



Digitally nudging users to explore off-profile recommendations: here be dragons

Gabrielle Alves¹ · Dietmar Jannach² · Rodrigo Ferrari de Souza¹ · Daniela Damian³ · Marcelo Garcia Manzato¹

Received: 19 July 2022 / Accepted in revised form: 21 July 2023 / Published online: 4 October 2023
© The Author(s) 2023

Abstract

In many application domains of recommender systems, e.g., on media streaming sites, one main goal of the provider of the recommendation service is to increase the engagement of users by helping them *discover* new types of content they like. Standard collaborative filtering algorithms by design often lead to a certain level of discovery. Nonetheless, in certain domains, it may be helpful to more actively promote content to users beyond their past preference profile (“off-profile”) and thereby help users explore new content. However, when showing such off-profile content to users in combination with more familiar content, the new content items may be overlooked. In this research, we explore to what extent *digital nudging*, i.e., subtly directing user choices in a specific direction, can help to raise the attention and interest of users for off-profile content. We conducted a user study ($N = 1064$) on a real-world social book recommendation app. We find that users who are nudged towards recommended books of their non-preferred genres significantly more often put these off-profile books on their reading lists, thus confirming the effectiveness of digital nudging in this application. However, we also found that digital nudges may negatively impact the users’ beliefs and attitudes towards the system and a more limited intention to use the system in the future. As a result, we find that digital nudging in recommendations, while effective in the short run, must be done with due care, keeping an eye on the overall quality perceptions by users and potentially harmful long-term effects.

Keywords Recommendation · Digital nudging · Discovery · Diversity · Novelty

✉ Dietmar Jannach
dietmar.jannach@aau.at

¹ Universidade de São Paulo, São Paulo, Brazil

² University of Klagenfurt, Klagenfurt, Austria

³ University of Victoria, Victoria, Canada

1 Introduction

Nowadays, recommender systems have been adopted by many extensive online services—including Amazon, Netflix, Spotify, or YouTube—where they are used to provide personalized item suggestions to users. Depending on the goals of an organization, these systems may serve various purposes (Jannach and Adomavicius 2016). The most common goal is to help users find content that matches their past preferences, e.g., movies of genres the user had preferred in the past. Often, however, an additional goal is to help users to *discover* new things, e.g., by suggesting items outside their historical tastes. On movie and music streaming websites, for example, such discovery support ultimately aims at increasing the users' engagement and joy, which generally is assumed to lead to higher retention, see Gomez-Uribe and Hunt (2015), Hansen et al. (2021) for related discussions for the cases of Netflix and Spotify.

While questions of discovery (or: exploration) are essential in practice, there is comparably little research in the academic literature that explicitly aims at understanding how discovery support affects the user experience and quality perception of users. An example of an early work that studies discovery in the music domain through a user study is Celma (2020). In a more recent work (Ludewig et al. 2021), discovery support was considered as one of several quality factors of a music recommender. More frequently, questions of discovery are studied through data analyses and offline experiments, e.g., Hansen et al. (2021), Kapoor et al. (2015). In addition, a more significant number of works exist that study related questions that *can be* related to discovery, i.e., questions regarding the role of diversity and novelty in recommender systems, see Kaminskas and Bridge (2016), Kunaver and Požrl (2017) for respective surveys. Again, however, most of these studies are based on offline experimentation, using, for example, the intra-list similarity (ILS) metric (Bradley et al. 2001; Ziegler et al. 2005) as a proxy for the users' diversity perception. However, to what extent metrics like ILS are actually good proxies for user perceptions is not always clear.

One question that one may ask is to what extent the inclusion of what we call “off-profile” items—i.e., items beyond a user's past tastes and thus aimed to support discovery—would impact users' actual quality perception and behavior. Ekstrand et al. (2014), for example, observed that including surprising or unexpected items may negatively impact the users' satisfaction with the recommendations. In Ge et al. (2011, 2012), on the other hand, the authors found indications that even adding unsuitable *diverse* items—in their case, they included comedies in a list of recommended action movies—may almost go unnoticed by users, depending on where these diverse items were positioned in the list.

Given in particular this latter observation, i.e., that users' might not even notice off-profile items or consider them irrelevant, we are interested in understanding the effects of subtly guiding users to these items. Specifically, we aim to examine the effects of *digitally nudging* (Caraban et al. 2019; Schneider et al. 2018) users to consider off-profile items in terms of (i) their actual choice behavior and (ii) their quality perception of the provided recommendations. Nudging, as defined in Thaler and Sunstein (2008), can be seen as a benevolent intervention aimed to direct user choices in a certain direction without changing the “choice architecture”, e.g., without changing the set of options.

In this work, we investigate the described questions in the context of book recommendations. Specifically, we conducted an online study involving 1064 users of a real-world social book recommendation site. In the study, participants were shown recommendation lists that contained both books that matched their preferred genres and off-profile recommendations from other genres. For the participants in the treatment group, we applied different types of nudges to the off-profile items to help them grow their spectrum of interest and discover books in genres that they did not report as preferred before the study. We emphasize that different from prior work, e.g., Berger et al. (2020), Jesse et al. (2021), our goal is not to find out which type of nudge works best in a given domain, but to analyze the combined effects of the applied nudges on user behavior.

Our study results suggest that the nudges were highly effective, and participants that were nudged placed significantly more off-profile items on their reading lists. However, we also found that some nudged participants were less satisfied with the items on their reading list, even though the options were identical in the control and treatment groups. Our study, therefore, points to interesting psychological phenomena that arise when off-profile items receive more attention through the nudges. As a result, we conclude that while nudging showed to be effective in this domain, at least in the short-term, we also find that digital nudges must be designed with due care in order to avoid potential negative long-term effects and a lower intention to rely on the recommendations in the future.

The paper is organized as follows. We discuss earlier work next in Sect. 2. The details of the conducted study are provided in Sect. 3. In Sect. 4, we analyze and interpret the obtained results. The paper ends with a discussion of research limitations and future works.

2 Related work

In this section, we first discuss earlier works on discovery and exploration in recommender systems (Sect. 2.1) and then review previous papers which investigate digital nudging in combination with recommendations (Sect. 2.2).

2.1 Discovery and exploration with recommender systems

Helping users discover relevant items previously unknown to them is central to most recommender systems applications.¹ There are differences, though, to what extent different algorithms are able to surface relevant content outside a user's past preference profile. Content-based filtering systems (Lops et al. 2011), for example, by design recommend items that are similar to the user's past profile. Collaborative filtering, in contrast, has more potential for the discovery of items outside a user's familiar taste profile—which is the focus of our present work—by relying on preference patterns

¹ Exceptions however exist, for example, in cases when the system recommends repeated purchases of consumables, recommends playing the user's favorite music, or reminds users of previously inspected items (Lerche et al. 2016).

in the user community. In Dias et al. (2008) and Lawrence et al. (2001), for example, it was found that recommender systems, when deployed in online supermarkets, can have indirect, inspirational effects and guide users to categories in which they had not made purchases before. Similar observations were made in Kamehkhosh et al. (2019) in the music domain. Finally, reinforcement learning-based approaches implement explore-exploit schemes to gather more information about users' preferences. Such recommender systems, therefore, sometimes purposefully recommend items for which they are uncertain about the suitability for the users, see, e.g. McNerney et al. (2018).

Independent of the choice of the algorithm, it has been recognized long ago that solely recommending items with high predicted relevance may not be enough (McNee et al. 2006). Correspondingly, various approaches have been proposed in the last years that aim to increase the diversity, novelty, and serendipity of the recommendations that are served to users. In a recent large-scale user study, Chen et al. (2019) investigated the perception of novel, serendipitous and diverse recommendations of over 3,000 customers of a large Chinese mobile e-commerce platform. The authors in particular found that *serendipity* significantly influenced user satisfaction and purchase intent, and that the effects of serendipity are larger than those that come from novelty, and more direct than diversity effects. Our present work is similar to theirs in that we study human perceptions by involving users of a real-world online service and that we consider off-profile recommendations. However, in our work, we explore the effects of nudging users to such off-profile items.

We note that increasing serendipity, diversity and novelty may *in principle* help users to explore off-profile items and, thereby, discover relevant novel content. To what extent a diversity-aware or novelty-aware recommender system supports this goal, however, depends on the particular way the concepts of diversity and novelty are operationalized, i.e., it may depend on the specific choice of a particular evaluation metric or optimization goal. Including items of different genres in the recommendations may, for example, help to increase the metrics such as intra-list similarity (Ziegler et al. 2005). However, it does not guarantee that these genres are new to users (off-profile). Likewise, in some works, the novelty of an item is defined in terms of its general popularity. The discovery of off-profile content is thus guaranteed in such an approach and would only be achieved when novelty is determined in terms of an item's distance to a user's past profile, see Castells et al. (2011) for a related discussion on beyond-accuracy metrics. Similar considerations apply to the concepts of serendipity or unexpectedness, which are often also difficult to operationalize in computational metrics.

There are, however, a number of research works that explicitly study questions of exploration and discovery. In Kapoor et al. (2015), the authors investigate how users' openness or willingness to explore new musical content changes over time. Data-based analyses confirm that such dynamics exist, and the authors then develop a method to predict the variable novelty preferences. Different from our work, the authors consider novelty at the item level, i.e., they analyze if users have listened to a track before within a certain time window and consider an item to be novel if it has not appeared in such a listening history. In our work, in contrast, we explore whether users can be nudged towards items that are not part of their declared preferred genres. Also, differently from Kapoor et al. (2015), we investigate our research questions through a user study.

Questions of discovery in the music domain were also addressed as one aspect in the user study in Ludewig et al. (2021). In their work, the authors compared several recommendation algorithms for the problem of creating dynamic playlists. Among other aspects, it was found that in particular the recommendations provided by the commercial platform Spotify excelled compared to all others in terms of helping users discover some unknown tracks that they liked. Interestingly, while the offline accuracy of Spotify's recommendation was very low, this did not lead to a significantly lower self-declared intent by the study participants to use the system again or recommend it to others. This can thus be seen as an indication of the importance and value of discovery support in this application domain.

The importance of these topics for the business success of a practical application is also highlighted in a recent work by researchers of Spotify, where the explicit goal is to shift users' consumption towards less popular content and to content that is different from their historic tastes (Hansen et al. 2021). In their approach, the distance of a given track to a user profile is based on comparing track and profile embeddings. Generally, the authors observe a trade-off between higher diversity—which in their work means both recommending less popular items and more unfamiliar items—and user satisfaction. Their computational analyses aim to identify suitable algorithmic approaches to balance this trade-off. Differently from our work, Hansen et al. (2021) rely solely on offline experiments. As a proxy for user satisfaction, they rely on historical listening logs and consider a user dissatisfied if they skipped a recommended track. Similar to our work, however, they also explore a diversification strategy where tracks with high predicted relevance are interleaved with tracks of high diversity.

In some application domains of recommender systems, diversification in terms of recommending content beyond the users' typical preferences may also have a societal impact. In Heitz et al. (2022), the authors developed a mobile news reader app and developed a customizable algorithm that can deliver news that is diversified in terms of their political orientation (left/right-leaning, liberal/conservative). The political orientation of users was assessed through an existing political scoring instrument, and the orientation of articles was based on the orientation of readers who liked these articles. Different strategies were then studied in a field test, among them rather diversified recommendations or recommendations that rather narrowly followed the participants' past orientation. The obtained results showed in many dimensions no significant differences, e.g., for app usage or the perceived utility, indicating that diversification did not harm the user experience. Based on additional analyses, the authors ultimately concluded that their *“results suggest that recommendations enhance tolerance for opposing views”* and that *“diverse news recommendations may entail a de-polarizing capacity for democratic societies”*. Our present work is similar to Heitz et al. (2022) in that we aim to guide users outside what is often called a *filter bubble* in news recommendations or social media. Like in their work, our work is also based on a mobile app and real users. However, unlike their work, we actively try to nudge users towards off-profile items.

