively they supported users in selecting healthier foods. Structural equation modelling revealed that users engaged more extensively with the multi-modal system that was annotated with labels, compared to the system with a single modality, and in turn evaluated it as more effective.

**Keywords**

Personalization, Health, Food recommendation, Digital Nudges, Nutrition labels

## 1. Introduction and Context

Conversational recommender systems (CRSs) represent a hotly debated area of study in the field of information seeking [1, 2]. They combine the power of recommendation algorithms with conversational strategies. Using multi-turn conversations, CRSs are able to collect users' nuanced and dynamic preferences in more depth, which can enhance recommendation outcomes and user experience. CRSs are utilized in a variety of domains, including medical diagnosis [3], e-commerce [4], and entertainment [5, 6]. Only a few studies have investigated their merit for food recommendation [7], and in particular for encouraging users to make *healthier* food decisions.

Over 60% of all deaths are caused by non-communicable diseases, which are preventable by tackling risk factors, such as attaining a healthy food intake [8]. While our food decisions are driven by our overall preferences, the food selection process is extremely contextual and influenced by a variety of factors, such as the user's mood and dietary constraints. Moreover, many of the decisions are made spontaneously and consumers' judgments are influenced by factors unrelated to the food content, such as their perception of the food's visual characteristics [9]. For instance, the packaging of items with nutritional labels can serve to highlight the nutritious nature of the food (cf. [10]). Moreover, people generally prefer food that has a more visually appealing presentation, such as food that is presented in an attractive way [11]. People are willing to pay extra for food whose ingredients are tastefully/attractively organized, and restaurants strive to generate Instagram-friendly photographs by enhancing the color composition of their plates.

To surface effective and healthy food recommendations it is crucial to understand these underlying decision factors. Regrettably, the large majority of existing conversational recommender systems [12, 13] only consider a single type of interaction, such as natural language or click-based interaction, thereby neglecting a wealth of information in the actual imaging of meals [14]. The goal of the present work at hand is to employ a new conversational model for food recommendation that permits more natural, multi-modal user-system interaction.

*4th Edition of Knowledge-aware and Conversational Recommender Systems (KaRS) Workshop @ RecSys 2022, September 18–23 2023, Seattle, WA, USA.*

g.castiglia@studenti.poliba.it (G. Castiglia); ayoub.majjodiu@uib.no (A. E. Majjodi); f.calo8@studenti.poliba.it (F. Calò); yashar.deldjoo@poliba.it (Y. Deldjoo); fedelucio.narducci@poliba.it (F. Narducci); alain.starke@uib.no (A. Starke); christoph.trattner@uib.no (C. Trattner)
https://www.christophtrattner.info/ (C. Trattner)
0000-0002-7478-5811 (A. E. Majjodi); 0000-0002-6767-358X (Y. Deldjoo); 0000-0002-9255-3256 (F. Narducci); 0000-0002-9873-8016 (A. Starke); 0000-0002-1193-0508 (C. Trattner)

To attain this goal, this paper introduces a *multi-modal conversational food recommender system* (MM-CFRS). It implements different user-system interaction modes, along with nutrition labelling in order to assist the user in making dietary decisions. Our objective is to examine the effects of three distinct interaction modes: pure textual, multi-modal (text plus visuals), and multi-modal supplemented with *nutritional labeling*. While multi-modal conversational information seeking (MMCIS) is gaining attention by the research in the RecSys/IR/HCI communities [15, 1, 16], only a few practical studies have been published that focus on topics other than food and health, such as conservational systems on tourism [17] and fashion [18, 19]. In the field of food recommendation, Elsweiler et al. [20] provide a good frame of reference for recent advances in the field of food recommender systems in general. Specifically for conversational systems, Barko-Sherif et al. [21] investigate the possibility for conversational preference elicitation in a food recommender environment, using a Wizard of Oz study design (see also [22]). Using a between-groups approach, they compare spoken and text-input chat interfaces and reported that such interfaces are useful for users to describe their needs and preferences. In other studies, Samagaio et al. [23] present a RASA-based chatbot that can recognize and categorize user intentions in the conversation aimed to elicit food preferences for recommendation purposes. Another study of Samagaio et al. [24] applies more knowledge-based elements based on word embedding to optimize conversational ingredient retrieval. These studies, however, focus less on aspects pertaining to health, health labelling, or elicitation modalities. In a non-conversational recommender context, El Majjodi et al. [25] recently indicated that nutritional labels can reduce user's choice difficulty in non-conversational context. The primary distinction between our work and previous studies is the lack of multiple modalities (typically only text is used), as well as that only a few studies (e.g., [25]) have used nutrition labelling.

To summarize, the goal of this study is to compare the impact of three user-system interaction and explanation modalities (textual, multi-modal, and multi-modal with nutritional labels) on both behavioral aspects (what type of recipe is chosen? How healthy is that recipe?) and evaluation aspects (how does the user evaluate the system or their chosen recipe?). Using a mediation analysis (structural equation modelling), we answer the following research question:

- *RQ:* To what extent do different interaction modalities affect a user's recipe choices and evaluation in a conversational food recommendation scenario?

To address this question, we consider different dimensions of analysis. This includes system interaction length, presentation time, healthiness of recipes chosen and a user's level of choice satisfaction and experienced system effectiveness.

## 2. System Design

In this section we describe the features of our conversational food recommender system, which supports users in making healthier choices.[1]

We designed a system-driven conversation in which the system requires user feedback (response/input) to continue. The main steps of the conversational flow are shown in Figure 1. Users can interact with the system using both buttons and textual messages[2]. The main steps of the interaction are reported below:

- *Food category acquisition*: The user was presented with a choice of *four* different food categories that were considered in this work: Pasta, Salad, Dessert, and Snack.
- *User constraints acquisition*: The user was then prompted to indicate any potential dietary constraints. Initially, the system used an interface with a single checkbox for each of the most prevalent *intolerances* and *allergies*: Lactose, Meat, Alcohol, Seafood, Reflux, Cholesterol, Diabetes. Afterwards, the system asked the user to disclose a list of ingredients she could not consume.
- *Preference elicitation*: According to the constraints specified by the user, the user was prompted to submit preferences for five of the dishes proposed by the system. Each dish was accompanied with two buttons: "Like" and "Skip". The skip option was provided to encourage users to inspect an addition dish, which was retrieved from the randomly sorted menu. The retrieval was based on a random active learning strategy. This way, users were encouraged to like five dishes they were interested in, after which the user profile was built by the system.
- *Processing*: The system constructed the *user profile* by analyzing the user's five preferences from the previous stage. The cosine similarity was computed between the user profile and each of the available foods in the catalog, to provide a list of dishes from which recommendations would be selected. The algorithm also provided a list of dishes ranked according to their healthiness (based on their FSA score; see Section 3).

