On the Negative Perception of Cross-domain Recommendations and Explanations

Denis Kotkov
Department of Computer Science
University of Helsinki
Helsinki, Finland
kotkov.denis.ig@gmail.com

Alan Medlar
Department of Computer Science
University of Helsinki
Helsinki, Finland
alan.j.medlar@helsinki.fi

Yang Liu
Department of Computer Science
University of Helsinki
Helsinki, Finland
yang.liu@helsinki.fi

Dorota Głowacka
Department of Computer Science
University of Helsinki
Helsinki, Finland
dorota.glowacka@helsinki.fi

ABSTRACT

Recommender systems typically operate within a single domain, for example, recommending books based on users' reading habits. If such data is unavailable, it may be possible to make cross-domain recommendations and recommend books based on user preferences from another domain, such as movies. However, despite considerable research on cross-domain recommendations, no studies have investigated their impact on users' behavioural intentions or system perceptions compared to single-domain recommendations. Similarly, while single-domain explanations have been shown to improve users' perceptions of recommendations, there are no comparable studies for the cross-domain case.

In this article, we present a between-subject study (N=237) of users' behavioural intentions and perceptions of book recommendations. The study was designed to disentangle the effects of whether recommendations were single- or cross-domain from whether explanations were present or not. Our results show that cross-domain recommendations have lower trust and interest than single-domain recommendations, regardless of their quality. While these negative effects can be ameliorated by cross-domain explanations, they are still perceived as inferior to single-domain recommendations without explanations. Last, we show that explanations decrease interest in the single-domain case, but increase perceived transparency and scrutability in both single- and cross-domain recommendations. Ourfi ndings offer valuable insights into the impact of recommendation provenance on user experience and could inform the future development of cross-domain recommender systems.

CCS CONCEPTS

• **Human-centered computing → Graphical user interfaces**; • **Information systems → Recommender systems**.

SIGIR '24, July 14–18, 2024, Washington, DC, USA

ACM ISBN 979-8-4007-0431-4/24/07
https://doi.org/10.1145/3626772.3657735

KEYWORDS

cross-domain explanations, explanations, cross-domain recommendations, user study, recommender systems

ACM Reference Format:
Denis Kotkov, Alan Medlar, Yang Liu, and Dorota Głowacka. 2024. On the Negative Perception of Cross-domain Recommendations and Explanations. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24), July 14–18, 2024, Washington, DC, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3626772.3657735

1 INTRODUCTION

Recommendations are typically generated based on users' preferences for items within the same domain [47]. A recommender system might, for example, suggest books based on users' reading habits, book purchases, or ratings. However, it is also possible to make cross-domain recommendations: recommending items from one domain based on users' preferences in another, such as recommending books based on movie preferences. In particular, cross-domain recommendation has been positioned as a solution to the cold start problem, where there is insufficient information about users' interests in the target domain to make personalized recommendations [48]. Despite the wide variety of cross-domain recommendation algorithms [12, 63], to our knowledge there are no studies of how cross-domain recommendations are perceived by users compared to those from single-domain recommenders. In a related topic, recommendation explanations have been shown to increase users' interest in recommended items as well as their perceived transparency with respect to how recommendations were generated [14], but there are no empirical studies of explanations in the cross-domain setting [59], where recommendations in the target domain are justified in terms of users' interests in another domain[1].

In this article, we investigate how cross-domain recommendation and explanations influence users' perceptions of book recommendations. We consider four scenarios: (i) a recommender system that makes book recommendations based on users' book interests, (ii)

[1]We note, however, that cross-domain explanations have been mentioned in several articles as possible future work (e.g., [63, 64]), but, to our knowledge, these ideas were not investigated.

a cross-domain recommender that makes book recommendations based on users' movie interests, (iii) the recommender from (i) with additional explanations and (iv) the cross-domain recommender from (ii) with cross-domain explanations. By considering all four scenarios, we are able to understand how user perceptions and behavioural intentions are impacted, while disentangling the effects of whether recommendations were single- or cross-domain from whether explanations were present or not. We, therefore, investigate the following research questions:

- **RQ1:** How does the cross-domain setting affect user perceptions of recommendations?
- **RQ2:** How do recommendation explanations affect user perceptions of single- and cross-domain recommendations?

To answer these research questions, we identified numerous experimental confounders that influenced our study design. First, **it must be unambiguous what information is (and is not) available for generating recommendations**. In the cross-domain setting, for example, users must believe that book recommendations are made solely on the basis of their movie interests. Therefore, we used a between-subject study design, where each participant is situated in only one of the four scenarios outlined above (see Figure 1 for an overview). In this way, there is no potential for information to be shared between experimental conditions. Second, **the quality of recommendations must be constant across all scenarios** and cannot be systematically worse, for example, in the cross-domain case. Therefore, irrespective of whether a participant provides information about movies or books, they are always shown a randomly generated list of recommendations. By not making actual recommendations based on user preferences, we were able to isolate the impact of the presumed origin of recommendations from the quality of the recommendations themselves. Lastly, to make statements about the effect of recommendation explanations, **we need to verify that our explanations provide a credible rationale for a given recommendation**. We, therefore, additionally asked participants to rate the quality of single- and cross-domain recommendation explanations for books they had already read to assess the quality of explanations.

The results of our study show that the mere fact that users believed book recommendations were generated based on movie preferences decreased trust in the recommender system and their interest in recommendations. The negative effects of cross-domain recommendations were ameliorated when accompanied by cross-domain explanations, but overall were still perceived as inferior to single-domain recommendations. Our study also shows that explanations improved perceived transparency and scrutability of recommendations in both the single- and cross-domain cases, but in the single-domain case this was at the expense of decreased interest in recommendations. Lastly, the presence of recommendations that were already familiar to users increased perceived transparency, trustworthiness and persuasiveness, but decreased perceived efficiency. The article has three main contributions:

- **Study design:** We present a novel study design to investigate how cross-domain recommendations and explanations influence user perceptions and behavioural intentions compared to single-domain recommendations while avoiding numerous confounding factors.
- **Perception of cross-domain recommendation:** We show that cross-domain recommendations are at a disadvantage compared to single-domain recommendations across various dimensions of user perception.
- **Impact of recommendation explanations:** Prior research has demonstrated the benefits of recommendation explanations. We extend this work to the cross-domain case, highlighting the potential for explanations to improve users' perceptions of cross-domain recommendations.

2 RELATED WORK

In this section, we describe related work in cross-domain recommendation and recommendation explanations. We emphasize the popularity of cross-domain recommender systems, while noting the absence of user studies. Similarly, despite numerous studies on explanations in recommender systems, there are no user studies of cross-domain explanations [63, 64].

2.1 Cross-domain Recommendation

2.1.1 Approaches to Cross-domain Recommendation. Cross-domain recommender systems attempt to address the shortcomings of single-domain recommenders, including user cold start and data sparsity, by sharing knowledge between source and target domains [63]. We focus on methods that attempt to resolve the cold start problem – where users have ratings in a source domain (e.g., movies), but are missing them in the target domain (e.g., books) – as it is the closest scenario to our study design.

Of the methods that aim to address cold start, a majority are based on inter-domain shared/correlated tags [16, 31, 50, 62] or the embedding and mapping approach [13, 24, 35, 60, 68, 70]. Methods based on shared tags between domains were initially based on collaborativefi ltering [16] and similiarity matrices between items and users [50]. However, as domains can use different vocabularies, subsequent approaches sought to identify semantically equivalent tags between domains. For example, Kumar et al. [31] proposed a semantic clustering-based cross-domain recommendation algorithm that used ontologies to map semantic relationships between tags in different domains. Similarly, Yang et al. [62] used data from online encyclopedias to build a multi-partite graph representing the similarity of tags in different domains. Later approaches were based on the embedding and mapping framework proposed by Man et al. [35]. Embedding and mapping consists of creating latent representations of items and users, and creating a mapping function between domains. In the original approach, latent representations were created using matrix factorization and the mapping function was derived from Bayesian personalized ranking [35]. Follow-on work improved on this approach by applying different latent representations [60] and developing novel mapping functions [24, 70]. Recent approaches have extended the embedding and mapping approach by incorporating aspect correlations based on user/item reviews [68]. For a more comprehensive overview of cross-domain recommendation algorithms, we refer readers to recent surveys by Khan et al. [25] and Zang et al. [63].