2.2 Digital nudging and recommender systems

Nudge theory was popularized in 2008 through an influential book by Thaler and Sunstein (2008), where nudging is seen as a mechanism to influence people's behavior positively, without coercion. A typical example of applying nudging in the real world includes the more prominent positioning of healthy food in a cafeteria. Nudging can however also be applied in the online world, where it is often referred to as digital nudging (Caraban et al. 2019; Schneider et al. 2018). A typical digital nudge in the online world is the pre-selection of a default, e.g., when users can choose between different subscription plans for a service.

In Caraban et al. (2019), provided a review on digital nudging techniques in human-computer interaction research. The authors identified 23 ways of nudging, which they organized into six groups. Interestingly, the authors also included nudge mechanisms that can be used in a non-benevolent way and deceive users. This stands to some extent in contrast to the original conception of nudges, which were commonly designed for benevolent purposes. Some of the nudges identified by Caraban et al. could therefore rather be viewed as covert *persuasion*. Persuasion techniques (Fogg 2002) are related to nudging and are often also based on certain psychological phenomena that can be used to influence the behavior of humans. Mols et al. (2015) also see nudges as related yet different concepts. They consider nudges to be effective when targeting to influence people's behavior in a particular setting. In contrast, persuasion targets people's underlying beliefs and preferences and should lead to a more sustained behavior change.

Jesse and Jannach (2021) reviewed the topic of nudging in combination with recommender systems. They identified 87 nudging mechanisms in the literature and listed a set of psychological phenomena that are commonly used to explain why nudges work. In their work, the authors argue that any recommender system, in some ways, can be seen as a tool that nudges users because recommender systems implement mechanisms that are common in digital nudging. These mechanisms include that the salience of some items is increased, that positioning effects are used, or that the ease and convenience of accessing certain items is higher than for others. The terminology in the literature is however not always consistent. Sometimes recommendations are considered persuasive even in cases when they are not targeting the underlying beliefs of users, e.g., in Köcher et al. (2019), where the authors consider recommendations to be "hidden persuaders". For an in-depth treatment on persuasion and recommender systems, we refer readers to Yoo et al. (2013).

In our present work, we combine recommendations with nudging, in our case by applying nudges to certain items in a recommendation list. Several earlier works investigated nudging in conjunction with search and recommender systems. The food (recipe) domain was the focus of earlier research in Elsweiler et al. (2017) or Starke et al. (2021). In Elsweiler et al. (2017), the authors showed through a user study that a suitable selection of food images can nudge users to choose healthier recipes. In Starke et al. (2021), a related study on choices in a recipe search scenario showed that participants tended to pick healthier options when these have visually attractive images attached. Another study in the food domain can be found in Jesse et al. (2021).

In their work, the authors examined the effectiveness of different types of nudges. They found that a hybrid nudge that combined two types of mechanisms, setting a default and providing a social cue, was the most effective one to steer food choices without negatively impacting the choice satisfaction of the user.

The focus of the work in Starke et al. (2020) was in the energy domain. Here, the authors studied the effect of a “social norm” nudge in a recommender system that suggests energy-saving measures. Somewhat mixed results were obtained, and not all hypotheses were confirmed, e.g., that the number of selected measures can be increased with nudging. However, some critical observations were made in the study, e.g., that the perceived feasibility of a measure, i.e., how difficult it would be to implement it, has a mediating influence on the effect of the nudge.

In the music domain, Liang et al. (2021) conducted an exploratory study to determine whether the use of digital nudging can encourage users to explore genres that are distant from their current preferences. They found that users are more likely to select such distant genres if they appear at the top of the list, which can be explained by order effects and the corresponding position bias. In contrast to Liang et al. (2021), we interspersed off-profile items in the entire list with items from the user preferences in the control and treatment groups. Moreover, in our study we applied additional nudges in the treatment group to encourage users to interact with off-profile items.

Vermeulen (2022) recently discusses the potential role of recommender systems as a means to achieve *public policy goals*. According to Article 10 of the European Convention on Human Rights, governments are required to ensure that citizens have access to diverse news sources and information via audiovisual media. Such requirements have however not yet been established for digital media. Vermeulen noted that recommender systems and digital nudging could be a promising tool to promote online access diversity and reduce filter bubbles and echo chambers. These latter effects are often a result of what is called a *confirmation bias*, where (online) users tend to mainly consume content that is aligned with their prior beliefs. In this context, Rieger et al. (2021) recently explored how digital nudging could mitigate confirmation bias and help users to make more informed decisions. Our study is related to these works as we aim to help users discover relevant content outside their previous preferences, but in a very different domain.

Other areas where recommender systems and nudging or persuasive techniques were combined include the recommendation of environment-friendly routes (Bothos et al. 2016), news recommendation (Gena et al. 2019), or the recommendation of followers on social media (Verma et al. 2018), see also Jesse and Jannach (2021). To our knowledge, nudging has not been applied previously in the context of book recommendation, which is the target application in our present work.

3 Study design

In this section we first provide an overview of our study in Sect. 3.1 and then provide details of the experiment in Sect. 3.2.

3.1 Study overview

Our research aims to examine the effects of actively directing users towards off-profile recommendations. Specifically, in our study, we investigate the effectiveness of different digital nudging mechanisms in terms of influencing users to actually consider such off-profile recommendations. Moreover, we aim to study the effects of the applied nudges on the users' quality perception of the recommendations and on the users' future behavioral intentions.

As a target domain for our work, we consider the domain of book recommendations, and we study the problem with the help of users of a real-world Brazilian social book recommendation site. One primary mission of this book recommendation site is to increase the literacy level in society. Overall, with this work, we aim to create new insights regarding effective means to diversify users' reading interests and engage them more in reading and increase their literacy.

Our study implements a between-subjects design. Participants both in the treatment and control groups were using a mobile book recommendation app that was developed for the purpose of the study. Participants in both groups were then presented with recommendations that contained both books from their previous preferred genres and off-profile books. For the participants in the treatment group, the user interface was enhanced with different types of digital nudges attached to each off-profile book. The users' task was to select books to add to their reading lists. To understand the effectiveness of the nudges, we compared both the participants' objectively observed behavior and their subjective statements regarding the quality of the recommendations and the choice process.

Let us emphasize here that we relied on an own app for the study for two main reasons. First, from a user experience perspective, one main goal was to serve the target users in the most convenient way through a mobile app. Second, while we aimed to provide a realistic user experience, it was important to us that the participants were aware that they interacted with the app for the purpose of a study.²

3.2 Experiment details

Experiment Flow Fig. 1 illustrates the overall flow of the experiment. The study participants, who were recruited through postings on social media, first downloaded the mobile app and were then informed about the tasks in the study. In the cover story, participants were told that the study was designed to develop a better recommendation system for book lovers and people seeking to establish a reading habit. In order to participate in the study, users could download the app from the Google Play App Store. In the app, they were then able to create a profile and specify their favorite genres. The participants were encouraged to choose between three and five books recommended by the app to put on their reading list. They were however not required to select a specific number of books. Participants were then randomly assigned to either the treatment or the control group.

² This was declared when the subjects agreed to participate, and the subjects thus were made aware that they took part in an experiment in which different forms of recommendations were explored.



Fig. 1 Experiment flow

After providing informed consent, the participants were then presented with a list of the six book genres, and they were asked to select their preferred genres. Then, a list of recommended books was shown to the participants, where every second was from a non-preferred genre. Participants could then add as many books as they wanted to their reading list. After completing their selection, participants were asked by email to answer a number of questions in a post-task questionnaire. To entice users to complete the questionnaire, every participant that completed the post-task survey entered a lottery draw with prizes consisting of bookmarks and books.

Preference Elicitation & Recommendation List Design The six genres (or: categories) for which participants could express their preferences were *Fantasy*, *Horror*, *Nonfiction*, *Romance*, *Suspense*, and *Young Adult*. These categories were the six most popular ones on Amazon.com.br in the books section at the time of the study. We selected popular categories to increase the chances that our study participants will prefer at least some of them.

The recommended books in each category were selected manually based on various factors. We both considered books based on average reader ratings and included new releases from independent publishers and debut authors. This was done to make sure that the recommender can support the discovery of new books and that it is not limited to bestsellers. The exact list of recommended books can be found in the “Appendix”.

Both the participants in the treatment group and in the control group were recommended 24 books. Half of the books were chosen from a category that the participant had indicated to be a preferred one in the previous step. The order of the books was randomized once and kept static across users. The other half was chosen from non-preferred categories.

In order to maintain a consistent experience for all users, the interleaving rule and the number of books on the list remain static. However, the recommended books are based on the genres selected by the user. For example, the user selects four genres, there are only two non-preferred genres, and the list interleaves twelve random books from the four genres on the preferred list with twelve random books from the two non-preferred genres.

In case a participant has selected all six categories as preferred ones in the previous step, we did not consider this participant in our analysis, because such users were only shown books from their preferred genres.

To avoid order effects and presentation biases, we adopted an *interleaving* approach that was used earlier, e.g., in Joachims (2003), for the problem of *click-through* analysis in information retrieval settings. It is well known from earlier research that recommen-

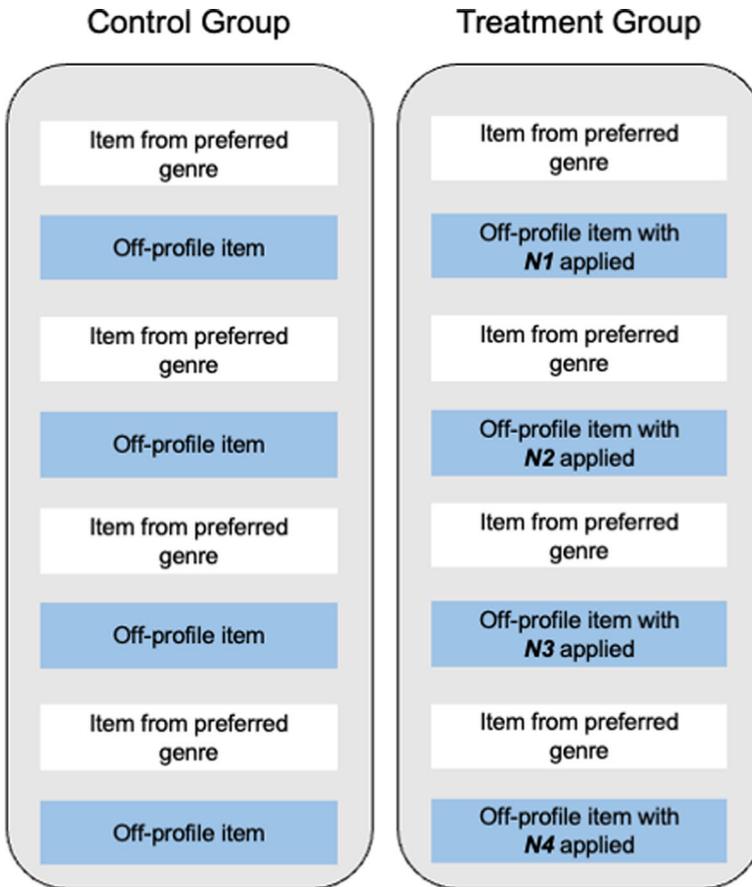


Fig. 2 General Structure of Interleaved Recommendation Lists (only showing the first four of 24 items)

dations or search results that appear higher up in a list have a higher chance of being seen and inspected by users. Thus, a click event on such items may not be an absolutely reliable indicator of relevance (Baeza-Yates et al. 2011; Oosterhuis 2020). Comparing two separate ranked lists can therefore be difficult. To avoid this problem and get more unbiased users' feedback, we show participants recommendation lists where items of preferred genres are interleaved with off-profile items, i.e., every second item in the list is an off-profile item.

Figure 2 illustrates this interleaving approach. We iterate that this interleaving is done both for participants in the control and participants in the treatment group. The difference between the two groups is that we attached a digital nudge to each off-profile item in the treatment group.