[1]Code and recipe data used for implementing the chatbot are available at https://github.com/giocast/MMCFRS

[2]A video demo of the three versions of our system is available at https://tinyurl.com/mtzxr2sw

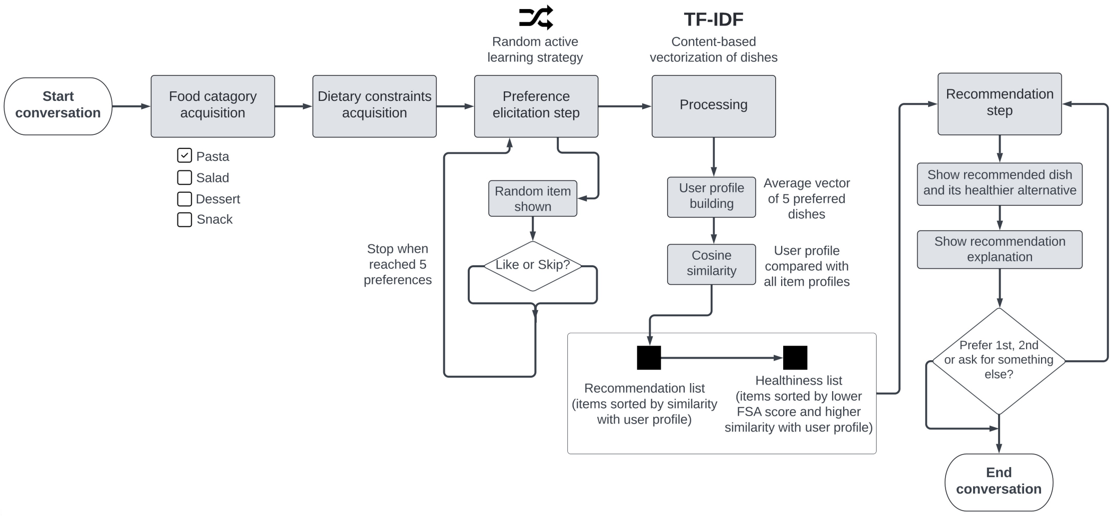



**Figure 1:** Our conversational recommender system flow.

For each food category we built a matrix containing the TF-IDF representation (dish vs. ingredient) of dishes in the catalog. The higher the TF-IDF score, the greater the ingredient's significance to this dish (as opposed to other dishes).

- *Recommendation and explanation*: The system provided two personalized recommendations, based on the user's preferences. The system constrained the retrieval to ensure that the two options differed in terms of healthiness, so that one option was healthier than the other. Thus, the algorithm provided a description of the suggested dishes. Specifically, it explained why the second dish was healthier than the first and why the advice was made. The user would then be prompted to select one or request a new recommendation. The two recommended dishes were chosen using the following strategy: The first dish would be the most similar to the user profile, while the second dish (the healthier alternative) was selected from a list of most similar dishes ranked on their FSA scores, selecting the healthiest one (i.e. with the lowest FSA score).

Three different interaction modes were implemented by modifying the values associated with the two manipulated variables: interaction *I* and explanation *E*, according to Table 1.

In the *Pure text* version (T + T), the system communicates with the user solely through text, displaying simply the dish titles and offering textual explanations of the food recommendations. In the *Multi-modal* version (MM + T), the system engages the user in a multi-modal manner by displaying the name and image of each dish throughout the dialogue. However, the supplied explanation remains textual. For the first dish, the explanation can be like "I recommend these dish because I know that you have diet constraints due to: meat, zucchini. The first dish I proposed contains ingredients that you might like: carrot, lemon, tuna, olive oil". For the second recommendation, the explanation further provides information about macro nutrients quantities of the two recommended dishes and can be in the form of "The second dish I proposed has less calories (54 Kcal) than the first one (123 Kcal) and has less fats than the first one. The third version *MM-Label* (MM + MM) likewise employs a multi-modal interaction approach, but it also makes use of nutritional explanations in the form of a front-of-package nutrition label with FSA's Multiple Traffic Lights (MTL) [25]. MTL nutrition labels depicted the intake adequacy of a dish in terms of energy and nutritional content, along five dimensions: energy (kcal), fat, saturates, sugars, and salt. This adequacy, per serving and per 100g, was depicted using the colors green, yellow and red, where green indicated a dish to adhere to the nutritional intake guideline, while red indicated that the content was unacceptable. These labels were generated

**Table 1**
Differences between three implementations of the system.

| Interaction Mode | I | E |
|---|---|---|
| Pure text (T) | T | T |
| Multi-modal (MM) | MM | T |
| Multi-modal with labels (MM-Label) | MM | MM |