Despite the existence of numerous cross-domain recommendation algorithms for the scenario we are investigating, we decided to use randomly generated recommendations as it was important to ensure that the quality of recommendations was the same in both single- and cross-domain recommendation settings.

2.1.2 User Perception of Cross-domain Recommendations. User perception of recommendations plays an important role in recommendation acceptance [1, 53, 58]. Cross-domain recommender systems, however, have only been evaluated offline using precollected data sets [12]. Despite increasing interest in cross-domain recommendation [25], we could not identify a single study where cross-domain recommendations were explicitly evaluated with real users (based on recent surveys [7, 17, 25, 63, 69], individual works [49, 61] and a comprehensive literature review[2]). Meanwhile, additional factors of cross-domain recommender systems may affect user interactions and perceptions of recommendations compared to a single-domain recommender system, highlighting the need for user experiments to explore these dynamics further.

2.2 Explanations in Recommender Systems

Explanations have been shown to affect user perception of recommendations in single-domain recommender systems [23, 53, 57], but remains unexplored for cross-domain recommendations [63, 64].

2.2.1 Explanation Evaluation. Explanations can be evaluated online or offline [65]. Offline evaluation uses precollected data sets and evaluation metrics, such as BLEU [11, 44], whereas online evaluation is conducted through user studies orfi eld experiments with real users [5, 23, 58]. User studies for explanation evaluation have two common experimental settings: only showing users the explanations [5, 19, 39], and showing users explanations together with item metadata [42, 54, 58]. For example, Bilgic and Mooney evaluated explanations for book recommendations by asking users to fill out a questionnaire initially with only explanations and, subsequently, based on book content [5]. In contrast, Vig et al. displayed explanations along with movie metadata from MovieLens [58].

In our study, we decided to display explanations along with the titles and descriptions of recommended books as it more closely mirrors real-life scenarios.

2.2.2 User Perception of Explanations. In a recent survey of evaluation methods for recommendation explanations, Chen et al. identified 118 studies, of which 55 contained an online evaluation study [11]. Of these studies, only 7 were case-control studies where one of the experimental conditions corresponded to the absence of explanations [14, 30, 37, 38, 42, 56, 66]. These studies investigated a majority of the seven possible aims of recommendation explanations enumerated by Tintarev and Masthoff[51]:

- Transparency [14, 42, 66]: user understanding of how recommendations are generated
- Scrutability: ability of users to tell if the system is wrong
- Trust [42]: user confidence that the system is right
- Effectiveness [37]: user ability to make good decisions
- Persuasiveness [30, 66]: convincing users to consume items
- Efficiency [37]: user ability to make quick decisions
- Satisfaction [30]: enjoyment from using the system

[2]We searched ACM digital library using the search query: title contains "cross-domain" and ("recommendation" or "recommender"), which resulted in 188 articles. None of these articles contained user studies.

More concretely, explanations were shown to increase transparency in mobile e-commerce [66], image recommendation [14], and e-learning [42]. Zhang et al. highlighted the persuasiveness of explanations generated from user reviews [66]. Millecamp et al. observed that explanations can increase effectiveness and efficiency of music recommendations depending on users' level of music sophistication [37]. Kouki et al. compared several explanation presentation styles and showed they were capable of increasing transparency (referred to as "understandability"), satisfaction and persuasiveness [30]. Ooge et al. investigated effects of explanations on recommended mathematics exercises [42], where they increased trust and transparency in the recommender system. In contrast, explanations can also have a negative impact on a recommender system's persuasiveness [37], effectiveness [38, 56] and efficiency [38].

3 METHODS

This section presents our study design, the generation of recommendations and explanations, the experimental procedure and details about study participants.

3.1 Study Design Overview

Figure 1 shows an overview of our study design with the pages in thefi gure corresponding to the pages in a survey. We designed a between-subject study consisting of two parts: (i) an investigation into user perceptions of cross-domain recommendations and explanations (Figure 1, Pages 2–4) and (ii) a validation of our method for generating explanations (Figure 1, Pages 5–6).

Thefi rst part of the study had four experimental conditions. Depending on the experimental condition, we asked users to select eitherfi ve books orfi ve movies that they enjoyed consuming (Figure 1, Page 2). After selecting items, participants were presented with ten book recommendations and asked to rate each recommendation in terms of their reading interest. In two of the experimental conditions, recommendations were shown with explanations (Figure 1, Page 3). Next, participants answered a questionnaire to indicate their overall impressions of these recommendations (Figure 1, Page 4).

In the second part of the study, all participants performed the same explanation validation task: we asked participants to select five books they had already read (Figure 1, Page 5) for which we generated explanations based on the books or movies selected at the start of the survey on Page 2. Participants were then asked to rate how well these explanations described the contents of each book (Figure 1, Page 6).

3.2 Data

To generate recommendations and explanations, we used the tag genome data structure that encodes items based on their relevance to various tags [26, 29]. To date, there are two publicly available tag genomes:

(1) **Tag genome for movies** [26]: contains relevance scores for the 9,734 most popular movies in MovieLens[3] with 1,084 tags.
(2) **Tag genome for books** [29]: contains relevance scores for 9,374 popular books from Goodreads[4] with 727 tags.

[3]https://movielens.org/
[4]https://www.goodreads.com/

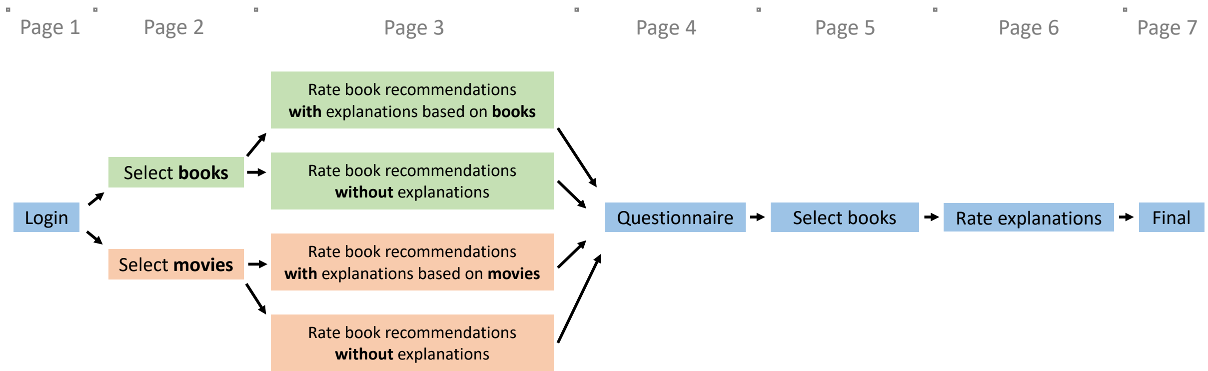

Figure 1: Overview of our between-subject study with four experimental conditions.

We used the item-tag relevance scores generated by the TagDL method due to its higher precision [26], extracting 491 common tags that appear in both tag genome data sets. As tag genome for movies lacks the covers and descriptions of movies, we extracted cover images from the HetRec data set [6] and movie descriptions from IMDB[5]. This resulted in 8,380 movies for our experiment.

3.3 Generating Recommendations

Irrespective of the experimental condition, participants were shown randomly generated recommendations. In this section, we justify this experiment design decision and describe the method used to produce recommendations.