Selected Nudges We considered four types of digital nudges from the literature, which are nowadays also widely used in practical applications. Given that our experiments took place in the environment of a social book recommendation site, we specifically focused on *social nudges* (Jesse and Jannach 2021), i.e., nudges that are based on

social psychology theories, such as social comparison and conformity (Gena et al. 2019). According to those theories, people who are unsure about how to behave in a situation actively search for information about what others did when they were in a similar situation. This information then shapes their behavior and attitudes. The specifics of each nudge are as follows.

- *Nudge N1 “Hybrid: Following the herd and Increase Salience”*: An example of this nudge, as found in the literature (Thaler and Sunstein 2008; Mirsch et al. 2017), is to show users a message that indicates some other users have favorited or selected a particular item. The underlying hypothesis is that since most users do not want to stand out from the crowd, they are more inclined to follow the trend of the majority, and this nudge is based on this behavior. In our specific implementation, the nudge was visually highlighted and indicated that many other people currently put this book on their reading list, e.g., “*Favorited 36 times today*”.³ We selected this nudge as it is commonly used in practical applications, e.g., on flight reservation websites.
- *Nudge N2 “Hybrid: Increase salience of attribute and Argumentum-Ad Populum”*: This nudge increases the visual salience of the nudged item to increase the attention they receive. Moreover, we display a message based on the hypothesis that a participant may have a higher inclination to accept an opinion when the majority agrees with it, e.g. by stating that “*90% liked this book*”.⁴ Among other reasons, we selected this hybrid nudge because hybrids were found to be particularly effective in a recent study in the food domain (Jesse et al. 2021). Presenting the number of *likes* to users is common in practice, e.g., on social media.
- *Nudge N3 “Hybrid: Increase salience of attribute and Messenger Effect”*: This nudge is a variation of N2. This time, instead of emphasizing how many other users liked a book, we presented messages such as “*Netflix will adapt this book*”, “*Bestseller of the week on Amazon*”. Such labels are also common in the off-line world, e.g., “*New York Times Bestseller*”.
- *Nudge N4 “Social reference point”*. This nudge is referring to the assessment of a book by opinion leaders. The underlying assumption is that users may often trust the opinions of prominent and influential other people. In our study, the opinion leaders that we cited were: George R R Martin, Neil Gaiman, Obama, Stephen King, Lupita Nyong’o, Reese Witherspoon, Emma Watson, Dakota Fanning, Mark Zuckerberg, and Bill Gates. We created this list of these opinion leaders in a way that participants with different backgrounds and demographics should be impacted. The message displayed in one nudge for example was “*Recommended by Stephen King*”. We note that providing endorsement statements by opinion leaders or celebrities is a widely used instrument in traditional marketing.

The order in which the four nudges were placed was randomly determined before the experiment and kept static throughout the study. Since each recommendation list contained 12 items with nudges in the treatment group, the static order was repeatedly applied after the fourth and the eighth off-profile item. We note that all messages in the

³ We note that the entire app was in Portuguese; the examples provided here are translated to English.

⁴ The shown numbers were slightly varied for higher realism, see also Jesse et al. (2021).

Table 1 Question items in the post-task questionnaire

ID	Construct	Question
Q1	Accuracy	The items recommended to me matched my interests
Q2	Attractiveness	The app provided some really good recommendations
Q3	Diversity	The items recommended to me are diverse
Q4	Novelty	The recommendations helped me discover new books which I find interesting
Q5	Serendipity	I was surprised by some of the recommendations
Q6	Information Sufficiency	The amount of information available was sufficient to make the decisions
Q7	Transparency	I understood why the items were recommended to me
Q8	Usefulness	Finding books to put on my reading list was easy
Q9	Choice Satisfaction	My overall satisfaction with my selection of books on my reading list is high
Q10	General Satisfaction	My overall satisfaction with the app is high
Q11	Intention to use	I like the app, and I would use it again in the future
Q12	Intention to “purchase”	I am eager to read the book(s) I have put on the reading list

nudge statements regarding the ratings and the opinion leaders were real, except for nudge N1, where the number of people who favorited an item recently was invented.

We iterate that we decided to consider nudges from the “Social Decision Appeal” category from Jesse and Jannach (2021) based on our specific application setting. In three of the cases, we furthermore additionally increased the salience of the information. In these cases, we implemented a hybrid nudging approach, which has been shown to be effective in the past in Jesse et al. (2021), Renaud and Zimmermann (2019). Several other types of nudges from the literature could be applied as well. In our present work, we however limited ourselves to one category of nudges, because a comparison of the effectiveness of different types of nudges—as done, e.g., in Berger et al. (2020), Jesse and Jannach (2021)—was not the focus of our research.

In terms of the general categorization of nudges (e.g., “Social Decision Appeal” in Jesse and Jannach (2021)), we note that the border between the categories is not always sharply defined in the literature. In addition, as observed also in Jesse and Jannach (2021), there can be more than one psychological phenomenon that is related to a certain type of nudge. In the context of our study, some of the nudges are related in certain ways. Nudge N3 and N4, for example, both refer to an authority in some sense. Overall, however, all four nudges in our study share that they have some social aspect.

Objective Measurements of User Behavior To understand in which ways the nudges affect users’ choice behavior, we first of all recorded which and how many books the participants put on their reading lists. Moreover, we logged the positions of the selected books and if they are from a preferred category or if they are off-profile items.

To understand the participants' behavior in more detail, we also recorded how many books the participants inspected their details (by clicking on a "show more" button for a book), and we tracked how far participants scrolled down the list. Finally, we measured how long participants took to make a selection.

Assessing Subjective User Perceptions To investigate our second research question on the impact of the nudges on users' perceptions, we asked participants some related questions in the post-task questionnaire. Our questionnaire is based on the user-centric recommender systems evaluation framework ResQue proposed in Pu et al. (2011). The ResQue framework is a validated and widely used instrument to assess human perceptions of recommender systems at different levels. It first connects *user-perceived quality factors* (first level) such as the accuracy or the diversity of the recommendations, with *user beliefs* (second level), e.g., about the system's perceived usefulness, transparency or ease-of-use. These beliefs may then impact the *user attitudes* (third level) in terms of user satisfaction and trust. Finally, these factors may ultimately affect the *behavioral intentions* (fourth level) of users in terms of intended purchases or intended use.

The ResQue framework also provides a validated questionnaire, which we used as a basis for our study. However, since our study was conducted on a mobile setting with comparably small screen sizes and high user expectations regarding the ease-of-use of the app, we decided only to use a subset of the original questionnaire. The specific questions for the different dimensions are shown in Table 1.⁵ Participants could answer on a five-point Likert scale from 1 (completely disagree) to 5 (entirely agree). We note that the constructs shown in the table are measured with only one questionnaire item in order to not overwhelm participants in the mobile app.⁶

In addition to the questionnaire items focusing on the users' perceptions, we included a number of questionnaire items regarding the demographics of the participants and their general interest in reading books.

Technical Implementation Fig. 3 shows the main screen of the application after they provided their preferred genres. The application has two tabs. On the first one, a list of recommendations is shown. In this screen capture, nudge N4 is applied to one book ("Recommended by Stephen King"). By clicking on "show more", participants could inspect more details about the book. On the details page, participants were shown the full cover, the title, author, publisher and price of the book, along with a short synopsis. Figure 4 shows an example of a book details page.

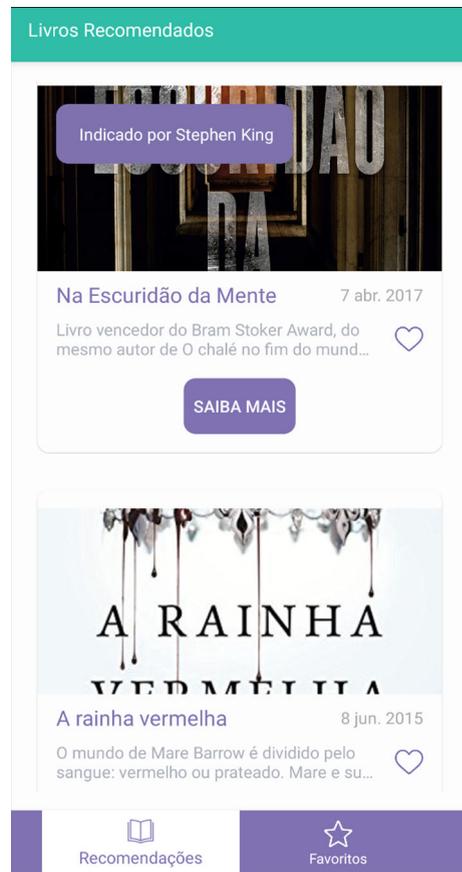
Putting an item on the reading list is accomplished by clicking on the heart symbol. On the second tab, the currently favorited items (i.e., those on the reading list) could be inspected.

For running the experiments, we developed an Android application using the Kotlin language. We focused on the Android platform due to its large number of users and the ease of distributing the application through the Google Play app store. The architecture used for the development of the application was the Clean Architecture (Martin 2022) along with the MVVM (Model-View-ViewModel) design pattern. This approach led to

⁵ These questions were translated to Portuguese.

⁶ Using only one item per construct may represent a possible research limitation. We however believe that the resulting risks, e.g., regarding a misinterpretation of the questions, are low, as we use questions from the validated inventory of the ResQue framework.

Fig. 3 Screen Capture taken from the Treatment Group, with nudge N4 applied (in Portuguese)



an architecture that is easy to maintain, testable, and with good readability. The back-end responsible for bringing the list of books to be displayed in the application was implemented as a simple online document. This way, updating the list was simpler. The same software was used to build the two versions of the software (Treatment and Control). For the treatment group, a switch was enabled to display the respective nudges. All user interactions with the application were recorded using the Mixpanel tool,⁷ which facilitates the process of logging and analyzing application log events.

4 Results and discussion

Our study's major findings are summarized in this section. First, we describe how the study was executed and how the participants were recruited. Afterwards, we investigate the effects of nudges on user behavior and user perceptions.

⁷ <https://mixpanel.com>.

Fig. 4 Screen Capture of a Book Details Page



4.1 Study execution and participants

The study was conducted between 11 February 2022 and 08 March 2022. We recruited users of the Brazilian social book recommendation site “Livros & Citações”⁸ as study participants. We posted announcements of the study on social media, e.g., on the Instagram account of the book recommendation site. To take part in the study, participants had to download and install the app from the Google Play app store.

1064 participants used the app and added at least one recommended book in the reading list. 520 of them were in the treatment group and 544 in the control group. 762 subjects successfully completed the post-task questionnaire. 367 of them were in the treatment group, and 395 were in the control group. More than 90% of our participants were female. We found that more than 50% of the subjects were between 18 and 25 years old regarding the age group. Less than 10% were older than 40, which reflects the average population of users of the book recommendation site. The majority of the participants can be considered engaged book readers. More than 65% of the

⁸ <https://www.livrosecitacoes.com>.

Table 2 Demographic information and reading habits of participants

Variable	Control (%)	Treatment (%)
<i>Gender</i>		
Female	95.19	93.73
Male	4.05	5.99
Other	0.76	0.27
<i>Reading habits</i>		
1–5 books per year	8.10	7.36
6–20 books per year	33.16	38.42
21–50 books per year	34.94	31.06
51–100 books per year	18.48	16.35
>100 books per year	5.32	6.81

Table 3 Absolute numbers of books placed on reading list (favorited)

	Control	Treatment	Σ
From preferred genres	1718	1535	3253
Off-profile	751	873	1624
Σ	2469	2408	4877

Numbers printed in bold represent sums

participants stated that they read between 6 to 50 books per year. More than 15% said they read even more than 50 books a year. The detailed information about gender and reading habits is shown in Table 2. The differences between the control and treatment groups for these characteristics are not statistically significant.⁹

We note that our general goal was to obtain highly reliable results based on a sample size that is large enough for a robust analysis.

4.2 Effects of nudging on user behavior

We analyzed the potential effects of the digital nudges from different perspectives.¹⁰ *Effect on Resulting Reading Lists* Table 3 shows how many items the participants of different groups have placed on their reading lists. The table also shows how many of them were from the participants' preferred genres and how many were off-profile items. Table 4 provides the mean number of items that were favorited *per user*, considering all items and only off-profile items.