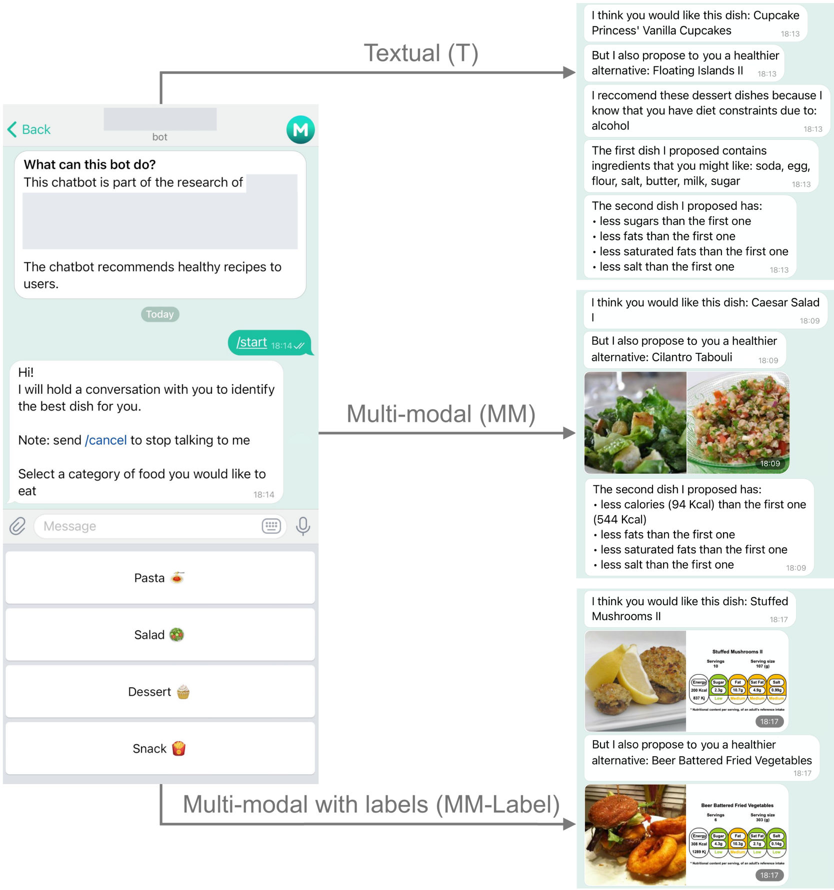



**Figure 2:** The three implementations of the system. Some details displayed on the interface, such as the chatbot's and authors' names are anonymized and will be added after peer review.

for each dish by following the directives of Food Standard Agency and UK department of health [26].

Figure 2 depicts a snapshot of the chatbot prototype, visualizing the different interaction phases.

In the *Textual (T)* version, the user received recommendations identified by only the names of the dishes (e.g., Cupcake Princess' Vanilla Cupcakes, Floating Island II). The recommendations were followed by textual explanations, based on the ingredients in the dish that the user likes. A comparative analysis of the nutritional facts (e.g., 'less sugars') would also be provided. In the *Multi-modal (MM)* version, the system additionally provided images of the recommended dishes. The explanation was similar to the one presented in the *T* version. Finally, the *Multi-modal with labels (MM-Label)* version provided nutritional labels that were annotated to the depicted images (e.g., Sugar 2.3g, Fat 10.7g, etc.) presented with red, yellow, and/or green colors according to the FSA score. As stated previously, following the presentation of the recommendations, we provide the user with an explanation that helps her comprehend the health benefits of the second alternative above the first, which is the dish that best matches her preferences. This is accomplished either by text (T and MM variants) or a multiple traffic light nutritional label (MM-Label).

The user can accept one of the two dishes proposed or can ask for another recommendation.

## 3. Experimental Evaluation

To evaluate the extent to which different versions of the chatbot affected users' evaluations and decisions, we recruited 195 participants from Amazon MTurk to use our system. Participants had to have a hit rate of 95% at least and were compensated with 2 dollars. On average, user required around 15 minutes to complete the study.[3] Users

[3] The research conformed to the ethical standards of the Norwegian Centre for Research Data (NSD). The collected data is available in

**Table 2**
Questionnaire items used in the confirmatory factor analysis. Alpha denotes Cronbach's Alpha, AVE denotes the Average Variance Explained, indicating construct validity if AVE > 0.5. Items in gray and without loading were omitted from analysis. Choice Satisfaction did not form a sensible aspect, because of a lack of construct validity.

| Aspect | Item | Loading |
|---|---|---|
| Choice Satisfaction | I think, I would enjoy eating the dish I have chosen in the end | |
| | I would recommend the dish I've chosen in the end to others | |
| | My chosen dish could become my favorite | |
| System Effectiveness | It was easy to make my final choice on the dish | 0.737 |
| | I interacted a lot with the system before getting the dish of my choice | |
| | The explanation influenced my final choice of dish | |
| | I think, that I would use this system frequently | |
| Alpha = 0.740 | I found the system easy to use and understand | 0.724 |
| AVE = 0.534 | I felt very confident using the system | 0.661 |
| | I would imagine that most people would learn to use this system very quickly | 0.722 |

performed the processes outlined in Section 2, interacting with our chatbot for preference elicitation, evaluating recipe recommendations, selecting one recipe, and evaluating the experience. A user's experience was evaluated through choice satisfaction and system effectiveness, using questionnaire items that were evaluated on 5-point Likert scales.

Chosen recipes were evaluated according to their healthiness. This was evaluated using the FSA score [27]. Each recipe was scored between 4 and 12, where 4 indicated that all four nutrients (sugar, fat, saturated fat, salt) adhered to nutritional guidelines per 100g [9, 28], while 12 would indicate that a recipe was unhealthy because of all nutritional contents being too high.

The responses to the evaluation questionnaire item were submitted to a confirmatory factor analysis (CFA; see Table 2). Unfortunately, we could not infer a reliable construct for choice satisfaction, as the variance explained by the questionnaire items was too low, while Cronbach's Alpha was only acceptable (0.60). Other items were dropped from the system effectiveness aspect because of low factor loadings.

We organized the different factors (e.g., conversation time, condition factors) and aspects (i.e., system effectiveness) into a path model using Structural Equation Modelling. Figure 3 depicts the resulting model, which had decent fit statistics: $\chi^2(17) = 28.064$, $p < 0.05, CFI = 0.969$, $TLI = 0.954$, $RMSEA = 0.058$, $90\% - CI$: $[0.009, 0.095]$. The relevant AVEs of the aspects was sufficiently high to form a path model [29].

Our analysis revealed that the MM-Label condition with nutrition labels (MM-label) stood out in terms of how long users interacted with our chatbot. Figure 3 illustrates this, while the use of multi-modal approaches alone had no effect on the interaction or evaluation factors considered. For MM-Label, our mediation analysis suggested that in the MM-Label condition, the conversation duration was significantly longer ($p < 0.05$) than in the text-based condition . This indicated that the usage of nutrition labels affected conversation time, on top of the other modalities.

The duration of the conservation affected, in turn, the evaluation of the user. Inferred from our confirmatory factor analysis (cf. Table 2), users who interacted with the chatbot for longer periods of time indicated greater levels of system effectiveness ($p < 0.01$). This indicated that an extended engagement did not frustrate users. Instead, it indicated that they were enthusiastic about using the system. Figure 3 also shows that the healthiness of chosen recipes was not significantly related to any of the other aspects or factors. Note that the MM-Label condition led the healthiest recipe choices, but the differences with the other conditions were not significant.

## 4. Conclusion and Future Work

We have presented a novel chatbot-like recommender system that introduces multi-modality in interaction with user, presentation of results and explanation of the recommendations with nutrition labels in a conversational scenario. We have designed and analyzed the impact of three distinct version of our chatbot: pure textual, multi-modal (use of text and images), and multi-modal supplemented with nutritional labels.