3.3.1 Rationale for Randomly Generated Recommendations. In our study, we were not interested in how users responded to actual cross-domain recommendations, but in how the cross-domain setting itself impacts user perception of recommendations. We, therefore, needed to ensure that the quality of recommendations was constant across all experimental conditions so that any observed differences in user responses were solely attributable to their expectations of different recommendation scenarios. This implied that we could not use personalized recommendations as different combinations of data sources, algorithms and users would impact their quality.

Instead, we decided to use randomly generated recommendations, but explicitly told participants that recommendations were based on the information they provided about movies or books at the beginning of the study. By using randomly generated recommendations, we needed to make the assumption that recommendation quality would not affect our overallfi ndings. Namely, that users would respond authentically to recommendations even if those recommendations had limited similarity to the items they provided information about. As there were numerous significant differences between experimental conditions and we were able to reproduce results from the (single-domain) recommendation explanation literature [14, 30, 42], this assumption appears justified. Furthermore, generating recommendations randomly is a common strategy employed in user studies that evaluate recommendation explanations [8, 15, 33, 53]. Lastly, we acknowledge that "perfect" recommendations will likely override any misgivings users have about the origin of recommendations, but this was not the research question we were seeking to answer.

3.3.2 Generating Random Recommendations. We used tag genome for books to create randomized lists of book recommendations. We generated lists of book recommendations using the following procedure:

(1) **Filter popular items:** To ensure that participants' feedback primarily reflects the recommendation setting rather than their recollections of books, following [58], we wanted to decrease the chance that participants had already consumed recommended items. We removed the top 25% of books ranked by popularity (i.e., number of ratings [21, 67]) from tag genome for books, which left 7,030 books in our candidate list.
(2) **Filter unpopular items:** Following [3, 58], we alsofi ltered out unpopular books, so that if participants indicated low interest in a book, it was due to the information we provided and not, for example, because the book was about an unappealing or obscure topic. We, therefore, removed the bottom 25% of books in terms of popularity because these books are more likely to be uninteresting to the average user, leaving 4,677 books in our candidate list.
(3) **Generate diversified lists:** We wanted each recommendation list to cover a wide range of topics to allow users to express a range of interest levels. We diversified recommendation lists using the topic diversification algorithm [71] using the whole candidate list as input. The algorithm starts with the most relevant item and iteratively populates the list of recommendations based on relevance and average dissimilarity to items already in the list. As we selected items randomly, we did not use the relevance term. We used cosine similarity as the similarity measure. Each book in the candidate list was used as a starting point for generating a list of 10 diversified recommendations, resulting in 4,677 recommendation lists.
(4) **Filter niche items:** Some books were highly dissimilar to all other books and, therefore, appeared in up to 79% of diversified recommendation lists. As these books tended to represent niche interests, we removed the top 10% of books ordered by

[5]https://www.imdb.com/

descending frequency of the number of lists they appeared in and re-generated the recommendation lists. This resulted in 4,209 diversified lists of books for use as recommendations.

(5) **Select recommendation list:** For each participant, we randomly sampled one of the pre-generated lists of book recommendations without replacement.

3.4 Generating Recommendation Explanations

In this section, we describe our method for generating recommendation explanations. We applied tag-based explanations, incorporating a contrastive feature—showing both common and uncommon tags in recommended items. This choice is based on its prevalence in the literature [52] and its effectiveness in supporting users to compare different recommendations [10, 15].

3.4.1 Tag Selection. To generate explanations, we constructed a combined tag genome by merging the two genomes based on the 491 common tags. We used exact matches between tag names for merging, manually removing tags that were not descriptive, such as "interesting", "awesome" and "story", resulting in a set of 469 tags. To manually remove tags, two annotatorsfl agged tags for exclusion with an inter-rater agreement of 0.581 (Cohen's kappa, $P < 10^{-10}$), corresponding to moderate agreement. Both annotators agreed on thefi nal list of tags to exclude. Disagreements were generally due to tags having different semantics in different contexts, e.g., "epic" may refer to epic poetry in the context of books or colloquially as a synonym for "awesome".

3.4.2 Generating Explanations. In our study, participants selected eitherfi ve books orfi ve movies depending on the experimental condition and received recommendations of books along with explanations in the following format (inspired by [32, 53, 66]):

> This book has the following themes in common with your favourite **[items]**: **[common tags]**, and the following themes that are uncommon to those **[items]**: **[uncommon tags]**

In this format, "[items]" is replaced with either "movies" or "books", depending on the experimental condition, while "[common tags]" and "[uncommon tags]" correspond to two lists of tags. Common tags are those with high relevance in a participant's favourite items and the current book recommendation, whereas uncommon tags are tags with high relevance to only the current book recommendation. For each tag, we calculated two scores: $score_c$ and $score_u$ to measure how common and uncommon each tag is between favourite items and recommendations. For common and uncommon tags, we displayed the top-3 tags with the highest scores.

To generate scores for common tags, we used term frequency-inverse document frequency ($tfidf$) as follows:

$$score_c(t, b, prof) = rel(t, prof) * tfidf(t, b), \qquad (1)$$

where $prof$ corresponds to a user profile that we created by calculating the mean relevance for each tag across the favourite items selected by the user, and $rel(t, prof)$ is the relevance of tag t in the user profile. $tfidf$ is calculated as follows:

$$tfidf(t, b) = rel(t, b) * log\left(\frac{||B||}{\sum_{i \in B} rel(t, i)}\right), \qquad (2)$$

where $rel(t, b)$ represents term frequency and corresponds to the relevance of tag t to book b. The second part of the equation corresponds to inverse document frequency, where B is the set of all the books in the data set.

To generate scores for uncommon tags, we simply subtract the relevance scores in the user profile from those of the book recommendation:

$$score_u(t, b, prof) = rel(t, b) - rel(t, prof). \qquad (3)$$

We used this approach because the methods proposed in the literature tend to assume the existence of ordinal ratings (e.g., [8, 40, 41]), whereas we only have unary ratings in our study (i.e., participants only selected items they liked). Following [58], we additionally verified that this approach generates explanations of acceptable quality (detailed in Section 4).

3.5 Procedure

We designed our study for Amazon Mechanical Turk[6] (MTurk). In general, MTurk has been shown to be a reliable platform for conducting user studies [4, 36, 43]. It has been shown that many of the risks associated with user studies, such as multiple responses from a single participant and experimenter effects, are low for MTurk [43]. Moreover, MTurk workers represent the general population better [4, 43] and have been shown to be more attentive [22] than participants recruited through traditional subject pools. Lastly, MTurk allows for the recruitment of certain categories of participants, such as book readers, movie goers and music fans.

3.5.1 User Profile Creation. On Page 1 of our survey (Figure 1), participants were presented with details of the experiment and asked to provide their MTurk ID. On Page 2, participants were randomly assigned to one of the four experimental conditions and asked to select eitherfi ve books they enjoyed reading orfi ve movies they enjoyed watching. Participants couldfi nd items using a search box and could not transition to the next page until they had selected exactlyfi ve items.

We selectedfi ve as the number of items study participants provided at the start of the experiment by manually assessing the quality of generated explanations for different numbers of items. To decide on the range of the number of items, we relied on past literature, where this number varied between 3 [5, 39] and 15 [19]. To evaluate this design decision, we verified the quality of explanations with participants in the study. According to the results in Section 4, the quality of explanations is acceptable.

3.5.2 Rating Book Recommendations. On Page 3 (Figure 1), depending on the experimental condition participants were assigned to, each participant received 10 book recommendations with or without explanations. The instructions explicitly stated that these recommendations were generated based on the items the user selected in the previous step. For each condition, the recommendations came from a randomly sampled list of diversified books (see Section 3.3), while the explanations were generated based on the items participants selected on Page 2 (see Section 3.4.2).