From Table 4, we find that participants in the treatment group, on average, placed slightly more items from the total on their reading list. However, the observed differences were not statistically significant according to a Student's *t* test with $\alpha = 0.05$. We advised users to favorite 3 to 5 books before starting to use the app; so this might indicate they were simply following this suggestion.

⁹ According to a Student's *t* test with $\alpha = 0.05$. We use this common alpha-value also for subsequent statistical tests if not noted otherwise.

¹⁰ We share the anonymized collected data for reproducibility online at <https://github.com/gaahbie/unbook>.

Table 4 Average number of books placed on reading list (favorited)

	Control	Treatment	<i>p</i> value
Mean (std) of favorited items per user (total)	4.54 (1.09)	4.64 (1.01)	0.12
Mean (std) of fav'ed items per user (off-profile)	1.38 (1.10)	1.68 (1.13)	< 0.001

Table 5 Numbers of users who placed at least one off-profile item on their reading list

	Control	Treatment	Σ
At least one off-profile item	407 (74.82%)	440 (84.62%)	847
No off-profile item	137 (25.18%)	80 (15.38%)	214
Σ	544	520	1064

Numbers in bold represent sums

However, looking only at off-profile items that participants placed on the list reveals a difference. The nudges applied to these items were effective, participants in the treatment group adding considerably more off-profile items on their reading list. A chi-squared test given that data in Table 3 revealed that the differences were statistically significant, with $p < 0.001$.

In addition to the chi-squared test, we further compared the *mean* number of off-profile favorited items in the treatment and control groups. A Student's *t*-test revealed statistical significance ($p < 0.001$) and thus provides additional indications that the nudges were effective in getting users to engage with off-profile content. Table 5 completes this analysis and shows how many of the participants have placed *at least one* off-profile item on their reading list. The data show that the nudges stimulated more participants to pick at least one off-profile item than the control group. Again, the results are statistically significant ($p < 0.001$) according to a chi-squared test. This means that the nudges were effective motivating more users to explore off-profile items during this experiment.¹¹

This main result supports our central hypothesis that *digital nudges can be an effective means to stimulate users to explore items from genres that they were previously not among their preferred ones*. We note here that participants are already comparably open to exploration in our study without the nudges. In the control group, about 30% of the reading list items were off-profile in the control group, and this value was increased to about 36% in the treatment group. Regarding the considerable fraction of off-profile items that we observe in the control group, we hypothesize that this phenomenon may be, at least in part, a result of position/order bias that was also present in the control condition, with every second item being an off-profile item.

In terms of the selected books, we analyzed whether the nudges helped mitigate popularity biases to some extent. In many domains, we observe long-tail distributions where most of the attention by users is placed on a small set of items (the "short

¹¹ In order to validate that the statistical significance does not merely result from a potentially overpowered study, we repeated the analyses on three randomly created, much smaller subsamples ($N = 532$, with 272 in control, and 260 in treatment). In all cases, the differences were also statistically significant with $\alpha = 0.05$.

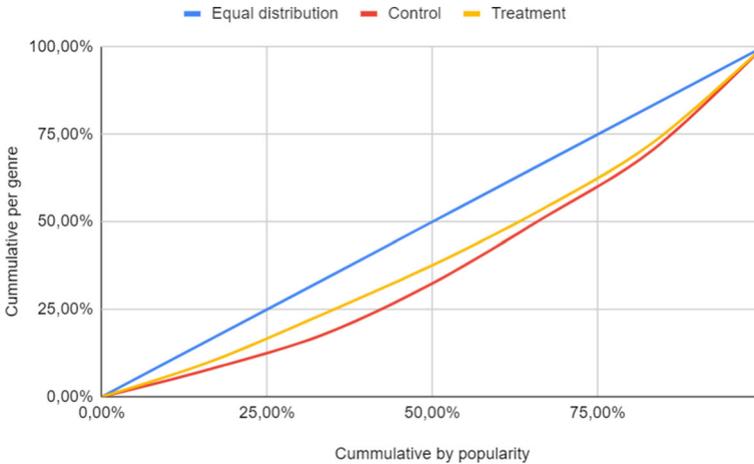


Fig. 5 Gini index of popularity of genres in reading lists

Table 6 Summary of other recorded user interactions. Numbers in parentheses show the average number of “show more” clicks per user

	Control	Treatment
Inspect item details (Preferred)	439 (0.80)	503 (0.96)
Inspect item details (Off-Profile)	224 (0.41)	322 (0.62)
Avg. nb. “inspect item details” events per user	1.21	1.58

head”). Recommender systems have the potential to reinforce such effects, leading to an effect where the *rich gets richer*, which may be undesirable both from a business perspective or the perspectives of discovery and fairness. To assess if nudges can help to counteract such biases, we computed the Gini coefficient concerning the popularity of the genres of the selected books for the treatment and the control groups, as done in Jannach et al. (2015). The Gini index is a metric that lies between zero and one and indicates how imbalanced the data are distributed, with higher numbers indicating higher concentration.

Figure 5 shows a Lorenz curve that visualizes the distributions. The Gini index corresponds to the perfect equality and perfect inequality ratio. We observe that the curve for the treatment group is closer to the perfect line, thus indicating that the genre preferences are more evenly distributed in the treatment group with the nudges. The Gini index for the treatment group is 0.17, and the one for the control group is 0.23, with the latter indicating higher concentration. Looking at the data, we found that nudging, in particular, led to an increased selection of books from the genre “non-fiction”. In exchange, the popularity of the genre “romance” lowered in the treatment group.

Effect on Exploration Behavior Besides additions to the reading lists, we recorded a number of other user interactions during the study. The most interesting statistic in

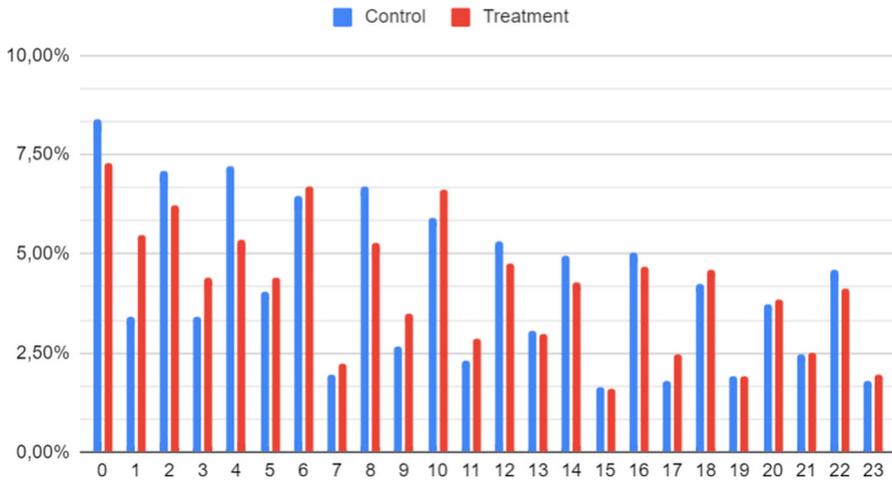


Fig. 6 Distribution of percentages of books being placed in reading lists at different list positions

this context is how often participants inspected the details of a book recommendation by pushing the “show more” button. Table 6 shows the collected data.

The results show that participants in the treatment group inspected about 30% more items than participants in control (from 1.21 to 1.58 “show more” clicks). Interestingly, participants in the treatment group also inspected more items from their preferred genres, even though their presentation was not different from the control group. Again, a chi-squared test given the data in Table 6 revealed that the differences are statistically significant ($p=0.037$).

We furthermore recorded how often participants *removed* items from their reading lists. The numbers were very similar in treatment and control, both regarding the removal of off-profile items and the removal of items from the participants’ preferred genre. For participants in the control group, we recorded 25 item removals; and there were 31 such removals in the treatment group. Interestingly, the differences mainly come from the increased removal of preferred genre books in the treatment group. However, the absolute numbers of recorded events are small, and the differences were not statistically significant. Overall, from the results, we have no indication that the nudges had a negative effect, where participants would initially place off-profile items on their reading lists to remove them later on.

Another aspect we considered is from *where* in the list participants picked the books they placed on their reading lists. Figure 6 illustrates how often an item from each position in the list was selected in the treatment and control group normalized by group size. For the control group, a relatively clear order effect can be observed for the items from the preferred genres (i.e., those with even position numbers). Items in odd position numbers are off-profile items, and the alteration between the types of items can be clearly observed in the figure.

For the treatment group, we also observe an order effect, but we also notice that the differences between the neighboring items are often much smaller, in particular at the beginning of the list. Later in the list, however, we can observe that the nudges’

effects often seem to become smaller. We also logged how far participants scrolled down the list and which one was the last book they viewed. It turned out that there were no differences between the groups, and about 70% of subjects in both groups scrolled down to the very end of the 24-item lists.

Finally, we looked at how much *time* participants needed to make their selection. On average, participants in the control group took 252.41 s—around 4.2 min—to select items for their reading lists (std = 187.66). On the other hand, participants in the treatment group needed 331.30 s—around 5.5 min—for the same task (std = 218.56). This is an increase of over 30%, which is also statistically significant ($p < 0.001$) according to a Welch's t test and a robust statistic analysis of the heteroscedastic data. We note that we removed outliers for this analysis that were beyond three standard deviations from the mean. The differences between control and treatment groups are similar and statistically significant also when not removing the outliers.

The observed increase in the time needed to select items for the reading list is expected given that participants in the treatment group interacted with more items, as shown in Table 6. We note, in general, that an increased number of interactions and more time spent with a list of recommendations can be caused by different things. In a favorable interpretation, participants were more engaged through the nudges, exploring more options in detail. In an unfavorable interpretation, the nudges might have raised participants' attention to some items that were later found to be irrelevant. Thus, the nudges may have caused unnecessary effort and distractions for the participants. To shed more light on these questions, we will analyze the results of the post-task questionnaire next.

Relationship between Prior Interest Diversity and Exploration For the last analysis in this subsection, we investigated whether the prior diversity of interests plays a role concerning the effects of the nudges. One hypothesis in this respect could be that participants who declared more preferred genres at the beginning of the study might also be more easily nudged to off-profile items because their predisposition is to be more open. To explore this aspect, we first computed the correlation between the number of initially preferred genres and the *fraction of off-profile* items in the final reading lists. Across all users, we found no such correlation ($\rho = 0.08$), and no correlation was found either when considering the treatment and control groups individually.¹²

In order to study if differences can be observed for the *extreme* groups in terms of their genre preference diversity, we split the users based on the median number of declared genre preferences. Figure 7 shows the distribution of the number of declared genre preferences. Correspondingly, we separately analyzed the data for participants who only declared one or two preferred genres for participants who had four or five preferred genres.

Again, however, in none of the subgroups we find a strong correlation between the participants' prior disposition in genre preferences and their selection behavior in terms of off-profile items. In all cases, whether considering both a low or a high number of preferred genres, and in both the treatment and control groups, the observed correlation

¹² Since the average number of favorited books was not different for the treatment and control groups, no correlation existed between the number of preferred genres and the absolute number of selected off-profile items either.