Our experimental evaluation reveals that our chatbot is the most effective when accompanied by explanatory labels. This is indicated by the length of conversation, as well as by the user's evaluation of the system effectiveness.

Limitations to this study could be viewed from different viewpoints. In terms of analysis, we have been unable to infer the choice satisfaction evaluation aspect. Other research have demonstrated that decision satisfaction is a good predictor of post-interaction engagement with selected item, such as for household energy con-

the project's GitHub repository.

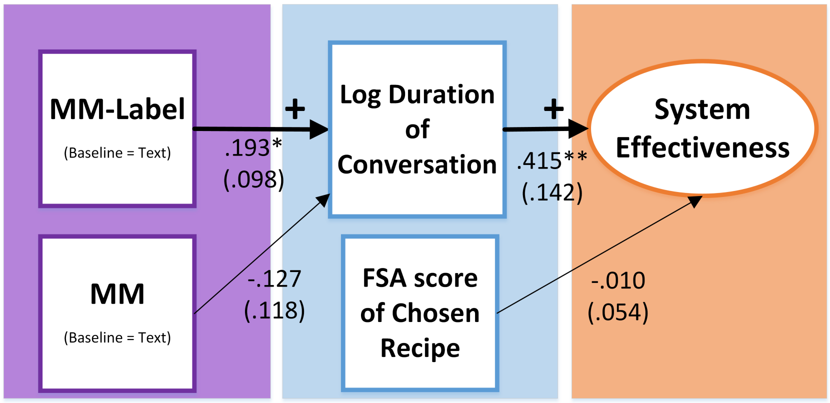



**Figure 3:** Structural Equation Model (SEM). Numbers on the arrows represent the $\beta$-coefficients, standard errors are denoted between brackets. Effects between the subjective constructs are standardized and can be considered as correlations, other effects show regression coefficients. Aspects are grouped by color: Objective system aspects are purple, behavioral indicators are blue (note: the FSA score represents recipe unhealthiness) and experience aspects are orange. The thinner arrows are non-significant relations, in addition: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

servation [30]. Moreover, rather than relying solely on system-driven interaction, it might be intriguing and natural to investigate *user-driven* scenarios in which users might query the system with an image and textual query. The food categories considered in this work (pasta, salad, dessert, snack) could additionally be expanded to include more meal categories and their combinations, such as to create a complete meat (first dish, second dish and vegetables). On top of that, the distinctions between various label modalities are an additional intriguing topic we wish to investigate more in-depth [31].

## References

[1] H. Zamani, J. R. Trippas, J. Dalton, F. Radlinski, Conversational information seeking, arXiv preprint arXiv:2201.08808 (2022).
[2] D. Jannach, A. Manzoor, W. Cai, L. Chen, A survey on conversational recommender systems, ACM Computing Surveys 54 (2022) 1–36. doi:10.1145/3453154.
[3] P. Cordero, M. Enciso, D. López, A. Mora, A conversational recommender system for diagnosis using fuzzy rules, Expert Systems with Applications 154 (2020) 113449. doi:10.1016/j.eswa.2020.113449.
[4] D. Griol, J. Milina, From voicexml to multimodal mobile apps: development of practical conversational interfaces, ADCAIJ Adv. Distrib. Comput. Artif. Intell. J. 5 (2016) 43.
[5] F. Narducci, P. Basile, M. de Gemmis, P. Lops, G. Semeraro, An investigation on the user interaction modes of conversational recommender systems for the music domain, User Model. User Adapt. Interact. 30 (2020) 251–284. URL: https://doi.org/10.1007/s11257-019-09250-7. doi:10.1007/s11257-019-09250-7.
[6] A. Iovine, F. Narducci, G. Semeraro, Conversational recommender systems and natural language: : A study through the converse framework, Decis. Support Syst. 131 (2020) 113250. URL: https://doi.org/10.1016/j.dss.2020.113250. doi:10.1016/j.dss.2020.113250.
[7] C. Trattner, D. Elsweiler, Food recommendations, in: Collaborative recommendations: Algorithms, practical challenges and applications, World Scientific, 2019, pp. 653–685.
[8] R. Y. Toledo, A. A. Alzahrani, L. Martinez, A food recommender system considering nutritional information and user preferences, IEEE Access 7 (2019) 96695–96711.
[9] A. D. Starke, M. C. Willemsen, C. Trattner, Nudging healthy choices in food search through visual attractiveness, Frontiers in Artificial Intelligence 4 (2021) 621743.
[10] E. J. Van Loo, C. Grebitus, J. Roosen, Explaining attention and choice for origin labeled cheese by means of consumer ethnocentrism, Food Quality and Preference 78 (2019) 103716.
[11] Y. Peng, J. B. Jemott III, Feast for the eyes: Effects of food perceptions and computer vision features on food photo popularity., International Journal of Communication (19328036) 12 (2018).
[12] C. Zhou, Y. Jin, K. Zhang, J. Yuan, S. Li, X. Wang, Musicrobot: Towards conversational context-aware music recommender system, in: International Conference on Database Systems for Ad-