Each book recommendation was accompanied by the book's title, a description from tag genome for books [29] and, in corresponding

[6]https://www.mturk.com/

experimental conditions, explanations. We did not include book covers as they were missing for a large subset of books. However, we included book covers during the profile building step. For each book, the user was asked to rate the statement: *"I am interested in reading this book"*. The user could provide the rating on a 5-point Likert response scale ("Strongly disagree"–"Strongly agree"). Participants were also asked to indicate if they had already read the book: *"Have you read this book?"*. The available answers to this question were "Yes", "No" and "No, but I know the plot".

3.5.3Qu estionnaire. After rating book recommendations, participants were asked to rate their level of agreement to the statements in Table 1 (Figure 1, Page 5), which were adapted from the ResQue questionnaire [46] and studies on explanations in recommender systems [39, 45, 58]. The available responses were on a 5-point Likert response scale ("Strongly disagree"–"Strongly agree") and an "I don't know" option.

3.5.4 Recommendation Explanation Validation. We validated the quality of our explanations by asking participants to select 1-5 books that they had already read and were, therefore, able to judge whether we provided meaningful recommendation explanations for those books (Figure 1, Page 5). If participants had previously been required to select books during profile building, then those books were removed from the list of available books.

On Page 6 (Figure 1), participants were asked to rate their agreement with the common and uncommon tags generated for each book selected on the previous page (Figure 1, Page 5) based on the items they selected in the beginning of the study (Figure 1, Page 2). The available ratings were on a 5-point Likert response scale ("Strongly disagree"–"Strongly agree").

3.5.5Qu ality Control. To identify andfi lter out responses of participants who failed to pay close attention to the survey, we implemented a common strategy known as catch-trials [43]. First, we included two fake items, which we added to the list of available items on Page 2 (Figure 1) under the heading "Recent Picks". The list was displayed as the default set of items under the search box. Second, we included two trick questions in the questionnaire (Figure 1, Page 4), including *"To prove that you have read this statement, please answer Disagree"*.

Participants who selected fake items or answered the control questions incorrectly were shown a message that the survey is disabled and they will not receive payment even if they retake the survey. We excluded these participants from our data set. We also removed participants who provided suspicious responses. For example, we excluded responses from participants who indicated they had read all of the recommended books or completed the survey in less than three minutes, which we estimated as the minimum amount of time necessary to complete the study.

3.6 Participants

We published our survey on Amazon Mechanical Turk to participants with an approval rating of at least 80% (particularly attentive participants) and who had indicated they buy books online. After removing participants who provided suspicious responses, our data set contained responses from 237 participants assigned to experimental conditions as follows: 57 selected books and received no explanations, 60 selected books and received explanations, 63 selected movies and received no explanations, and 57 selected books and received no explanations.

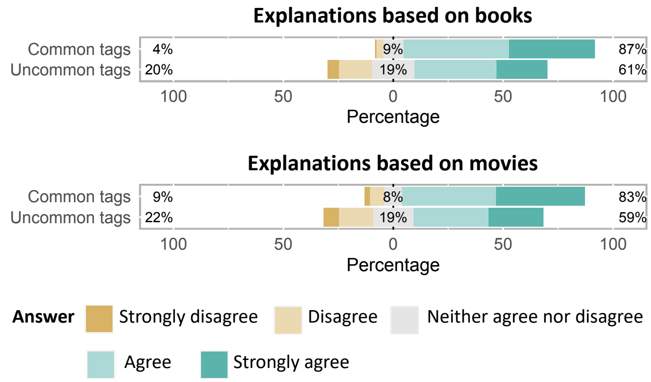

Figure 2: Participant agreement with explanations for known books based on either books (top) or movies (bottom). The values correspond to percentages of participant responses.

4 EXPLANATION VALIDATION

In the second half of the study, participants were asked to rate their agreement with the explanations we generated for books they had already read. The explanations were based on either books or movies dependent on which experimental condition participants were assigned to at the beginning of the experiment (Section 3.5).

Figure 2 shows that participants agreed with explanations in a majority of cases for both common and uncommon tags for both single- (based on books) and cross-domain (based on movies) explanations. Based on these data, we concluded that the quality of explanations is sufficient to consider participant's questionnaire responses as authentic examples of users responding to the presence of recommendation explanations.

The lower user agreement with uncommon tags compared to common tags could be caused by diverse items selected by the user on Page 2 of the survey (Figure 1). If the participant selects items covering a wide range of topics, such as "romance", "horror" and "adventure", then these may cover all the core themes of a given book and lead to the uncommon tags not being truly representative of the contents.

5 RESULTS

We analysed the data from 237 users. Each user responded to the 7 questionnaire statements (Table 1), rated their interest in reading the 10 recommended books and indicated their familiarity with each recommendation. Participants reported not having read a majority of recommended books. To the question "*Have you read this book?*": 75% of responses were "No", 11% were "No, but I know the plot" and 14% were "Yes". The most common recommended book that participants had already read was "Winnie the Pooh" by A. A. Milne. The most common book selected to build the user profile was "Harry Potter and the Sorcerer's Stone" by J. K. Rowling, while the most commonly selected movie was "Titanic" directed by James Cameron.

Table 1: Statements from the questionnaire for experimental conditions without explanations. In experimental conditions with explanations, the phrase "(title, description)" is replaced with "(title, description, explanation)".

Aim	Statement	Adapted from
Transparency	The information included with each book recommendation (title, description) made it clear why the books were recommended to me	[2, 33, 34, 39, 46]
Scrutability	The information included with each book recommendation (title, description) allowed me to understand if the system made an error in making this recommendation	[45, 55]
Trust	The information included with each book recommendation (title, description) made me trust the recommendations	[2, 39, 46]
Effectiveness	The information included with recommendations (title, description) helped me decide whether I would like to read each book	[2, 15, 18, 38, 45, 58]
Persuasiveness	The information included with each book recommendation (title, description) made me interested in reading the book	[2, 34, 39, 45]
Efficiency	The information included with each book recommendation (title, description) helped me quickly decide whether I was interested in reading the book	[2, 9, 20, 38]
Satisfaction	Overall, I was satisfied with the information included with each book recommendation (title, description)	[34, 46]

5.1 User Perceptions

We used ordinal logistic regression to investigate the relationships between user perceptions of recommendations and our experimental conditions. Wefi tted seven regression models where the dependent variables corresponded to the questions from Table 1 on a 5-point Likert response scale. The independent variables encoded our experimental conditions: *crossdomain* was set to 1 if movies were used to recommend books and 0 otherwise, *explanations* was set to 1 if explanations were present and 0 otherwise, and *crossdomain:explanations* is an interaction term indicating that both cross-domain recommendations and explanations were shown. We additionally included the number of books the participant was already familiar with in a variable called *familiar*[7] on the basis of improved modelfi t using AIC (data not shown).

All models werefi tted using the maximum likelihood method implemented in the Ordinal R package. Table 2 shows the coefficients from each model and their corresponding statistical significance (based on the Wald test). We converted all coefficients from logs odds to odds ratios, as they are easier to understand. We can draw the following conclusions from these analyses:

- **Cross-domain recommendations decrease trust (RQ1).** The odds of participants trusting cross-domain recommendations was 0.51 times that of single-domain recommendations. We reiterate that both single- and cross-domain recommendations were randomly generated, so this difference in trust represents participants' perception of the provenance of recommendations and not an unbiased assessment of their relevance.
- **Explanations increase perceived transparency and scrutability for all recommendations (RQ2).** The odds that participants perceived recommendations as transparent and scrutable was respectively 2.3 and 2.6 times more likely when explanations were present. This was the case for both single- and cross-domain explanations.
- **Cross-domain explanations influenced user perceptions the same as single-domain explanations (RQ2)**. None of the interaction terms were statistically significant (see row for *crossdomain:explanations* in Table 2), suggesting any differences in user perception of single- and cross-domain explanations has a negligible effect size.
- **Familiar recommendations have a positive influence on transparency, trust and persuasiveness, but negatively impact efficiency**. An increase in the number of familiar recommendations was associated with 1.16-1.22 times increase in the odds ratios for transparency, trust and persuasiveness. We speculate that the drop in perceived efficiency could be due to users carefully considering their ratings for books they were already aware of (all participants said they bought books online and are, therefore, likely to be genuinely interested in reading books).
- **Neither cross-domain recommendations nor the presence of explanations influenced perceived effectiveness or satisfaction**. While no coefficients were statistically significant for effectiveness or satisfaction, the effect sizes may simply be too low to detect with the sample size of the present study.