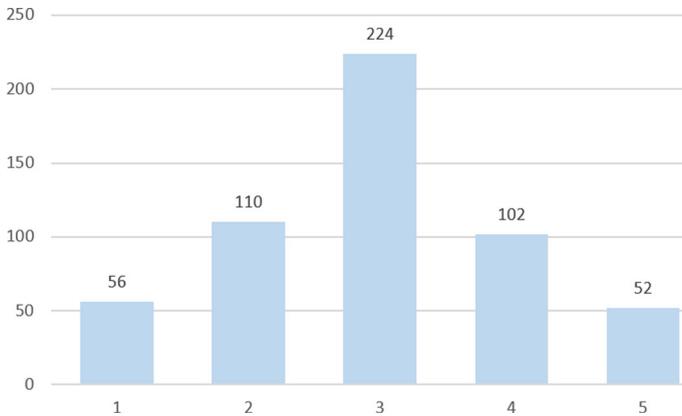


Fig. 7 Distribution of number of declared preferred genres. The x-axis shows the number of preferred genres, and the y-axis the number of participants per group

value did not exceed 0.2, which is commonly considered a very weak correlation in the best case. As a result, we have *no indication that the effectiveness of digital nudges depends on the prior preference diversity of the participants.*

4.3 Effects on quality perception and behavioral intentions

Main Observations The results of the post-task questionnaire on the participants' perception of the recommendations, their beliefs, satisfaction, and behavioral intentions are shown in Table 7. The table shows the means and the standard deviations for the different groups as well as the p value obtained with a Wilcoxon test¹³ We also note that we test 12 independent hypotheses here, based on constructs from the ResQue model (Pu et al. 2011). P-values lower than 0.05 are marked with two stars in the table.¹⁴

Generally, we observe that the differences in the means given a 5-point response scale are generally small and all below 5%. In terms of *user-perceived qualities* (Q1 to Q6), the first level of the ResQue model, no statistically significant differences at the alpha-level of 0.05 could be found. We recall that the recommendations were identical for participants with the same genre preferences. Among these quality factors, the largest numerical difference was observed for Q1, which might indicate some trend towards a lower accuracy perception in the presence of nudges. We can speculate that such an effect may happen in case the nudges raise the attention of some users who then found these off-profile items of little relevance to them. The calculated p-value (0.095) was however above the chosen alpha-level also in this case. Another interesting observation in terms of user-perceived qualities is that the perceived *diversity* of the

¹³ Shapiro–Wilk tests revealed that the data is not normally distributed ($p < 0.01$).

¹⁴ We additionally performed an analysis with robust Winsorized estimates of means and standard deviations and Welch's t tests. The significance results were well aligned with the outcomes of the Wilcoxon test.

Table 7 Mean responses and standard deviations for post-task questionnaire

		Control	Treatment	<i>p</i> value
Q1	The items recommended to me matched my interests	4.24 (1.04)	4.06 (1.08)	0.095
Q2	The app provided some really good recommendations	4.20 (0.96)	4.11 (0.94)	0.293
Q3	The items recommended to me are diverse	4.38 (0.87)	4.31 (0.91)	0.567
Q4	The recommendations helped me discover new books which I find interesting	3.97 (1.18)	4.05 (1.13)	0.287
Q5	I was surprised by some of the recommendations	3.74 (1.21)	3.71 (1.16)	0.350
Q6	The amount of information available was sufficient to make the decisions	3.87 (1.16)	3.91 (1.21)	0.525
Q7	I understood why the items were recommended to me	4.18 (1.09)	4.03 (1.1)	0.025**
Q8	Finding books to put on my reading list was easy	4.07 (1.19)	3.89 (1.23)	0.058
Q9	My overall satisfaction with my selection of books on my reading list is high	4.19 (1.03)	4.05 (1.08)	0.007**
Q10	My overall satisfaction with the app is high	4.25 (0.94)	4.10 (0.98)	0.029**
Q11	I like the app, and I would use it again in the future	4.48 (0.87)	4.31 (0.95)	0.017**
Q12	I am eager to read the book(s) I have put on the reading list	4.37 (0.93)	4.32 (0.97)	0.004**

p values lower than 0.05 are marked with **

recommendations did *not* increase, which one might have expected in a situation where the off-profile items were emphasized through the nudges.

On the second level of the ResQue framework, the *user beliefs*, we found a statistically significant drop in terms of perceived transparency (Q7) when the nudges were applied, i.e., they on average found it less clear why the items were recommended to them. One possible explanation of this effect may be that the nudges indeed raised the attention of the participants on the off-profile items in the list, which the participants may then have found unexpected, given the preferences they had specified at the beginning of the experiment. A slight drop was also observed for Q8 on the *ease-of-use* of the system, which however did not reach statistical significance ($p=0.058$). A potential drop in terms of the ease-of-use might be caused by the cognitively more complex user interface and the additional information provided by the nudges.

On the third and fourth levels of the ResQue framework (*user attitudes* and *behavioral intentions*), finally, significant differences between the treatment and control groups can be found for all examined aspects (Q9 to Q12). Including the nudges led to lower satisfaction with the made choices (Q9) and with the app in general (Q10). Furthermore, the participants expressed a slightly lower intention to use the app again in the future (Q10) and to read the books on their reading lists (Q11). We note here that while the differences are statistically significant, the observed drop is not large in terms of absolute values, which are all above four on the five-point scale.

Free-form Inputs As a final step of our analyses, we examined the free-text input that participants could provide at the end of the experiments. About 100 participants in each group provided feedback. Those comments that referred to the recommendations—and not to the app—were often about a particular wish to receive more recommendations. 16% of the users in the control group mentioned that the recommended list was too short, and 6% noticed that the recommendations included books from genres that they did not select. For the treatment group, the numbers were similar, with 15% highlighting that a list of 24 books may be too short and 5% of the users mentioning that the recommendations had off-profile items. Generally, we did not observe any differences in the feedback given in the control and treatment groups. However, an interesting observation in the treatment group was that no single comment was related to the nudges, even though the nudges effectively influenced the behavior and the choices of the participants.

4.4 Discussion

Practical Implications Our study, which involved more than 1,000 users of a Brazilian book recommendation site, leads to different important insights. Most importantly, we found that digital nudges proved to be a very effective means to help online users engage with content outside their past habits and stated preferences. Our participants not only explored the provided off-profile content, but they also found this off-profile content relevant to the extent that they placed off-profile items on their reading list. We recall here that the final average number of items in the reading lists did not vary largely across the experimental groups. Ultimately, this individual behavior of participants at the micro-level led to a shift in the distribution of items on the reading list at the

macro level, where a lower concentration on some popular genres was observed in the treatment group.

In terms of the absolute size of the observed effects, we emphasize that we did not expect a radical behavior shift. In the end, the intervention in the treatment group was limited to applying small visual cues to an otherwise unchanged recommendation list. In that light, the observed effects actually appear quite relevant and in a range that is often observed for nudging interventions in the real world. The review and meta-analysis in Arno et al. (2016) for example reports an average increase in healthier eating choices of 15.3%.

However, while the nudges proved to be highly effective in steering user behavior in specific directions in the short term, our study also revealed indications that “over-doing” it may be problematic in the long run. While the immediate user-perceived aspects of the recommendations, e.g., in terms of the quality or the diversity of the recommendations, were not largely affected by the nudges, we found that the nudges had significant effects on the users’ beliefs, their attitudes, and future behavioral intentions. Ultimately, these findings suggest that the nudges may in the worst case lead to a lower intention to use the system in the future, even though it was generally found useful.

As a result, we conclude that digital nudges as persuasive cues must be designed with care and with an eye on long-term effects on user satisfaction. This finding is in line with the trade-off between diversity and accuracy reported recently in Hansen et al. (2021). Clearly, our experimental design represents an extreme case because there was a digital nudge for every second item on the list, and also, because the nudges were only applied for off-profile items. In real-world applications, nudges should probably be applied more selectively and sparingly. On a more general level, our work adds to the growing body of literature that emphasizes the importance of considering the longitudinal effects of recommendations on user behavior (Zhang et al. 2019; Ferraro et al. 2020).

An interesting side-observation of our study is that some study participants assessed the accuracy of comparable sets of recommendations slightly differently in the presence of digital nudges. To our knowledge, such a phenomenon has not been reported in the literature before, and therefore, more research is required in terms of which factors influence the users’ perception of accuracy, which is a central element that drives the perceived usefulness of a recommender system.

Focus on Short-term Effects Like many studies on digital and “offline” nudging, we can only report the short-term effects of the nudges, which is a limitation also of most earlier works. By asking participants about their future behavioral intentions, we hope to have obtained some indications about the potential longitudinal effects of applying the nudges. However, these indications must be confirmed in future works. Otherwise, it may remain unclear if the used nudges have a lasting effect on the reading preferences of the users. In a recent work, Liang et al. (2022) studied longitudinal effects of nudging users towards increased genre exploration in the music domain. While the authors found that the effects of a *default nudge* faded quickly, they also observed that the user profiles “*did move somewhat towards the chosen genre*”. Given the similarity of the music and book domains, we are optimistic to observe some lasting effects of the nudges in our application setting. Generally, only few works exist that

study longitudinal effects of nudges. In the offline setting, Van Gestel et al. (2018) for example replicated the “food positioning nudge” mentioned in Thaler and Sunstein (2008), and they observed that re-positioning the food options had a measurable effect on food choices after several weeks. In the online world, Renaud and Zimmermann (2019) found that a hybrid nudge led to stronger password choices in a study that ran over a full year. Independent of the use of digital nudging, longitudinal studies of the effects of recommender systems on users are rare. Even reports from industry often only cover field tests that last a few weeks, see Jannach and Jugovac (2019) for a survey. An example for an academic study in the movie domain can be found in Taijala et al. (2018).

Ethical considerations Ethical considerations are crucial in the context of nudging, especially when it comes to the presentation of off-profile items as recommendations and the potential manipulation of user behavior. Kuyer and Gordijn (2023), delve into the controversial nature of nudging and its ethical justification, highlighting the importance of evaluating the ethical goals of nudging given its potential impact on consumers and citizens. The violation of autonomy is indeed a significant ethical concern associated with nudges. We acknowledge the debate surrounding nudges and autonomy, which encompasses notions of freedom of choice, agency, and self-constitution. It is important to note that in our study participants were fully informed about the experiment and had the option to withdraw their consent at any time, thereby addressing the issue of autonomy as a freedom of choice. Regarding autonomy as agency, we observed that the implementation of nudge mechanisms in our interface led to an increase in the average time spent on the experiment. This suggests that users had more time for reflection and decision-making, which aligns with the notion of autonomy as agency.

According to Bovens (2009) and to Kanev and Terziev (2017), nudges can be considered ethically acceptable when they are transparent, enabling individuals to perceive their presence. Transparency is considered a crucial factor in determining whether a nudge is manipulative. In our study, we observed that the implementation of nudge mechanisms in our interface resulted in an extended average duration of the experiment, indicating that users took more time to reflect and make their choices. Furthermore, the nudges employed in our study were designed to be transparent. Schmidt and Engelen (2020) also raised concerns about the manipulation of human behavior and the importance of transparency and equitable use of nudges are valid considerations. We made efforts to ensure transparency by providing a consent form before the start of the experiment, and we carefully selected the nudges implemented in the system to avoid disadvantaging already recommended items. The aspect of autonomy as self-constitution, which emphasizes the importance of individuals making choices aligned with their own values, is pertinent. Our intention was not to force users to deviate from their true preferences but rather to encourage exploration and discovery by highlighting items outside their usual preferences. We acknowledge that items matching the user’s profile should not be considered as inherently bad choices. Bovens (2009) emphasizes the importance of respecting individual values and preferences in ethical nudging practices. We agree that those responsible for implementing nudges bear a moral responsibility to consider the ethical consequences of their interventions and ensure that nudges are used to promote well-being and positive outcomes. By

carefully considering ethical aspects, we strive to use nudges in a beneficial manner that respects the values and choices of individuals.

5 Threats to validity

Ecological Validity In terms of ecological validity (realism), our study shares potential limitations with similar user studies in which (a) participants do not interact with a real production system that they use regularly and where (b) participants may not be “naturally” engaged in the study, but were invited to participate. A potential threat to validity in this context may therefore arise when a substantial fraction of study participants are not genuinely interested and engaged when interacting with the system. We however have no strong indications that this might have been the case. First, we recall that the subjects were genuine users of a book recommendation platform and their participation in the study was voluntary and unpaid. Second, we observed high rating responses given by participants with respect to the usefulness of the app and their intention to use the app in the future. Third, an analysis of the interaction data furthermore revealed that participants on average spent about five minutes to add items to their reading list, and more than half of the participants inspected at least one item detail page. Adding an arbitrary item to quickly complete the task could in contrast probably be done in much less time. An analysis of the scrolling behavior also showed that almost 80% of the participants scrolled down to the end of the 24-item list. Overall, we therefore believe that the ecological validity of our study is high.¹⁵

Generalizability Our present study focused on one particular domain (book recommendations), and we therefore cannot conclude with certainty to what extent the obtained findings would generalize to other settings.