vanced Applications, Springer, 2018, pp. 817–820.
[13] J. Schaffer, T. Hollerer, J. O'Donovan, Hypothetical recommendation: A study of interactive profile manipulation behavior for recommender systems, in: The Twenty-Eighth International Flairs Conference, 2015, pp. 507–512.
[14] Y. Deldjoo, M. Schedl, P. Cremonesi, G. Pasi, Recommender systems leveraging multimedia content, ACM Computing Surveys (CSUR) 53 (2020) 1–38.
[15] Y. Deldjoo, J. R. Trippas, H. Zamani, Towards multimodal conversational information seeking, in: Proceedings of the 44th International ACM SIGIR conference on research and development in Information Retrieval, 2021, pp. 1577–1587.
[16] R. G. Sousa, P. M. Ferreira, P. M. Costa, P. Azevedo, J. P. Costeira, C. Santiago, J. Magalhaes, D. Semedo, R. Ferreira, A. I. Rudnicky, et al., ifetch: Multimodal conversational agents for the online fashion marketplace, in: Proceedings of the 2nd ACM Multimedia Workshop on Multimodal Conversational AI, 2021, pp. 25–26.
[17] L. Liao, L. H. Long, Z. Zhang, M. Huang, T.-S. Chua, Mmconv: an environment for multimodal conversational search across multiple domains, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 675–684.
[18] S. Moon, S. Kottur, P. A. Crook, A. De, S. Poddar, T. Levin, D. Whitney, D. Difranco, A. Beirami, E. Cho, et al., Situated and interactive multimodal conversations, arXiv preprint arXiv:2006.01460 (2020).
[19] Y. Yuan, W. Lam, Conversational fashion image retrieval via multiturn natural language feedback, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 839–848.
[20] D. Elsweiler, H. Hauptmann, C. Trattner, Food recommender systems, in: Recommender Systems Handbook, Springer, 2022, pp. 871–925.
[21] S. Barko-Sherif, D. Elsweiler, M. Harvey, Conversational agents for recipe recommendation, in: Proceedings of the 2020 Conference on Human Information Interaction and Retrieval, 2020, pp. 73–82.
[22] A. Steinfeld, O. C. Jenkins, B. Scassellati, The oz of wizard: simulating the human for interaction research, in: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, 2009, pp. 101–108.
[23] Á. Mendes Samagaio, H. Lopes Cardoso, D. Ribeiro, A chatbot for recipe recommendation and preference modeling, in: EPIA Conference on Artificial Intelligence, Springer, 2021, pp. 389–402.
[24] Á. M. Samagaio, H. Lopes Cardoso, D. Ribeiro, Enriching word embeddings with food knowledge for ingredient retrieval, in: 3rd Conference on Language, Data and Knowledge (LDK 2021), Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.
[25] A. El Majjodi, A. D. Starke, C. Trattner, Nudging towards health? examining the merits of nutrition labels and personalization in a recipe recommender system, in: Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, 2022, pp. 48–56.
[26] Department of Health and Social Care UK, Front of Pack nutrition labelling guidance, 2016. URL: https://www.gov.uk/government/publications/front-of-pack-nutrition-labelling-guidance.
[27] D. of Health UK, F. S. Agency, Guide to creating a front of pack (fop) nutrition label for pre-packed products sold through retail outlets (2016). URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/566251/FoP_Nutrition_labelling_UK_guidance.pdf.
[28] C. Trattner, D. Elsweiler, Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems, in: Proceedings of the 26th international conference on world wide web, ACM, New York, NY, USA, 2017, pp. 489–498.
[29] B. P. Knijnenburg, M. C. Willemsen, Evaluating recommender systems with user experiments, in: Recommender systems handbook, Springer, 2015, pp. 309–352.
[30] A. Starke, M. Willemsen, C. Snijders, Effective user interface designs to increase energy-efficient behavior in a rasch-based energy recommender system, in: Proceedings of the eleventh ACM conference on recommender systems, 2017, pp. 65–73.
[31] Y. Deldjoo, M. Schedl, B. Hidasi, Y. Wei, X. He, Multimedia recommender systems: Algorithms and challenges, in: Recommender systems handbook, Springer, 2022, pp. 973–1014.

# Manuscript VI

## Advancing Visual Food Attractiveness Predictions for Healthy Food Recommender Systems.

*Ayoub El Majjodi, Sohail Ahmed Khan, Alain D. Starke, Mehdi Elahi, Christoph Trattner.*
*HealthRecSys@ RecSys 2024,*

# Advancing Visual Food Attractiveness Predictions for Healthy Food Recommender Systems

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

Sohail Ahmed Khan
MediaFutures, University of Bergen
Bergen, Vestland, Norway
sohail.khan@uib.no

Alain D. Starke
Amsterdam School of Communication Research (ASCoR), University of Amsterdam
Amsterdam, Netherlands
MediaFutures, University of Bergen
Bergen, Vestland, Norway
a.d.starke@uva.nl

Mehdi Elahi
MediaFutures, University of Bergen
Bergen, Vestland, Norway
mehdi.elahi@uib.no

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

## ABSTRACT

The visual representation of food has a significant influence on how people choose food in the real world but also in a digital food recommender scenario. Previous studies on that matter show that small change in visual features can change human decision-making, regardless of whether the food is healthy or not. This paper reports on a study that aims to understand further how users perceive the attractiveness of food images in the digital world. In an online mixed-methods survey ($N = 192$), users provided visual attractiveness ratings on a 7-point scale and provided textual assessments of the visual attractiveness of food images. We found a robust correlation between fundamental visual features (e.g., contrast, colorfulness) and perceived image attractiveness. The analysis also revealed that cooking skills predicted food image attractiveness among user factors. Regarding food image dimensions, appearance and perceived healthiness emerged to be significantly correlated with user ratings for food image attractiveness.

## KEYWORDS

Food recommender systems, User modeling, Image attractiveness, Health, Personalization, Digital nudges

**ACM Reference Format:**
Ayoub El Majjodi, Sohail Ahmed Khan, Alain D. Starke, Mehdi Elahi, and Christoph Trattner. 2024. Advancing Visual Food Attractiveness Predictions for Healthy Food Recommender Systems. In *Proceedings of 6th Workshop on Health Recommender Systems co-located with ACM RecSys 2024, October 18th, 2024, Bari, Italy*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn


*HealthRecSys '24, October 18th, 2024, Bari, Italy*

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

Visual cues and attractiveness play a crucial role in everyday food choices [21]. Even when only presented with a food image, humans tend to instantly assess a food's energy density, expected taste and other characteristics [17]. As such, images are one of the key affective determinants of food preferences [17, 19], tapping into emotional and hedonic processes of an individual [2].

The importance of visual attractiveness also applies to digital choice context, including food recommender systems [5]. Our previous research has shown the capability of recommender systems to influence food behaviors via visual features, including the promotion of either high-fat or low-fat food choices [6], as well as encouraging the search for healthier options [19]. Additionally, our earlier work has established that visual attractiveness significantly contributes to predicting the online popularity of food items [20], and these visual features can also be leveraged to infer cultural backgrounds [23]

What is currently missing is in-depth examination of image feature modelling. Although previous studies have extracted image features and examined the relation between image features, visual attractiveness and user preferences [6, 19], these models have not been optimized. Moreover, to date, image features have not been related to user characteristics (e.g., demographics, food knowledge), which are also important determinants of food preferences [15].

We present the results of a mixed-method study that explores the determinants of visual attractiveness in digital recipe images more comprehensively. Our approach builds upon previous work by modeling perceived visual attractiveness based on low-level image features [10, 14, 19]. Additionally, we seek to optimize this model by integrating user characteristics that have been employed in knowledge-based food recommender systems to promote healthier recipe choices [4, 11, 18].