5.2 Behavioural Intentions

We analysed participants' behavioural intentions, i.e., their interest in reading each book recommendation, using the same ordinal logistic regression model as the previous section with a few modifications: (i) instead of controlling for the number of familiar books, we controlled for all levels of familiarity for each book (yes/no/no, but I know the plot) as a categorical variable and (ii) we added an additional random effect for each book to model the fact that the same

[7]We considered a book to be familiar if participants replied "Yes" or "No, but I know the plot" to the question *"Have you read this book?"*

Table 2: Odds ratios of seven ordinal logistic regression models, where each dependent variable corresponds to ratings to statements regarding the aims of explanations. Each column corresponds to a different regression model. P-values are indicated in brackets. Significance codes: * < 0.05, ** < 0.01, * < 0.001.**

Independent variables	Transparency	Scrutability	Trust	Effectiveness	Persuasiveness	Efficiency	Satisfaction
familiar	**1.155**** (0.001)	1.03 (0.506)	**1.156**** (0.001)	0.931 (0.122)	**1.219***** (3×10^{-5})	**0.896*** (0.020)	1.039 (0.401)
crossdomain	0.527 (0.059)	0.899 (0.755)	**0.505*** (0.043)	0.748 (0.405)	0.706 (0.315)	0.651 (0.227)	1.275 (0.496)
explanations	**2.257*** (0.018)	**2.604**** (0.006)	1.083 (0.818)	0.731 (0.386)	0.817 (0.558)	1.259 (0.527)	1.454 (0.295)
crossdomain:explanations	1.514 (0.387)	0.449 (0.100)	1.861 (0.191)	1.868 (0.219)	1.249 (0.645)	1.642 (0.326)	0.853 (0.751)

Table 3: Odds ratios of the ordinal logistic regression model, where the dependent variable corresponds to ratings participants gave to express their interest in reading each book recommendation. Significance codes: * < 0.05, * < 0.001.**

Independent variable	Coefficient
familiar (knows the plot)	**2.482***** (4×10^{-13})
familiar (read the book)	**5.929***** (2×10^{-16})
crossdomain	**0.701***** (7×10^{-4})
explanations	**0.783*** (0.022)
crossdomain:explanations	**1.662***** (7×10^{-4})

book can appear in multiple recommendation lists. These changes resulted in the best modelfi t in terms of AIC (data not shown). Table 3 shows the coefficients (odds ratios) of the model, all of which are statistically significant. We can draw several conclusions from these results:

- **Cross-domain recommendations decrease interest (RQ1).** The odds of participants being interested in cross-domain recommendations was 0.7 times that of single-domain recommendations. Similar to trust, the cross-domain setting itself diminishes user interest in recommendations independent of recommendation quality.
- **Explanations increase interest in cross-domain recommendations, but not to the extent of single-domain recommendations without explanations (RQ2).** The odds of participants being interested in cross-domain recommendations with explanations was 0.91 times that of single-domain recommendations without explanations[8].
- **Explanations decrease interest in single-domain recommendations (RQ2).** Despite explanations increasing interest in cross-domain recommendations, the odds of participants being interested in single-domain recommendations with explanations was only 0.78 times that of single-domain recommendations without explanations. This could be caused by our inclusion of uncommon tags in explanations (which we speculate might also contribute to the increased scrutability highlighted in Table 2).
- **Familiarity with recommendations increases interest.** Both familiarity with a book's plot and having read the book increases participants' stated interest in the book. Furthermore, these effects are strong, e.g., the odds of being interested in any recommendation is 5.93 times more likely if the participant has already read the book compared with not even being familiar with the book's plot.

6 DISCUSSION

We presented a study to investigate the impact of cross-domain recommendations and recommendation explanations on user perceptions and behavioural intentions. To the best of our knowledge, this was thefi rst user study on cross-domain recommendation and on cross-domain explanations. We conducted a user study on Amazon Mechanical Turk that was designed to avoid several experimental confounders. These included a between-subject design to ensure that each participant knew unambiguously what data was available for generating recommendations and using randomly sampled recommendation lists to control for recommendation quality. Further, we verified that both single- and cross-domain explanations provided a credible rationale for recommendations.

We found that cross-domain recommendations decrease perceived trust and user interest in recommendations compared to the single-domain setting. While explanations increased user interest in cross-domain explanations, it was still lower overall than for single-domain recommendations without explanations. In general, recommendation explanations increased users' perceptions of transparency and scrutability for both single- and cross-domain explanations. In this section, we discuss the implications of these results and the limitations of our study.

6.1 Cross-domain Recommendations

Our results showed that simply telling users recommendations were based on information from a different domain significantly altered their perception of the recommender system. We identified a significant difference in user interest between cross-domain and single-domain recommendations, with the odds of being interested in cross-domain recommendations 0.7 times that of single-domain recommendations (see Table 3). As we used random recommendations to control for recommendation quality between experimental settings, we conclude that the difference between behavioural intentions is based solely on users' belief that recommendations were cross-domain. This conclusion is further supported by how cross-domain recommendations were associated with decreased trust in

[8]This calculation required the odds ratios for *crossdomain*, *explanations* and *crossdomain:explanations* from Table 3 to give $0.701 \times 0.783 \times 1.662 = 0.91$ (we multiply because the original coefficients were in terms of log odds, but these numbers are odds ratios).

the recommender system (see Table 2). Indeed, one of the study participants commented that *"... It is also based on a shaky premise: that the user's taste in movies correlates with the user's taste in books. Is this a justified assumption?"*, which implies that even when cross-domain recommendations are highly relevant, users could still be dismissive of those recommendations due to their provenance.

In our study, we investigated a scenario where cross-domain recommendation is used to overcome cold start, however, there are other scenarios that could also have a detrimental effect on trust or perceived performance. For example, when cross-domain recommendation is used to reduce data sparsity, it could have similar repercussions as those observed in our study if users are aware of the lack of data in the target domain. Similarly, cross-domain recommenders based on other definitions of domain could provide relevant recommendations, but still give users a negative impression of the system if they defy users' expectations. In different attribute-level domains, for example, this could take the form of users who mainly consume a single movie genre (e.g., sciencefi ction) receiving recommendations from a genre they would otherwise avoid (e.g., romantic comedy). In instances such as this, it is difficult for users to evaluate whether a recommendation is potentially serendipitous [27, 28] or an error from the recommender system.

We believe that negative user perceptions are an important confounder that should be taken into account when evaluating cross-domain recommender systems. Ourfi ndings motivate further experiments on whether highly relevant recommendations can overcome negative user perceptions and, if so, how much better cross-domain recommendations need to be in order to be perceived as similar in quality to those from single-domain recommenders. Of course, in scenarios where recommendation quality is bounded, such as during cold start, we must rely on additional interventions (e.g., explanations) to improve user perceptions of recommendations.

6.2 Cross-domain Explanations

Previous research has demonstrated the practical importance of explanations and their ability to impact user perception of recommendations. Our study design had four experimental conditions to allow us to disentangle the effect of recommendations being single- or cross-domain from whether explanations were present or not by modelling separate interaction effects (see Tables 2 and 3).