Regarding the participants, we iterate that they were genuine users of a book recommendation platform and thus an essential and representative subset of the user population of a book recommender system. It is clear, though, that our participants are probably not fully representative of the average population of large online book stores or general e-commerce sites like Amazon.com, which have vast assortments of books. On average, our study participants read a few dozen books per year. They may generally also be more open to exploration than more occasional readers or readers with a narrow range of interests. Moreover, our study participants were primarily female, and it is left for future work to assess if there might be gender-specific differences concerning the effectiveness of the nudges.

Generally, however, our findings are in line with similar observations in terms of the effectiveness of nudges that were reported in earlier work, for example, in the domains of movies, healthy food or energy-saving, as discussed in Sect. 2. Our study therefore contributes to the accumulated evidence and knowledge about digital nudging. Whether or not the same specific nudges would work in a different application setting is however an important question to explore in future work.

¹⁵ We recall here that our study focused on analyzing behavior change in terms of adding items to a reading list. Such a change was observed in the study. Clearly, our study cannot inform about whether the participants actually bought, read, or liked the selected books.

6 Outlook

The study presented in this paper provides evidence that digital nudging may help users explore content beyond their past preference profiles. At the same time, our work indicates that when applied too extensively, nudges may negatively impact the users' beliefs and attitudes and, correspondingly, their future use of the system. Therefore, a critical part of future work is to study the effects of varying the degree of nudging to understand better the potential trade-off between helping users explore while maintaining user satisfaction with the system.

Given that the participants in our study were homogeneous in different respects, another essential question to address in future work is to understand if the effects of nudging depend on the characteristics of individuals. For example, do men and women perceive nudges differently? Or, does the effectiveness of nudges depend on certain personality traits (e.g., openness) or the expertise or engagement of a user for a given domain? In that context, we can identify at least two forms of considering user personality traits in future studies. First, we may vary the level of diversity for each user as suggested, e.g., in Wu et al. (2013), Guo et al. (2020). This would correspond to adapting the amount of off-profile items in the recommendations. Second, we may also try to adapt the type of the applied nudge based on user personality traits, as done, e.g., in Guo et al. (2020).

In our current study, we focused on the combined effects of the applied nudges on the bookmarking behavior of the participants. For other domains, such as health and nutrition, prior work (Berger et al. 2020; Jesse et al. 2021) reports that not all types of nudges are similarly effective. An in-depth analysis of the effectiveness of different nudges for our application setting was beyond our current research scope but represents an interesting area for future work. Our present study design does not allow us to derive conclusions in that direction, because there may be confounders such as position effects. A first analysis however provides indications that the non-hybrid Nudge N4 ("Social reference point") may be somewhat less effective than the other nudges.¹⁶ A deeper analysis through a new study however remains to be done to further investigate such indications.

Generally, our work contributes insights into yet another domain where digital nudges have not been explored in depth so far and where the nudges may contribute to achieving the specific societal goal of increasing literacy in a country. Future studies are needed to explore the effectiveness of nudging in other areas of societal relevance, for example, to stimulate online news readers to consume articles expressing opposing viewpoints to break filter bubbles and avoid radicalization.

In particular in the context of news recommendation, the additional question may arise how different off-profile recommendations can or should be from a given user's past preferences. In our experiment so far, we considered every non-preferred genre as being equally distant from the preferred ones. In the area of news recommendations, and in particular when it comes to political or controversial topics, it might be helpful to avoid off-profile recommendations that are too far away from the reader's

¹⁶ This indication is in line with the findings in Jesse et al. (2021), where a hybrid nudge turned out to be the most effective one.

past tendencies. Following the theory of the Overton Window of Political Possibility (Lehman 2010), it might therefore be meaningful to first establish a spectrum of “acceptable” opinions, and then select off-profile content for nudging, which has a reasonable chance of being considered by readers. Furthermore, following the discussions in Vermeulen (2022), another area for future work could be to give users more control about the level of exploration they would like to experience.

Funding Open access funding provided by University of Klagenfurt. This study was partially financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) under Finance Code 001.

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

Appendix

Figures 8, 9, 10 and 11 show stylized and translated screenshots of the app (treatment group) with the nudges N1 to N4 applied. Tables 8, 9, 10, 11, 12 and 13 then list the books that were recommended for each genre and which nudges were applied.

Table 8 List of *fantasy* books and applied nudges

Title	Author	N1	N2	N3	N4
Como o Rei de Elfhame aprendeu a odiar histórias	Holly Black				X
Estilhaça-me: 1	Tahereh Mafi				X
O guia definitivo do mochileiro das galáxias	Douglas Adams				X
A Princesa Prometida	William Goldman			X	
O Conto da Aia	Margaret Atwood			X	
Os Pilares da Terra	Ken Follett			X	
1984: Edição com Postais + Marcador	George Orwell		X		
Sombra e Ossos	Leigh Bardugo		X		
Trono de vidro (Vol. 1)	Sarah J. Maas		X		
A maldição do mar	Shea Ernshaw	X			
O canto mais escuro da floresta	Holly Black	X			
O Labirinto Do Fauno	Guillermo Del Toro	X			
A Fonte	CS Luis				
Box A Arma Escarlate: 4	Renata Ventura				
Depois	Stephen King				
Em Algum Lugar nas Estrelas	Clare Vanderpool				
Fábulas árabes	M. M. Jarouche				
Harry Potter e a Pedra Filosofal	J.K. Rowling				
Kindred: laços de sangue	Octavia E. Butler				
Maldição	Marie O'Regan				
MARVEL - OS PRIMEIROS 80 ANOS	Culturama				
Necronomicon em Netvilly	Kleber Inácio Da Silva				
Os reis do Wyld	Nicholas Eames				
Trilogia da Fundação - Deluxe	Isaac Asimov				
Vingadores: Guerra Infinita -	Steve Behling				

Table 9 List of *horror* books and applied nudges

Title	Author	N1	N2	N3	N4
Drácula - Dark Edition	Bram Stoker				X
Nunca Saia Sozinho	Charlie Donlea				X
O Retrato de Dorian Gray	Oscar Wilde				X
Joyland	Stephen King			X	
Na Escuridão da Mente	Paul Tremblay			X	
Tempo Estranho	Joe Hill			X	
A estrada da noite - Edição Luxo	Joe Hill		X		
Box Terríveis Mestres	Edgar Allan Poe		X		
It: A coisa	Stephen King		X		
Bom dia, você está morto!	Chell Sant'Ana	X			
Carrie (Com brindes)	Stephen King	X			
Frankenstein: O clássico está vivo!	Mary Shelley	X			
A canção de Bêlit: a tigrésa e o leão					
A ilha do tesouro	Robert Louis Stevenson				
A pequena caixa de Gwendy	Stephen King				
Box HP Lovecraft	Howard Phillips Lovecraft				
Calmaria Forçada	Rosane Montalvão				
Edgar Allan Poe - Medo Clássico	Edgar Allan Poe				
Eu sei o que vocês fizeram no verão passado	Lois Duncan				
Frankenstein, ou o Prometeu Moderno	M W Shelley				
O bosque das coisas perdidas	Shea Ernshaw				
O Colecionador	John Fowles				
O médico e o monstro	Robert Louis Stevenson				
Round 6 - por dentro da série	Park Minjoon				
Tempo Estranho	Joe Hill				
Tumular: a sete palmos do inferno	Thunder Dellú				
Uma mulher na escuridão	Charlie Donlea				

Table 10 List of *non-fiction* books and applied nudges

Title	Author	N1	N2	N3	N4
Ame pessoas, use coisas	Joshua Fields Millburn				X
Ed & Lorraine Warren: Vidas Eternas	Robert Curran				X
Monstros	Simon Sebag Montefiore				X
A Bailarina de Auschwitz	Edith Eva Eger			X	
Persópolis	Marjane Satrapi			X	
Sapiens - Uma Breve História da Humanidade	Yuval Noah Harari			X	
Box Memórias da Segunda Guerra Mundial	Winston Churchill		X		
Serial Killers - Anatomia do Mal	Harold Schrechter		X		
Ted Bundy: Um Estranho ao Meu Lado	Ann Rule		X		
ARQUIVOS SERIAL KILLERS	Ilana Casoy	X			
Enquanto eu respirar	Ana Michelle Soares	X			
Lady Killers	Tori Telfer	X			
50 Cent - Minha História, Minha Verdade	50 Cent				
A marca da vitória	Phil Knight				
Alpha Girls	Julian Guthrie				
BTK Profile: Máscara da Maldade	Roy Wenzl				
Decolonialidade e pensamento afrodiaspórico	Joaze Bernardino-Costa				
Denali	Ben Moon				
Elvis Presley	Gillian G. Gaar				
Escritos de Uma Vida	Sueli Carneiro				
Jovens heróis da União Soviética	Alex Halberstadt				
Meus primeiros 21	Nikki Sixx				
Na estrada com os Ramones	Monte A. Melnick				
O homem mais rico da Babilônia	George Samuel Clason				
Por que escrever?	Philip Roth				
Quem tem medo do feminismo negro?	Djamila Ribeiro				
Rebelde - Autobiografia do Criador de Conan	Robert E. Howard				
Stalin: Uma biografia	Robert Service				
Teologia do Corpo	São João Paulo II				
The Beatles	Vários Autores				
Tolkien e a Grande Guerra	John Garth				

Table 11 List of *romance* books and applied nudges

Title	Author	N1	N2	N3	N4
A vida invisível de Addie LaRue	V.E. Schwab				X
Amor & Gelato	Jenna Evans Welch				X
O morro dos ventos uivantes	Emily Brontë				X
Americanah	Chimamanda Ngozi Adichie			X	
Daisy Jones & The Six	Taylor Jenkins Reid			X	
O Rouxinol	Kristin Hannah			X	
É Assim que Acaba	Colleen Hoover		X		
Os Bridgertons, um amor de famli	Julia Quinn		X		
Sem julgamentos	Meg Cabot		X		
História é tudo que me deixou	Adam Silvera	X			
Pássaro e serpente (Vol. 1)	Shelby Mahurin	X			
Teto Para Dois	Beth O'leary	X			
novembro, 9	Colleen Hoover				
A soma de todos os afetos	Fabíola Simões				
Até o verão terminar	Colleen Hoover				
Em outra vida, talvez?	Taylor Jenkins Reid				
Malibu renasce	Taylor Jenkins Reid				
Mil beijos de garoto	Tillie Cole				
Missão romance	Lyssa Kay Adams				
Mr. 365	Ruth Clampett				
O acordo	Elle Kennedy				
O conde enfeitado - Edição Luxo	Julia Quinn				
O lado feio do amor	Colleen Hoover				
O Palácio de Papel	Miranda Cowley Heller				
Quase Uma Família	Sherryl Woods				
Rosas Esquecidas	Martha Hall Kelly				
Todas as suas (im)perfeições	Colleen Hoover				
Todo esse tempo	Rachael Lippincott				
Um pai para o meu bebê?	Tamires Barcellos				

Table 12 List of *suspense* books and applied nudges

Title	Author	N1	N2	N3	N4
Deixada Para Trás	Charlie Donlea				X
Gente Ansiosa	Fredrik Backman				X
O último julgamento	Scott Turow				X
Amada	Toni Morrison			X	
Mestre das Chamas	Joe Hill			X	
O Chamado do Cuco	Robert Galbraith			X	
Mestre das Chamas	Joe Hill		X		
Um de nós está mentindo	Karen M. McManus		X		
Verity	Colleen Hoover		X		
Assassinato no Expresso do Oriente	Agatha Christie	X			
Billy Summers	Stephen King	X			
O Homem de Giz	C. J. Tudor	X			
A Contrapartida - Livro 2: O Contra-ataque	Uranio Bonoldi				
A garota na neve	Danya Kukafka				
A Lista de Convidados	Lucy Foley				
A paciente silenciosa	Alex Michaelides				
A Salvação: Uma história de Vampiro	Diego Sousa				
A Última Festa	Lucy Foley				
Box Sherlock Holmes	Arthur Conan Doyle				
Coleção Agatha Christie - Box 1	Agatha Christie				
Detalhe final (Myron Bolitar - Livro 6)	Harlan Coben				
E não sobrou nenhum	Agatha Christie				
Em fogo lento	Paula Hawkins				
Estado de alerta	David Klass				
Mentiras incendiárias	Jennifer Lynn Alvarez				
Não Confie em Ninguém	Charlie Donlea				
O Clube do Crime das Quintas-Feiras	Richard Osman				
O mistério de Agatha Christie	Benedict Benedict				
Um de nós é o próximo	Karen M. McManus				
Um Segredo em Provence	Walter Barbosa				
Ve de Veronica, O	Raphael Prats				