Finally, we inquire more qualitatively on user justifications for provided visual attractiveness ratings, asking to motivate their quantitative judgment. We formulate the following research questions:

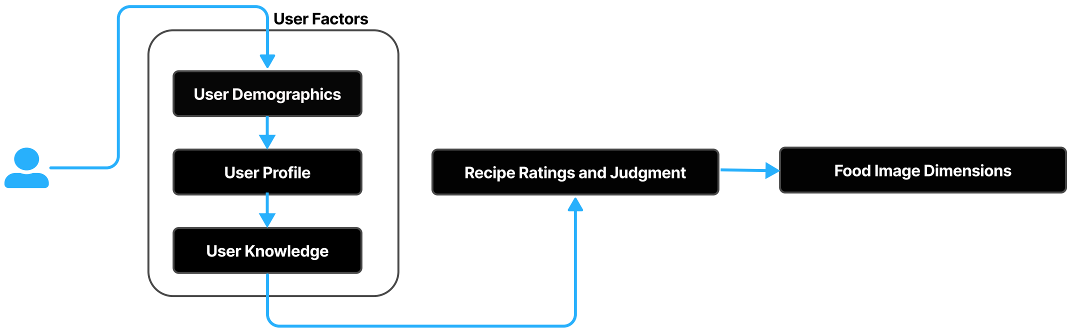



**Figure 1: Steps of the user flow designed for the online survey.**

- **RQ1**: To what extent do the latest deep learning methods predict visual attractiveness compared to state-of-the-art low-level features?
- **RQ2**: To what extent do user characteristics, including demographics, food knowledge, and eating goals, predict food image attractiveness?
- **RQ3**: What dimensions determine the attractiveness of food image?

### 1.1 Contributions

Compared to our extensive previous work in the field on visual attractiveness and food recommender systems [6, 18, 20, 23], this study offers novel insights into several key aspects:

- Previous work mostly relied on low-level image attractiveness features, while this study shows how new deep-learning models compare to these old features.
- This work, compared to any before, also shows as to what extent demographic features play a role in predicting visual food attractiveness. To our knowledge, no other work has shown this before.
- Finally, this study tries to go beyond traditional quantitative black box approaches and reveals why images are rated less or more attractive.

## 2 STUDY DESIGN

To perform our study, we employed a dataset sourced from the well-known recipes website AllRecipes.com, with the addition of new recipe photos [4, 19]. The dataset comprised various recipe features, including image URL, ingredients, amount of fats and sugar, and instructions and ingredients. To generate a diverse set of images, we randomly selected 200 recipes with relatively from the dataset of 58,000. As most images in this dataset were relatively unattractive [19], we used the recipe's title in search engines and image websites (e.g., Unsplash) to look for more attractive images for 100 of these recipes. To validate this process, three computational food researchers, including a co-author, voted on which of the two photos was the most attractive to ensure a diverse set of recipe images in terms of expected attractiveness.

The study involved a survey design, as depicted in Figure 1. Participants first provided demographic information, as well as responded to items that measured their subjective food knowledge (4 items) and cooking skills (6 items), using 5-point Likert scales based on earlier work [7, 8, 12]. We also used questions from earlier work on a knowledge-based food recommender [4], to inquire on other user characteristics, including recipe website usage and home cooking frequency, cooking experience and dietary goals. Afterwards, users were invited to rate the visual attractiveness of 12 semi-randomly selected recipe images, on 7-point attractiveness scales. In addition, to address [RQ3], they were asked to write at least one sentence about why they had given this rating. Finally, to support our examination of [RQ3], we used 5-point Likert scales on food image dimensions [24], to ask to what extent a recipe's appearance, expected taste, healthiness, and familiarity affected their attractiveness ratings.

We employed the Prolific crowdsourcing platform to recruit 192 users (65% male; $M_{age} = 33.54$) to participate in our study. The study took approximately 11 min to complete and participants were reimbursed with GBP 1.65 [1].

## 3 RESULTS

To address the research questions, we primarily employed linear regression models to understand the principal impacts of image attributes and user characteristics on image attractiveness derived from user ratings. For our thematic analysis, the images were split into attractive and unattractive based on the mid-point of the rating scale (4) ($M = 4.33$, $SD = 1.80$). Details of used materials and conducted analyses can be accessed through the following URL [1].

### 3.1 RQ1: Predicting Visual Attractiveness

We first modeled perceived visual attractiveness based on the underlying image features. We extracted diverse low-level visual features using the OpenIMAJ Java Framework (cf. [20]). Subsequently, we conducted a linear regression analysis to predict attractiveness based on these extracted visual features. The results are outlined in Table (1.A), revealing that several image features significantly affected the attractiveness of a recipe image: $F(8, 2100) = 32.66$, $p < 0.001$. Specifically, Colourfulness, Brightness, Naturalness, and Entropy demonstrated a positive association with image attractiveness. In contrast, Saturation, Sharpness, and RgbContrast negatively

[1]Our study complied with the ethical guidelines of the Research Council of Norway and the guidelines of University of Bergen for scientific research. It was judged to pass without further extensive review.

**Table 1: Linear regression models predicting visual attractiveness ratings for recipe images: (A) with low-level image visual features, (B) with deep learning-based visual features. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.**

**(A)**

| **Low-level Image Features** | $\beta$ $(S.E)$ |
|---|---|
| Colourfulness | 6.725 (1.521)$^{***}$ |
| Brightness | 2.136 (0.155)$^{***}$ |
| Naturalness | 1.925 (0.530)$^{***}$ |
| Entropy | 1.026 (0.154)$^{***}$ |
| Saturation | −3.976 (1.020)$^{***}$ |
| Sharpness | −1.182 (1.187)$^{*}$ |
| RGBContrast | −1.782 (3.808) |
| Contrast | 7.401 (11.101) |
| Constant | −6.884 (1.243)$^{***}$ |
| $R^2$ | 0.110$^{***}$ |
| RMSE | 1.753 |

**(B)**

| | **Image features Extractor** | | |
|---|---|---|---|
| | VGG16 | ResNet | Clip |
| $R^2$ | 0.351$^{***}$ | 0.349$^{***}$ | 0.357$^{***}$ |
| RMSE | 1.500 | 1.491 | 1.501 |

affected image attractiveness. In line with [19], these results suggested that users perceived colorful, bright, and naturalistic food images as more attractive.