For both single- and cross-domain recommendations, we showed that explanations had a positive impact on user perceptions of transparency (understanding why a recommendation was given) and scrutability (ability to tell if the system made an error). These results corroborate and compliment previous case-control studies from the literature (see Section 2.2.2), where increased transparency was the most commonfi nding [14, 30, 42], and increased scrutability was not previously observed. We found no additional effects on user perceptions of explanations on cross-domain recommendations as none of the interaction terms in Table 2 were statistically significant. This lack of interaction effects implies that users perceive explanations as independent from the source of recommendations (which is not the case as they were derived from the same data), but it is unclear whether this would change as users actually consumed cross-domain recommendations as a result of explanations.

In general, explanations decreased user interest in recommendations (see Table 3), which can be interpreted as being effective because additional information helped users to make informed decisions (our explanations included uncommon tags, i.e., tags associated with a recommended item that were absent from the user profile), which is supported by users reporting improved transparency and similar observations in prior work [37]. For cross-domain recommendations, explanations partially mitigated the decrease in user interest in recommended items. However, the positive impact of cross-domain explanations was insufficient to match user interest in single-domain recommendations without explanations. Indeed, we speculate that the impact of explanations is at least partially dependent on users' initial perceptions of recommendations, with cross-domain explanations providing a net improvement due to users' lower baseline interest in cross-domain recommendations. This feeds into our conclusions about cross-domain recommendations, with the presence of explanations being one interface component that can be used to make up for the shortfall in user interest during cold start or data sparsity. It is an open research question, however, whether there are types of explanation that are particularly effective in cross-domain recommendation but underperform for single-domain recommendations and vice versa.

6.3 Limitations and Future Work

Our study has several limitations. First, our study focused on one particular cross-domain recommendation scenario and it is unclear whether our results generalise to other scenarios, such as using cross-domain recommendations to overcome data sparsity or different definitions of domain. Second, our experiment was designed to measure the impact of the cross-domain setting and explanations on users. However, single-domain recommenders are likely to outperform cross-domain recommenders with the same quantity of data. We speculate that the difference in perception between single- and cross-domain recommendations will be more pronounced in real-life. Third, our study is confined to participants who buy books online, which may limit the applicability of ourfi ndings to other kinds of participants, such as those who are less enthusiastic about reading.

In future work, we plan to investigate how different definitions of domain impact user perceptions of cross-domain recommendations. For example, item-level domains, such as movies and books, are unrelated to one another (aside from both being forms of entertainment), but type-level domains can be complementary, such as purchase recommendations for different kinds of clothing items (e.g., due to being related in terms of style). We also want to explore how different explanation styles affect user acceptance of recommendations and whether it is possible to persuade users that cross-domain recommendations are of equal quality to single-domain recommendations.

ACKNOWLEDGMENTS

This work has been supported by Helsinki Institute for Information Technology HIIT.

REFERENCES

[1] Fernando Amat, Ashok Chandrashekar, Tony Jebara, and Justin Basilico. 2018. Artwork personalization at netflix. In *Proceedings of the 12th ACM Conference*

on Recommender Systems (Vancouver, British Columbia, Canada) *(RecSys '18)*. Association for Computing Machinery, New York, NY, USA, 487–488. https://doi.org/10.1145/3240323.3241729

[2] Krisztian Balog and Filip Radlinski. 2020. Measuring Recommendation Explanation Quality: The Conflicting Goals of Explanations. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, China) *(SIGIR '20)*. Association for Computing Machinery, New York, NY, USA, 329–338. https://doi.org/10.1145/3397271.3401032

[3] Krisztian Balog, Filip Radlinski, and Shushan Arakelyan. 2019. Transparent, Scrutable and Explainable User Models for Personalized Recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) *(SIGIR'19)*. Association for Computing Machinery, New York, NY, USA, 265–274. https://doi.org/10.1145/3331184.3331211

[4] Adam J. Berinsky, Gregory A. Huber, and Gabriel S. Lenz. 2012. Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis* 20, 3 (2012), 351–368. https://doi.org/10.1093/pan/mpr057

[5] Mustafa Bilgic and Raymond J Mooney. 2005. Explaining recommendations: Satisfaction vs. promotion. In *Beyond personalization workshop, IUI*, Vol. 5. 153. https://api.semanticscholar.org/CorpusID:3228904

[6] Iván Cantador, Peter Brusilovsky, and Tsvi Kuflik. 2011. 2nd Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011). In *Proceedings of the 5th ACM conference on Recommender systems* (Chicago, IL, USA) *(RecSys 2011)*. ACM, New York, NY, USA.

[7] Iván Cantador and Paolo Cremonesi. 2014. Tutorial on Cross-domain Recommender Systems. In *Proceedings of the 8th ACM Conference on Recommender Systems* (Foster City, Silicon Valley, California, USA). ACM, New York, NY, USA, 401–402. https://doi.org/10.1145/2645710.2645777

[8] Shuo Chang, F. Maxwell Harper, and Loren Gilbert Terveen. 2016. Crowd-Based Personalized Natural Language Explanations for Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) *(RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 175–182. https://doi.org/10.1145/2959100.2959153

[9] Mohamed Amine Chatti, Mouadh Guesmi, Laura Vorgerd, Thao Ngo, Shoeb Joarder, Qurat Ul Ain, and Arham Muslim. 2022. Is More Always Better? The Effects of Personal Characteristics and Level of Detail on the Perception of Explanations in a Recommender System. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (Barcelona, Spain) *(UMAP '22)*. Association for Computing Machinery, New York, NY, USA, 254–264. https://doi.org/10.1145/3503252.3531304

[10] Li Chen and Feng Wang. 2017. Explaining Recommendations Based on Feature Sentiments in Product Reviews. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) *(IUI '17)*. Association for Computing Machinery, New York, NY, USA, 17–28. https://doi.org/10.1145/3025171.3025173

[11] Xu Chen, Yongfeng Zhang, and Ji-Rong Wen. 2022. Measuring" Why" in Recommender Systems: a Comprehensive Survey on the Evaluation of Explainable Recommendation. *arXiv preprint arXiv:2202.06466* abs/2202.06466 (2022).

[12] Maurizio Ferrari Dacrema, Iván Cantador, Ignacio Fernández-Tobías, Shlomo Berkovsky, and Paolo Cremonesi. 2012. Design and evaluation of cross-domain recommender systems. In *Recommender Systems Handbook*. Springer, 485–516.

[13] Pedram Daee, Joel Pyykkö, Dorota Glowacka, and Samuel Kaski. 2016. Interactive Intent Modeling from Multiple Feedback Domains. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (Sonoma, California, USA) *(IUI '16)*. Association for Computing Machinery, New York, NY, USA, 71–75. https://doi.org/10.1145/2856767.2856803

[14] Vicente Dominguez, Pablo Messina, Ivania Donoso-Guzmán, and Denis Parra. 2019. The effect of explanations and algorithmic accuracy on visual recommender systems of artistic images. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 408–416. https://doi.org/10.1145/3301275.3302274

[15] Tim Donkers, Timm Kleemann, and Jürgen Ziegler. 2020. Explaining recommendations by means of aspect-based transparent memories. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (Cagliari, Italy) *(IUI '20)*. Association for Computing Machinery, New York, NY, USA, 166–176. https://doi.org/10.1145/3377325.3377520

[16] Manuel Enrich, Matthias Braunhofer, and Francesco Ricci. 2013. Cold-Start Management with Cross-Domain Collaborative Filtering and Tags. In *E-Commerce and Web Technologies*, Christian Huemer and Pasquale Lops (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 101–112.

[17] Ignacio Fernández-Tobías, Iván Cantador, Marius Kaminskas, and Francesco Ricci. 2012. Cross-domain recommender systems: A survey of the state of the art. In *Spanish Conference on Information Retrieval*.