Fig. 8 Screen Capture (Treatment Group), with nudge N1 applied

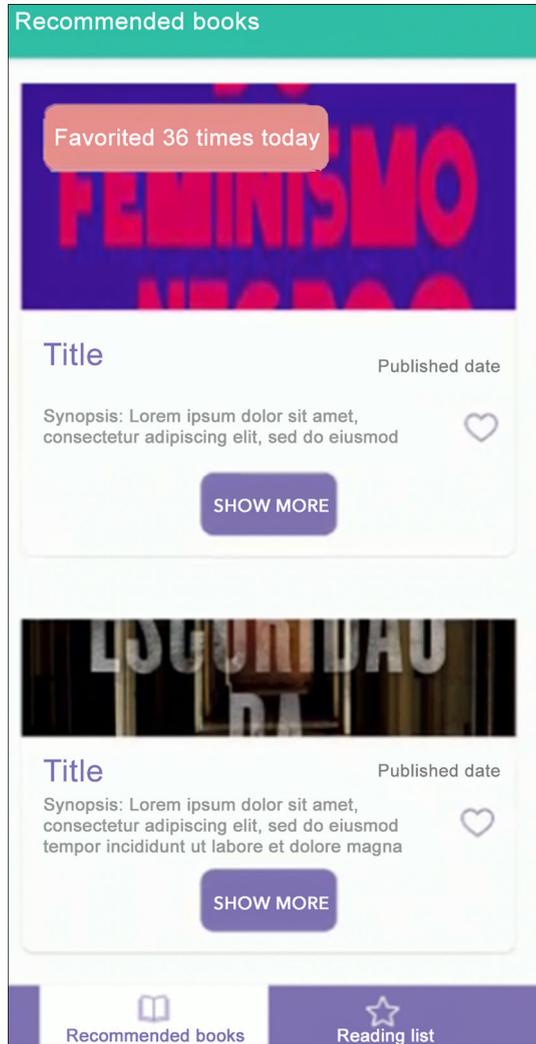


Fig. 9 Screen Capture (Treatment Group), with nudge N2 applied

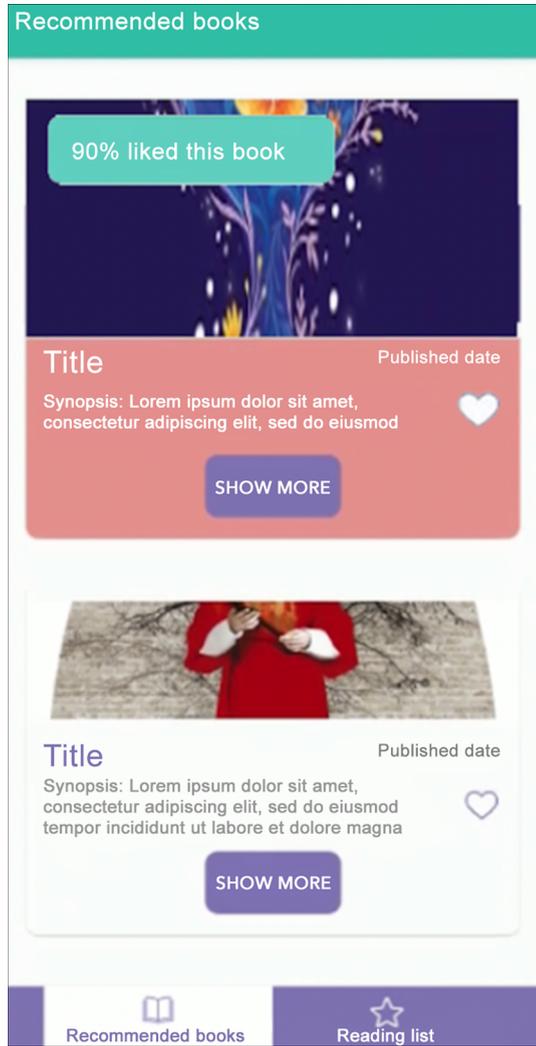


Fig. 10 Screen Capture (Treatment Group), with nudge N3 applied

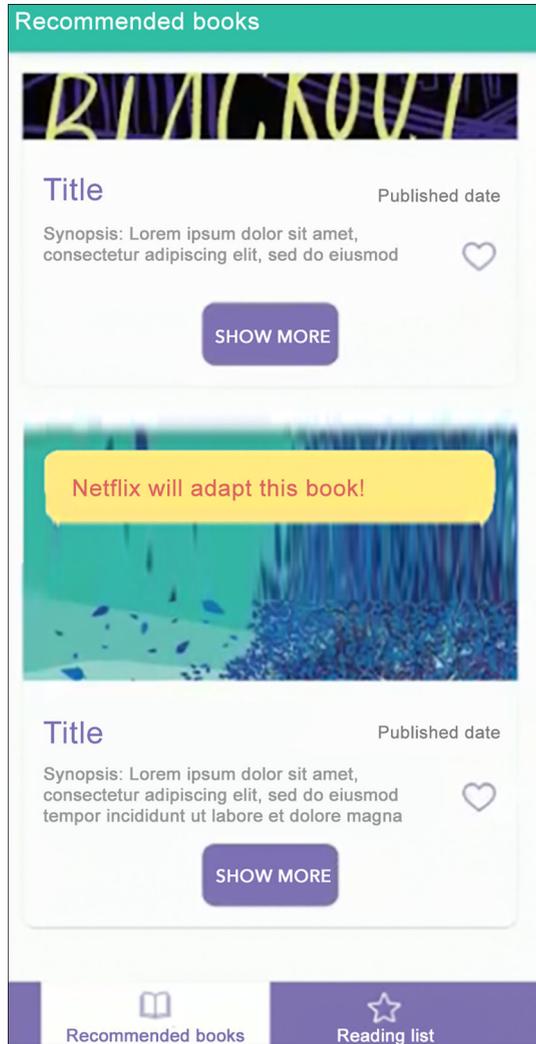


Fig. 11 Screen Capture (Treatment Group), with nudge N4 applied

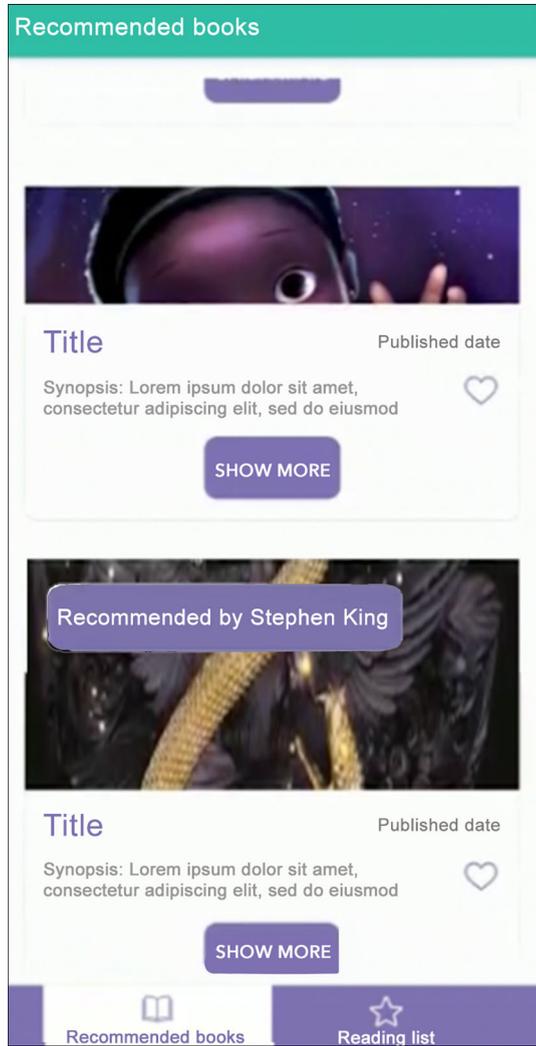


Table 13 List of books for *young adults* and applied nudges

Title	Author	N1	N2	N3	N4
Conectadas	Clara Alves				X
O príncipe cruel (Vol. 1 O Povo do Ar)	Holly Black				X
Por lugares incríveis	Jennifer Niven				X
O Livro da Selva	Rudyard Kipling			X	
O Ódio que Você Semeia	Angie Thomas			X	
Sulwe	Lupita Nyong'o			X	
A rainha vermelha	Victoria Aveyard		X		
A Seleção: 1	Kiera Cass		X		
Vermelho, branco e sangue azul	Casey McQuiston		X		
Coraline	Neil Gaiman	X			
Heartstopper: Dois garotos, um encontro (vol. 1)	Alice Oseman	X			
Última parada	Casey McQuiston	X			
A rainha do nada (Vol. 3 O Povo do Ar)	Holly Black				
Aristóteles e Dante...	Benjamin Alire Sáenz				
Ash	Malinda Lo				
Assim você me mata	Karen McManus				
Blackout: O amor também brilha no escuro	Dhonielle Clayton				
Bruxa Natural	Arin Murphy-Hiscock				
Confusão é meu nome do meio	Stephanie Tromly				
Coragem	Raina Telgemeier				
Corte de chamas prateadas	Sarah J. Maas				
Decifra-me	Tahereh Mafi				
Em fogo alto - com brinde exclusivo	Elizabeth Acevedo				
Fat Chance: A vez de Charlie Vega	Crystal Maldonado				
Harry Potter e a pedra filosofal	J.K. Rowling				
Metamorfose	Franz Kafka				
Os últimos jovens da Terra	Max Brallier				
Um dia a alma transborda	Marina Leme				
Unifica-me + brindes	Tahereh Mafi				

References

- Arno, A., Thomas, S.: The efficacy of nudge theory strategies in influencing adult dietary behaviour: a systematic review and meta-analysis. *BMC Public Health* **16**, 676 (2016)
- Baeza-Yates, R., Ribeiro-Neto, B.A.: *Modern Information Retrieval-The Concepts and Technology Behind Search*. Addison-Wesley, New York (2011)
- Berger, M., Müller, C., Nüske, N.: Digital nudging in online grocery stores: Towards ecologically sustainable nutrition. In: *Proceedings of the 41st International Conference on Information Systems (ICIS 2020)* (2020)
- Bothos, E., Apostolou, D., Mentzas, G.: A recommender for persuasive messages in route planning applications. In: *2016 7th International Conference on Information, Intelligence, Systems and Applications (IISA)*, pp. 1–5 (2016)