Going beyond low-level visual image features, we used deep learning architecture models. Our toolkit included established models, such as VGG16 [16] and ResNet [9], alongside the latest in neural network architectures for visual feature extraction Clip [13]. Table (1.B) outlines the performance of these different models, outperforming our regression model in terms of $R^2$ and RMSE. This aligns with previous research where deep learning embeddings also outperformed low-level visual features within the context of food [3, 20].

## 3.2 RQ2: User characteristics and Image Attractiveness

We further examined whether user factors affected the perceived visual attractiveness of images. Accordingly, we divided user characteristics into different categories: User demographics, User profile, which represented the backbone of a food knowledge-based recommender system, and User knowledge, which measures the user's food knowledge and cooking skills. A confirmatory factor analysis, reported in Table 2, showed that both subjective food knowledge and cooking skills adhered to internal consistency guidelines ($\alpha > .70$) while they also met the guidelines for convergent validity ($AVE > 0.5$).

Table (3. A) presents the outcomes of the linear regression model aimed at forecasting the attractiveness of image recipes: $F(9, 2090) = 3.60$. Among the various user factors examined, only two significantly affected recipe attractiveness: cooking skills ($\beta = 0.34$, p-value= 0.00021) and recipe website usage ($\beta = 0.18$, p-value= 0.020). However, none of the other user aspects affected user ratings for a given image recipe. Additionally, we also analyzed a combined model of image features and user factors, but this lead to results similar to the separate models reported in Tables (1 and 3.A). This suggested that low-level visual features had a more significant impact on food image attractiveness than user features, largely in line with preliminary findings in previous research [19, 24].

## 3.3 RQ3: Justifications for Visual Attractiveness

To assess the influence of different food image dimensions on user ratings for food images, we modeled visual attractiveness based on the reported importance of food image dimensions. Table 3 outlines the results of the regression model: $F(4, 21) = 2.41$.

Two factors significantly impacted attractiveness. First, appearance had a significant impact on user ratings ($\beta = 0.12$, $p = 0.03$). Second, the expected healthiness from the images also demonstrated a significant impact ($\beta = 0.07$, $p = 0.03$). However, perceived taste and familiarity did not show an impact on user ratings.

To understand why user ratings of visual attractiveness, we examined their qualitative justifications. We employed Natural Language Processing (NLP) techniques, including punctuation, repeated character and stopword removal, to analyze 2019 user justifications, given to both attractive and unattractive images. Based on their responses, we generated a two word clouds that highlighted the most prevalent terms. Figure 2 shows the most frequent responses for both attractive and unattractive images. We discuss these, based on the themes 'appearance' and 'health' (cf. Table (3.B)).

*3.3.1 Appearance-based justifications.* Figure 3 shows a few examples. Several participants, including user ($U_a$), expressed the term 'Crispy' in their assessments of attractive images, mainly referring to appearance. The word 'Simple' is frequently used by users, such as user ($U_b$), to convey the simplicity of recipe content. In contrast, 'mess' was more commonly associated with judgments of unattractive food images, indicating their unappealing appearance. Moreover, the repeated use of the term 'fat' suggested that fatty foods were generally perceived as unattractive, as in judgments by users ($Uc - d$).

**Table 2: Results of the principal component factor analysis across different subjective food knowledge and cooking skills. Items were measured on 5-point Likert scales. Cronbach's Alpha is denoted by $\alpha$, $AVE$ is the average variance explained. Items in grey and without loading were omitted.**

| Aspect | Item | Loading |
|---|---|---|
| Subjective Food Knowledge<br>$\alpha = 0.866$<br>$AVE = 0.858$ | Compared with an average person, I know a lot about healthy eating. | 0.777 |
| | I think I know enough about healthy eating to feel pretty confident when choosing a recipe. | 0.885 |
| | I know a lot about how to evaluate the healthiness of a recipe. | 0.773 |
| | I do not feel very knowledgeable about healthy eating. | 0.932 |
| Cooking skills<br>$\alpha = 0.783$<br>AVE = 0.591 | I can confidently cook recipes with basic ingredients. | 0.751 |
| | I can confidently follow all the steps of simple recipes. | |
| | I can confidently taste new foods. | 0.737 |
| | I can confidently cook new foods and try new recipes. | 0.869 |
| | I enjoy cooking food. | 0.655 |
| | I am satisfied with my cooking skills. | 0.816 |

**Table 3: Linear regression models predicting user rating for recipe image attractiveness: (A): with user factors, (B): with food image dimensions. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.**

**(A)**

| **User Factors** | $\beta$ (S.E) |
|---|---|
| **User Demographic** | |
| Age | −0.047 (0.116) |
| Education | −0.424 (0.320) |
| Gender | −0.077 (0.088) |
| **User Profile** | |
| Recipe Website Usage | $0.201 (0.086)^{*}$ |
| Home Cooking | −0.009 (0.078) |
| Cooking Experience | −0.052 (0.079) |
| Eating Goals | 0.019 (0.063) |
| **User Knowledge** | |
| Subjective Food Knowledge | −0.213 (0.138) |
| Cooking Skills | $0.315 (0.086)^{***}$ |
| Constant | $4.001 (0.570)^{***}$ |
| $R^2$ | $0.015^{***}$ |
| RMSE | 1.845 |

**(B)**

| **Food Image Dimension** | $\beta$(S.E) |
|---|---|
| Appearance | $0.129 (0.061)^{*}$ |
| Healthiness | $0.077 (0.035)^{*}$ |
| Taste | −0.005 (0.050) |
| Familiarity | 0.0231 (0.038) |
| Constant | $3.487 (0.365)^{***}$ |
| $R^2$ | $0.011^{***}$ |
| RMSE | 1.855 |

*3.3.2 Healthiness-based justifications.* Judgments related to health frequently appeared in connection with the food's appearance, such as by user ($U_e$) in Figure 4. The term 'restaurant' was employed in various user judgments, often associated with presentation and healthiness, as described by the user ($U_f$). Conversely, the concept of unhealthiness was linked to fatty foods and messy representation, as evident in the judgments of users ($U_{g-h}$) in Figure 4.

## 4 CONCLUSION & FUTURE WORK

This work has explored different aspects of the relationship between the user and food images. Through an online user study, we have found that various visual features can predict the attractiveness of a given image (i.e. colorfulness, brightness, naturalness). This prediction accuracy could be slightly improved using image features extracted using deep learning techniques (RQ1). In line with earlier work [14, 19, 24], this suggests that the visual attractiveness of food images can be enhanced by increasing their colorfulness, brightness, and naturalness, while decreasing other features, such as saturating and sharpness. Obviously, there may be tradeoffs between these features when altering them.