[18] Jingyue Gao, Xiting Wang, Yasha Wang, and Xing Xie. 2019. Explainable Recommendation through Attentive Multi-View Learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01, 3622–3629. https://doi.org/10.1609/aaai.v33i01.33013622

[19] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. 2014. How should I explain? A comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies* 72, 4 (2014), 367–382. https://doi.org/10.1016/j.ijhcs.2013.12.007

[20] Mouadh Guesmi, Mohamed Amine Chatti, Laura Vorgerd, Thao Ngo, Shoeb Joarder, Qurat Ul Ain, and Arham Muslim. 2022. Explaining User Models with Different Levels of Detail for Transparent Recommendation: A User Study. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (Barcelona, Spain) *(UMAP '22 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 175–183. https://doi.org/10.1145/3511047.3537685

[21] F. Maxwell Harper, Funing Xu, Harmanpreet Kaur, Kyle Condiff, Shuo Chang, and Loren Terveen. 2015. Putting Users in Control of their Recommendations. In *Proceedings of the 9th ACM Conference on Recommender Systems* (Vienna, Austria) *(RecSys '15)*. Association for Computing Machinery, New York, NY, USA, 3–10. https://doi.org/10.1145/2792838.2800179

[22] David J Hauser and Norbert Schwarz. 2016. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior research methods* 48, 1 (2016), 400–407. https://doi.org/10.3758/s13428-015-0578-z

[23] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining collaborativefi ltering recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (Philadelphia, Pennsylvania, USA) *(CSCW '00)*. Association for Computing Machinery, New York, NY, USA, 241–250. https://doi.org/10.1145/358916.358995

[24] SeongKu Kang, Junyoung Hwang, Dongha Lee, and Hwanjo Yu. 2019. Semi-Supervised Learning for Cross-Domain Recommendation to Cold-Start Users. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (Beijing, China) *(CIKM '19)*. Association for Computing Machinery, New York, NY, USA, 1563–1572. https://doi.org/10.1145/3357384.3357914

[25] Muhammad Murad Khan, Roliana Ibrahim, and Imran Ghani. 2017. Cross Domain Recommender Systems: A Systematic Literature Review. *ACM Comput. Surv.* 50, 3, Article 36 (jun 2017), 34 pages. https://doi.org/10.1145/3073565

[26] Denis Kotkov, Alexandr Maslov, and Mats Neovius. 2021. Revisiting the Tag Relevance Prediction Problem. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, Canada) *(SIGIR '21)*. Association for Computing Machinery, New York, NY, USA, 1768–1772. https://doi.org/10.1145/3404835.3463019

[27] Denis Kotkov, Alan Medlar, and Dorota Glowacka. 2023. Rethinking Serendipity in Recommender Systems. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval* (Austin, TX, USA) *(CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 383–387. https://doi.org/10.1145/3576840.3578310

[28] Denis Kotkov, Alan Medlar, Triin Kask, and Dorota Glowacka. 2024. The Dark Matter of Serendipity in Recommender Systems. In *Proceedings of the 2024 Conference on Human Information Interaction and Retrieval* (Sheffield, United Kingdom) *(CHIIR '24)*. Association for Computing Machinery, New York, NY, USA, 108–118. https://doi.org/10.1145/3627508.3638342

[29] Denis Kotkov, Alan Medlar, Alexandr Maslov, Umesh Raj Satyal, Mats Neovius, and Dorota Glowacka. 2022. The Tag Genome Dataset for Books. In *ACM SIGIR Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) *(CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 353–357. https://doi.org/10.1145/3498366.3505833

[30] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2017. User Preferences for Hybrid Explanations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (Como, Italy) *(RecSys '17)*. Association for Computing Machinery, New York, NY, USA, 84–88. https://doi.org/10.1145/3109859.3109915

[31] Anil Kumar, Nitesh Kumar, Muzammil Hussain, Santanu Chaudhury, and Sumeet Agarwal. 2014. Semantic clustering-based cross-domain recommendation. In *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*. 137–141. https://doi.org/10.1109/CIDM.2014.7008659

[32] Trung-Hoang Le and Hady W. Lauw. 2021. Explainable Recommendation with Comparative Constraints on Product Aspects. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (Virtual Event, Israel) *(WSDM '21)*. Association for Computing Machinery, New York, NY, USA, 967–975. https://doi.org/10.1145/3437963.3441754

[33] Qing Li, Sharon Chu, Nanjie Rao, and Mahsan Nourani. 2020. Understanding the Effects of Explanation Types and User Motivations on Recommender System Use. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 8, 1, 83–91. https://doi.org/10.1609/hcomp.v8i1.7466

[34] Hongyu Lu, Weizhi Ma, Yifan Wang, Min Zhang, Xiang Wang, Yiqun Liu, Tat-Seng Chua, and Shaoping Ma. 2023. User Perception of Recommendation Explanation: Are Your Explanations What Users Need? *ACM Trans. Inf. Syst.* 41, 2, Article 48 (jan 2023), 31 pages. https://doi.org/10.1145/3565480

[35] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-Domain Recommendation: An Embedding and Mapping Approach. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*.