- Bovens, L.: The ethics of nudge. In: *Preference Change*, pp. 207–219. Springer, Berlin (2009)
- Bradley, K., Smyth, B.: Improving recommendation diversity. In: *Twelfth Irish Conference on Artificial Intelligence and Cognitive Science*, pp. 85–94 (2001)
- Caraban, A., Karapanos, E., Gonçalves, D., Campos, P.: 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*, (CHI '19), pp. 1–15 (2019)
- Castells, P., Vargas, S., Wang, J.: Novelty and diversity metrics for recommender systems: choice, discovery and relevance. In: *International Workshop on Diversity in Document Retrieval* (2011)
- Celma, Ö.: *Music Recommendation and Discovery—The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Springer, Berlin (2020)
- Chen, L., Yang, Y., Wang, N., Yang, K., Yuan, Q.: How serendipity improves user satisfaction with recommendations? A large-scale user evaluation. In: *The World Wide Web Conference* (2019)
- Dias, M.B., Locher, D., Li, M., El-Deredy, W., Lisboa, P.J.G.: The value of personalised recommender systems to e-business: a case study. In: *Proceedings of the 2008 ACM Conference on Recommender Systems*, RecSys '08, pp. 291–294 (2008)
- Ekstrand, M.D., Harper, F.M., Willemsen, M.C., Konstan, J.A.: User perception of differences in recommender algorithms. In: *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, pp. 161–168 (2014)
- Elsweiler, D., Trattner, C., Harvey, M.: Exploiting food choice biases for healthier recipe recommendation. In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, pp. 575–584 (2017)
- Ferraro, A., Jannach, D., Serra, X.: Exploring longitudinal effects of session-based recommendations. In: *Proceedings of the 2020 ACM Conference on Recommender Systems (RecSys '20)* (2020)
- Fogg, B.J.: Persuasive technology: using computers to change what we think and do. *Ubiquity* **2002**(December), 2 (2002)
- Ge, M., Gedikli, F., Jannach, D.: Placing high-diversity items in top-n recommendation lists. In: *Proceedings of the Workshop on Intelligent Techniques for Web Personalization and Recommender Systems (ITWP 2011 at IJCAI 2011)* (2011)
- Ge, M., Jannach, D., Gedikli, F., Hepp, M.: Effects of the placement of diverse items in recommendation lists. In: *14th International Conference on Enterprise Information Systems (ICEIS 2012)*, pp. 201–208 (2012)
- Gena, C., Grillo, P., Lieto, A., Mattutino, C., Vernerio, F.: When personalization is not an option: an in-the-wild study on persuasive news recommendation. *Information* **10**(10), 300 (2019)
- Gomez-Uribe, C.A., Hunt, N.: The Netflix recommender system: algorithms, business value, and innovation. *Trans. Manag. Inf. Syst.* **6**(4), 13:1–13:19 (2015)
- Guo, Y., Zhang, Z., Guo, Y., Guo, X.: Nudging personalized password policies by understanding users' personality. *Comput. Secur.* **94**, 101801 (2020)
- Hansen, C., Mehrotra, R., Hansen, C., Brost, B., Maystre, L., Lalmas, M.: Shifting consumption towards diverse content on music streaming platforms. In: *Proceedings of the 14th ACM International Conference on Web Search and Data Mining, WSDM '21*, pp. 238–246 (2021)
- Heitz, L., Lischka, J.A., Birrer, A., Paudel, B., Tolmeijer, S., Laugwitz, L., Bernstein, A.: Benefits of diverse news recommendations for democracy: a user study. *Digital Journalism* 1–21 (forthcoming) (2022)
- Jannach, D., Adomavicius, G.: Recommendations with a purpose. In: *RecSys '16*, pp. 7–10 (2016)
- Jannach, D., Jugovac, M.: Measuring the business value of recommender systems. *ACM Trans. Manag. Inf. Syst.* **10**(4) (2019)
- Jannach, D., Lerche, L., Kamehkhosh, I., Jugovac, M.: What recommenders recommend: an analysis of recommendation biases and possible countermeasures. *User Model. User-Adap. Inter.* **25**(5), 427–491 (2015)
- Jesse, M., Jannach, D.: Digital nudging with recommender systems: Survey and future directions. *Comput. Human Behav. Rep.* **3** (2021)
- Jesse, M., Jannach, D., Gula, B.: Digital nudging for online food choices. *Front. Psychol.* **12**, 729589 (2021)
- Joachims, T.: Evaluating retrieval performance using clickthrough data. In: Franke, J., Nakhaeizadeh, G., Renz, I. (eds.) *Text Mining*. Physica (2003)
- Kamehkhosh, I., Bonnin, G., Jannach, D.: Effects of recommendations on the playlist creation behavior of users. *User Model. User-Adap. Inter.* **30**, 285–322 (2019)

- Kaminskas, M., Bridge, D.: Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Trans. Interact. Intell. Syst.* **7**(1), 1–42 (2016)
- Kanev, D., Terziev, V.: Behavioral economics: Development, condition and perspectives. *IJASOS Int. E-J. Adv. Soc. Sci.* **8**, III (2017)
- Kapoor, K., Kumar, V., Terveen, L., Konstan, J.A., Schrater, P.: “I like to explore sometimes”: Adapting to dynamic user novelty preferences. In: *Proceedings of the 9th ACM Conference on Recommender Systems, RecSys '15*, pp. 19–26 (2015)
- Köcher, S., Jugovac, M., Jannach, D., Holzmüller, H.: New hidden persuaders: an investigation of attribute-level anchoring effects of product recommendations. *J. Retail.* **95**(1), 24–41 (2019)
- Kunaver, M., Požrl, T.: Diversity in recommender systems—a survey. *Knowl.-Based Syst.* **123**, 154–162 (2017)
- Kuyer, P., Gordijn, B.: Nudge in perspective: a systematic literature review on the ethical issues with nudging. *Ration. Soc.* **35**, 191–230 (2023)
- Lawrence, R.D., Almasi, G.S., Kotlyar, V., Viveros, M.S., Duri, S.S.: Personalization of supermarket product recommendations. *Data Min. Knowl. Discov.* **5**(1–2), 11–32 (2001)
- Lehman, J.G.: An Introduction to the Overton Window of Political Possibility. Mackinac Center for Public Policy. <https://www.mackinac.org/12481> (2010)
- Lerche, L., Jannach, D., Ludewig, M.: On the value of reminders within e-commerce recommendations. In: *UMAP '16*, pp. 27–35 (2016)
- Liang, Y., Willemsen, M.C.: The role of preference consistency, defaults and musical expertise in users' exploration behavior in a genre exploration recommender. In: *Fifteenth ACM Conference on Recommender Systems* (2021)
- Liang, Y., Willemsen, M.C.: Exploring the longitudinal effects of nudging on users' music genre exploration behavior and listening preferences. In: *Proceedings of the 16th ACM Conference on Recommender Systems, RecSys '22*, pp. 3–13 (2022)
- Lops, P., de Gemmis, M., Semeraro, G.: Content-based recommender systems: state of the art and trends. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P. (eds.) *Recommender Systems Handbook*, pp. 73–105. Springer, Berlin (2011)
- Ludewig, M., Latifi, S., Mauro, N., Jannach, D.: Empirical analysis of session-based recommendation algorithms. *User Model. User-Adap. Inter.* **31**(1), 149–181 (2021)
- Martin, R.C.: Clean architecture. <https://blog.cleancoder.com/uncle-bob/2012/08/13/the-clean-architecture.html>. Accessed 28 April 2022
- McInerney, J., Lacker, B., Hansen, S., Higley, K., Bouchard, H., Gruson, A., Mehrotra, R.: Explore, exploit, and explain: personalizing explainable recommendations with bandits. In: *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys '18*, pp. 31–39 (2018)
- McNee, S.M., Riedl, J., Konstan, J.A.: Being accurate is not enough: How accuracy metrics have hurt recommender systems. In: *CHI '06 Extended Abstracts on Human Factors in Computing Systems, CHI EA '06*, pp. 1097–1101 (2006)
- Mirsch, T., Lehrner, C., Jung, R.: Digital nudging: altering user behavior in digital environments. *Wirtschaftsinformatik und Angewandte Informatik*, pp. 634–648 (2017)
- Mols, F., Haslam, S.A., Jetten, J., Steffens, N.K.: Why a nudge is not enough: a social identity critique of governance by stealth. *Eur. J. Polit. Res.* **54**(1), 81–98 (2015)
- Oosterhuis, H.: Learning from user interactions with rankings: a unification of the field. *ArXiv:2012.06576* (2020)
- Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework for recommender systems. In: *Proceedings of the 5th ACM Conference on Recommender Systems, RecSys '11*, pp. 157–164 (2011)
- Renaud, K., Zimmermann, V.: Nudging folks towards stronger password choices: providing certainty is the key. *Behav. Public Policy* **3**(2), 228–258 (2019)
- Rieger, A., Draws, T., Theune, M., Tintarev, N.: This item might reinforce your opinion: Obfuscation and labeling of search results to mitigate confirmation bias. In: *Proceedings of the 32nd ACM Conference on Hypertext and Social Media* (2021)
- Schmidt, A.T., Engelen, B.: The ethics of nudging: an overview. *Philos. Compass* **15** (2020)
- Schneider, C., Weinmann, M., vom Brocke, J.: Digital nudging: guiding online user choices through interface design. *Commun. ACM* **61**(7), 67–73 (2018)

- Starke, A.D., Willemsen, M.C., Snijders, C.: With a little help from my peers: Depicting social norms in a recommender interface to promote energy conservation. In: Proceedings of the 25th International Conference on Intelligent User Interfaces, IUI '20, pp. 568–578 (2020)
- Starke, A.D., Willemsen, M.C., Trattner, C.: Nudging healthy choices in food search through visual attractiveness. *Front. Artif. Intell.* **4** (2021)
- Tajjala, T.T., Willemsen, M.C., Konstan, J.A.: Movieexplorer: Building an interactive exploration tool from ratings and latent taste spaces. In: Proceedings of the 33rd Annual ACM Symposium on Applied Computing, pp. 1383–1392 (2018)
- Thaler, R.H., Sunstein, C.R.: *Improving Decisions About Health, Wealth and Happiness*. Yale University Press, Nudge (2008)
- Van Gestel, L., Kroese, F., De Ridder, D.: Nudging at the checkout counter—a longitudinal study of the effect of a food repositioning nudge on healthy food choice. *Psychol. Health* **33**(6), 800–809 (2018)
- Verma, A., Wadhwa, A., Singh, N., Beniwal, S., Kaushal, R., Kumaraguru, P.: Followee management: helping users follow the right users on online social media. In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 1286–1290 (2018)
- Vermeulen, J.: To nudge or not to nudge: news recommendation as a tool to achieve online media pluralism. *Digital Journalism* (2022)
- Wu, W., Chen, L., He, L.: Using personality to adjust diversity in recommender systems. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media, HT '13, pp. 225–229 (2013)
- Yoo, K.-H., Gretzel, U., Zanker, M.: *Persuasive Recommender Systems-Conceptual Background and Implications*. Springer, Berlin (2013)
- Zhang, J., Adomavicius, G., Gupta, A., Ketter, W.: Consumption and performance: Understanding longitudinal dynamics of recommender systems via an agent-based simulation framework. *Inf. Syst. Res.* (forthcoming) (2019)
- Ziegler, C.-N., McNee, S.M., Konstan, J.A., Lausen, G.: Improving recommendation lists through topic diversification. In: Proceedings of the 14th International Conference on World Wide Web, WWW '05, pp. 22–32 (2005)

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

Gabrielle Alves is pursuing her PhD in Computer Science at the University of São Paulo, Brazil. Her primary research delves into user behavior within human-computer interaction, especially concerning recommender systems.

Dietmar Jannach is a Professor of Computer Science at University of Klagenfurt, Austria. He has worked on different areas of artificial intelligence, including recommender systems, model-based diagnosis, and knowledge-based systems. He is the leading author of a textbook on recommender systems and has authored more than hundred research papers, focusing on the application of artificial intelligence technology to practical problems.

Rodrigo Ferrari de Souza is a mobile app developer at a financial company in Brazil and a master's student at the University of São Paulo, Brazil. His primary research area revolves around recommendation systems. He is researching calibration, fairness, and biases within recommendation systems.

Daniela Damian is a professor of software engineering in the Department of Computer Science at the University of Victoria in Victoria, British Columbia, Canada. There, she leads the Software Engineering Global InterAction Research Lab. Her research interests include empirical and collaborative software engineering, requirements engineering, and software ecosystems

Marcelo Garcia Manzato is Associate Professor at the Mathematics and Computer Science Institute of the University of São Paulo, SP, Brazil. His research interests are recommender systems, machine learning and human-computer interaction. He has been collaborating and supervising works related to recommender systems for more than 10 years.