Regarding user characteristics, none of the user demographics are related to food image attractiveness. In contrast, using online

(A) (B)


**Figure 2: Word cloud for terms in the user judgment: (A) : judgments for attractive images, (B): judgments for unattractive images.**

**(U$_a$):** *"looks juicy with nice crispy bits, which is nice and clear in the picture"*

**(U$_b$):** *"Interesting, slightly unusual, and does look visually appealing with simple ingredients presented well"*

**(U$_c$):** *"It looks messy and unappealing"*

**(U$_c$):** *"Too much carbs/fat"*

**Figure 3: Example of images used in the study, associated with users' textual judgment related to the appearance. (U$_{a-b}$) were textual justifications for attractive images, (U$_{c-d}$) were textual justifications for unattractive images.**

**(U$_e$):** *"Healthy salad option with balanced nutrients. It's is also quite colorful"*

**(U$_f$):** *"The dish looks very nice, like in a restaurant. It is colorful and looks very healthy"*

**(U$_g$):** *"It looks a bit mushy and brown and I don't like Turkey"*

**(U$_h$):** *"Chicken is unhealthy and gross"*

**Figure 4: Example images used in the study, associated with users' textual justifications related to the healthiness. (U$_{e-f}$) were for attractive images, (U$_{g-h}$) for unattractive images.**

recipe websites and cooking skills are positively associated with the attractiveness of food images (RQ2). More novel is our contribution on the user justifications, for which we have found image appearance and perceived healthiness to be important dimensions of visual attractiveness ratings (RQ3). It seems that attractiveness are related to the expect taste or hedonic food goals (e.g., 'crispy'), while unattractive images focused on poor presentation and disliked ingredients.

Our study offers valuable insights into techniques for image attractiveness selection for various goals and domains. In particular, these techniques can be leveraged to persuade or nudge users towards specific eating goals, such as health [19, 22]. We believe that leveraging the visual appeal of attractive images can address this issue. Our future studies will focus on designing image selection pipelines for the application of food recommender systems tailored to guide people toward healthy food choices without compromising the benefits of personalization. We aim to analyze and categorize the collected textual judgment through thematic analysis to build word dictionaries related to image dimensions. These dictionaries can then be used to train learning models, enabling the evaluation of food image attractiveness based on user textual inputs.

## ACKNOWLEDGMENTS

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

## REFERENCES

[1] 2024. Examining the Visual Attractiveness of Digital Recipe Images: Material. https://github.com/ayoubGL/Health-RecSys-2024

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

[3] Jing-jing Chen, Chong-Wah Ngo, and Tat-Seng Chua. 2017. Cross-modal recipe retrieval with rich food attributes. In *Proceedings of the 25th ACM international conference on Multimedia.* 1771–1779.

[4] Ayoub El Majjodi, Alain D Starke, Mehdi Elahi, Christoph Trattner, et al. 2023. The Interplay between Food Knowledge, Nudges, and Preference Elicitation Methods Determines the Evaluation of a Recipe Recommender System. In *Proceedings of the 10th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS 2023).* 1–18.

[5] David Elsweiler, Hanna Hauptmann, and Christoph Trattner. 2022. *Food Recommender Food recommenderSystems.* Springer US, New York, NY, 871–925. https://doi.org/10.1007/978-1-0716-2197-4_23

[6] David Elsweiler, Christoph Trattner, and Morgan Harvey. 2017. Exploiting food choice biases for healthier recipe recommendation. In *Proceedings of the 40th international acm sigir conference on research and development in information retrieval.* 575–584.

[7] Leisa Reinecke Flynn and Ronald E Goldsmith. 1999. A short, reliable measure of subjective knowledge. *Journal of business research* 46, 1 (1999), 57–66.

[8] Nike Frans. 2017. *Development of cooking skills questionnaire for EFNEP participants in Kansas.* Ph. D. Dissertation. Kansas State University.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 770–778.

[10] Aditya Khosla, Atish Das Sarma, and Raffay Hamid. 2014. What makes an image popular?. In *Proceedings of the 23rd international conference on World wide web.* 867–876.

[11] Cataldo Musto, Christoph Trattner, Alain Starke, and Giovanni Semeraro. 2020. Towards a knowledge-aware food recommender system exploiting holistic user models. In *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization.* ACM, New York, NY, USA, 333–337.

[12] Zuzanna Pieniak, Joris Aertsens, and Wim Verbeke. 2010. Subjective and objective knowledge as determinants of organic vegetables consumption. *Food quality and preference* 21, 6 (2010), 581–588.

[13] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning.* PMLR, 8748–8763.

[14] Jose San Pedro and Stefan Siersdorfer. 2009. Ranking and classifying attractiveness of photos in folksonomies. In *Proceedings of the 18th international conference on World wide web.* 771–780.

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

[16] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[17] Charles Spence, Kosuke Motoki, and Olivia Petit. 2022. Factors influencing the visual deliciousness/eye-appeal of food. *Food Quality and Preference* 102 (2022), 104672.

[18] Alain D Starke, Cataldo Musto, Amon Rapp, Giovanni Semeraro, and Christoph Trattner. 2023. "Tell Me Why": using natural language justifications in a recipe recommender system to support healthier food choices. *User Modeling and User-Adapted Interaction* (2023), 1–34.

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

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

[21] Iris Vermeir and Gudrun Roose. 2020. Visual design cues impacting food choice: A review and future research agenda. *Foods* 9, 10 (2020), 1495.

[22] Longqi Yang, Cheng-Kang Hsieh, Hongjian Yang, John P Pollak, Nicola Dell, Serge Belongie, Curtis Cole, and Deborah Estrin. 2017. Yum-me: a personalized nutrient-based meal recommender system. *ACM Transactions on Information Systems (TOIS)* 36, 1 (2017), 1–31.

[23] Qing Zhang, David Elsweiler, and Christoph Trattner. 2020. Visual cultural biases in food classification. *Foods* 9, 6 (2020), 823.

[24] Qing Zhang, David Elsweiler, and Christoph Trattner. 2023. Understanding and predicting cross-cultural food preferences with online recipe images. *Information Processing & Management* 60, 5 (2023), 103443.