2464–2470. https://doi.org/10.24963/ijcai.2017/343
[36] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23. https://doi.org/10.3758/s13428-015-0578-z
[37] Martijn Millecamp, Nyi Nyi Htun, Cristina Conati, and Katrien Verbert. 2019. To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 397–407. https://doi.org/10.1145/3301275.3302313
[38] Martijn Millecamp, Nyi Nyi Htun, Cristina Conati, and Katrien Verbert. 2020. What's in a User? Towards Personalising Transparency for Music Recommender Interfaces. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) *(UMAP '20)*. Association for Computing Machinery, New York, NY, USA, 173–182. https://doi.org/10.1145/3340631.3394844
[39] Cataldo Musto, Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. 2019. Justifying Recommendations through Aspect-based Sentiment Analysis of Users Reviews. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* (Larnaca, Cyprus) *(UMAP '19)*. Association for Computing Machinery, New York, NY, USA, 4–12. https://doi.org/10.1145/3320435.3320457
[40] Tien T. Nguyen, Daniel Kluver, Ting-Yu Wang, Pik-Mai Hui, Michael D. Ekstrand, Martijn C. Willemsen, and John Riedl. 2013. Rating support interfaces to improve user experience and recommender accuracy. In *Proceedings of the 7th ACM Conference on Recommender Systems* (Hong Kong, China) *(RecSys '13)*. Association for Computing Machinery, New York, NY, USA, 149–156. https://doi.org/10.1145/2507157.2507188
[41] Tien T. Nguyen and John Riedl. 2013. Predicting Users' Preference from Tag Relevance. In *User Modeling, Adaptation, and Personalization*, Sandra Carberry, Stephan Weibelzahl, Alessandro Micarelli, and Giovanni Semeraro (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 274–280.
[42] Jeroen Ooge, Shotallo Kato, and Katrien Verbert. 2022. Explaining Recommendations in E-Learning: Effects on Adolescents' Trust. In *Proceedings of the 27th International Conference on Intelligent User Interfaces* (Helsinki, Finland) *(IUI '22)*. Association for Computing Machinery, New York, NY, USA, 93–105. https://doi.org/10.1145/3490099.3511140
[43] Gabriele Paolacci, Jesse Chandler, and Panagiotis G. Ipeirotis. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5, 5 (2010), 411–419. https://doi.org/10.1017/S1930297500002205
[44] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 311–318.
[45] Gustavo Penha, Eyal Krikon, and Vanessa Murdock. 2022. Pairwise Review-Based Explanations for Voice Product Search. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) *(CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 300–304. https://doi.org/10.1145/3498366.3505828
[46] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (Chicago, Illinois, USA) *(RecSys '11)*. Association for Computing Machinery, New York, NY, USA, 157–164. https://doi.org/10.1145/2043932.2043962
[47] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender Systems: Introduction and Challenges. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, Boston, MA, 1–34. https://doi.org/10.1007/978-1-4899-7637-6_1
[48] Shaghayegh Sahebi and Peter Brusilovsky. 2013. Cross-Domain Collaborative Recommendation in a Cold-Start Context: The Impact of User Profile Size on the Quality of Recommendation. In *User Modeling, Adaptation, and Personalization*, Sandra Carberry, Stephan Weibelzahl, Alessandro Micarelli, and Giovanni Semeraro (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 289–295.
[49] Bracha Shapira, Lior Rokach, and Shirley Freilikhman. 2013. Facebook single and cross domain data for recommendation systems. *User Modeling and User-Adapted Interaction* 23, 2 (2013), 211–247. https://doi.org/10.1007/s11257-012-9128-x
[50] Yue Shi, Martha Larson, and Alan Hanjalic. 2011. Tags as Bridges between Domains: Improving Recommendation with Tag-Induced Cross-Domain Collaborative Filtering. In *User Modeling, Adaption and Personalization*, Joseph A. Konstan, Ricardo Conejo, José L. Marzo, and Nuria Oliver (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 305–316.
[51] Nava Tintarev and Judith Masthoff. 2007. A Survey of Explanations in Recommender Systems. In *2007 IEEE 23rd International Conference on Data Engineering Workshop*. 801–810. https://doi.org/10.1109/ICDEW.2007.4401070
[52] Nava Tintarev and Judith Masthoff. 2011. Designing and Evaluating Explanations for Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor (Eds.). Springer US, Boston, MA, 479–510. https://doi.org/10.1007/978-0-387-85820-3_15
[53] Nava Tintarev and Judith Masthoff. 2012. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction* 22, 4 (2012), 399–439. https://doi.org/10.1007/s11257-011-9117-5
[54] Thi Ngoc Trang Tran, Viet Man Le, Muesluem Atas, Alexander Felfernig, Martin Stettinger, and Andrei Popescu. 2021. Do Users Appreciate Explanations of Recommendations? An Analysis in the Movie Domain. In *Proceedings of the 15th ACM Conference on Recommender Systems* (Amsterdam, Netherlands) *(RecSys '21)*. Association for Computing Machinery, New York, NY, USA, 645–650. https://doi.org/10.1145/3460231.3478859
[55] Chun-Hua Tsai and Peter Brusilovsky. 2019. Evaluating Visual Explanations for Similarity-Based Recommendations: User Perception and Performance. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* (Larnaca, Cyprus) *(UMAP '19)*. Association for Computing Machinery, New York, NY, USA, 22–30. https://doi.org/10.1145/3320435.3320465
[56] Chun-Hua Tsai and Peter Brusilovsky. 2019. Explaining recommendations in an interactive hybrid social recommender. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 391–396. https://doi.org/10.1145/3301275.3302318
[57] Kosetsu Tsukuda and Masataka Goto. 2020. Explainable Recommendation for Repeat Consumption. In *Proceedings of the 14th ACM Conference on Recommender Systems* (Virtual Event, Brazil) *(RecSys '20)*. Association for Computing Machinery, New York, NY, USA, 462–467. https://doi.org/10.1145/3383313.3412230
[58] Jesse Vig, Shilad Sen, and John Riedl. 2009. Tagsplanations: explaining recommendations using tags. In *Proceedings of the 14th International Conference on Intelligent User Interfaces* (Sanibel Island, Florida, USA) *(IUI '09)*. Association for Computing Machinery, New York, NY, USA, 47–56. https://doi.org/10.1145/1502650.1502661
[59] Alexandra Vultureanu-Albişi and Costin Bădică. 2022. A survey on effects of adding explanations to recommender systems. *Concurrency and Computation: Practice and Experience* 34, 20 (2022), e6834. https://doi.org/10.1002/cpe.6834
[60] Xinghua Wang, Zhaohui Peng, Senzhang Wang, Philip S. Yu, Wenjing Fu, and Xiaoguang Hong. 2018. Cross-Domain Recommendation for Cold-Start Users via Neighborhood Based Feature Mapping. In *Database Systems for Advanced Applications*, Jian Pei, Yannis Manolopoulos, Shazia Sadiq, and Jianxin Li (Eds.). Springer International Publishing, Cham, 158–165.
[61] Pinata Winoto and Tiffany Tang. 2008. If You Like the Devil Wears Prada the Book, Will You also Enjoy the Devil Wears Prada the Movie? A Study of Cross-Domain Recommendations. *New Generation Computing* 26, 3 (2008), 209–225. https://doi.org/10.1007/s00354-008-0041-0
[62] Deqing Yang, Jingrui He, Huazheng Qin, Yanghua Xiao, and Wei Wang. 2015. A Graph-based Recommendation across Heterogeneous Domains. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* (Melbourne, Australia) *(CIKM '15)*. Association for Computing Machinery, New York, NY, USA, 463–472. https://doi.org/10.1145/2806416.2806523
[63] Tianzi Zang, Yanmin Zhu, Haobing Liu, Ruohan Zhang, and Jiadi Yu. 2022. A Survey on Cross-domain Recommendation: Taxonomies, Methods, and Future Directions. *ACM Trans. Inf. Syst.* 41, 2, Article 42 (dec 2022), 39 pages. https://doi.org/10.1145/3548455
[64] Yongfeng Zhang. 2014. Browser-oriented universal cross-site recommendation and explanation based on user browsing logs. In *Proceedings of the 8th ACM Conference on Recommender Systems* (Foster City, Silicon Valley, California, USA) *(RecSys '14)*. Association for Computing Machinery, New York, NY, USA, 433–436. https://doi.org/10.1145/2645710.2653367
[65] Yongfeng Zhang and Xu Chen. 2020. Explainable Recommendation: A Survey and New Perspectives. *Foundations and Trends® in Information Retrieval* 14, 1 (2020), 1–101. https://doi.org/10.1561/1500000066
[66] Yongfeng Zhang, Guokun Lai, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. 2014. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval* (Gold Coast, Queensland, Australia) *(SIGIR '14)*. Association for Computing Machinery, New York, NY, USA, 83–92. https://doi.org/10.1145/2600428.2609579
[67] Yuan Cao Zhang, Diarmuid Ó Séaghdha, Daniele Quercia, and Tamas Jambor. 2012. Auralist: introducing serendipity into music recommendation. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining* (Seattle, Washington, USA) *(WSDM '12)*. Association for Computing Machinery, New York, NY, USA, 13–22. https://doi.org/10.1145/2124295.2124300
[68] Cheng Zhao, Chenliang Li, Rong Xiao, Hongbo Deng, and Aixin Sun. 2020. CATN: Cross-Domain Recommendation for Cold-Start Users via Aspect Transfer Network. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, China) *(SIGIR '20)*. Association for Computing Machinery, New York, NY, USA, 229–238. https://doi.org/10.1145/3397271.3401169
[69] Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu. 2021. Cross-domain recommendation: challenges, progress, and prospects. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI 2021)*. 4721–4728. https://doi.org/10.24963/ijcai.2021/639
[70] Yongchun Zhu, Kaikai Ge, Fuzhen Zhuang, Ruobing Xie, Dongbo Xi, Xu Zhang, Leyu Lin, and Qing He. 2021. Transfer-Meta Framework for Cross-domain Recommendation to Cold-Start Users. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR*

'21). Association for Computing Machinery, New York, NY, USA, 1813–1817. https://doi.org/10.1145/3404835.3463010

[71] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web* (Chiba, Japan) *(WWW '05)*. Association for Computing Machinery, New York, NY, USA, 22–32. https://doi.org/10.1145/1060745.1060754