

Input or Output: Effects of Explanation Focus on the Perception of Explainable Recommendation with Varying Level of Details

Mouadh Guesmi^a, Mohamed Amine Chatti^a, Laura Vorgerd^a, Shoeb Joarder^a, Qurat Ul Ain^a, Thao Ngo^a, Shadi Zumor^a, Yiqi Sun^a, Fangzheng Ji^a and Arham Muslim^b

^aUniversity of Duisburg-Essen, Germany

^bNational University of Sciences and Technology, Pakistan

Abstract

In this paper, we shed light on two important design choices in explainable recommender systems (RS) namely, explanation focus and explanation level of detail. We developed a transparent Recommendation and Interest Modeling Application (RIMA) that provides on-demand personalized explanations of the input (user model) and output (recommendations), with three levels of detail (basic, intermediate, advanced) to meet the demands of different types of end-users. We conducted a within-subject study to investigate the relationship between explanation focus and the explanation level of detail, and the effects of these two variables on the perception of the explainable RS with regard to different explanation aims. Our results show that the perception of explainable RS with different levels of detail is affected to different degrees by the explanation focus. Consequently, we provided some suggestions to support the effective design of explanations in RS.

Keywords

recommender system, explainable recommendation, personalized explanation, explanation design choices

1. Introduction

Explanations in recommender systems (RS) have gained an increasing importance in the last few years. An explanation can be considered as a piece of information presented to the user to expose the reason behind a recommendation [1]. Explanations can have a huge effect on how users respond to recommendations [2]. Recent research focused on different dimensions and design choices of the explanation provided by the RS. These include the explanation aim (e.g., transparency, trust, effectiveness), explanation type or style (e.g., content-based, collaborative filtering, hybrid, social), and explanation format or display (textual, visual, hybrid) [3, 4, 5, 6, 7]. Additionally, other essential design choices must be considered, such as the focus and the level of detail of the explanation [8].

InTRS'21: Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, September 25, 2021, Virtual Event

✉ mouadh.guesmi@stud.uni.de (M. Guesmi); mohamed.chatti@uni-due.de (M. A. Chatti); Laura.vorgerd@stud.uni-due.de (L. Vorgerd); shoeb.joarder@uni-due.de (S. Joarder); qurat.ain@stud.uni-due.de (Q. U. Ain); thao.ngo@uni-due.de (T. Ngo); shadi.zumor@stud.uni-due.de (S. Zumor); yiqi.sun@stud.uni-due.de (Y. Sun); fangzheng.Ji@stud.uni-due.de (F. Ji); arham.muslim@seecs.edu.pk (A. Muslim)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

The focus of an explanation refers to the part that a RS is trying to explain, i.e., the recommendation input (i.e., user model), process (i.e., algorithm), or output (i.e., recommended items). Explainable recommendation focusing on the recommendation process aims to understand how the algorithm works. The explainability of the recommendation output focuses on the recommended items. This approach treats the recommendation process as a black box and tries to justify why the recommendation was presented. The explainability of the recommendation input focuses on the user model. This approach provides a description that summarizes the system's understanding of the user's preferences and allows the user to scrutinize this summary and thereby directly modify his or her user model [2]. Compared to explainability of the recommendation output or the recommendation process, focusing on the recommendation input is under-explored in explainable recommendation research [2, 4].

Another crucial design choice in explainable recommendation relates to the level of explanation detail that should be provided to the end-user. Results of previous research on explainable AI (XAI) and explainable recommendation revealed that for specific users or user groups, the detailed explanation does not automatically result in higher trust and user satisfaction. The reason is that the provision of additional explanations increases cognitive effort, and different users have different needs for explanation [9, 10, 11, 12, 13].

In this paper, we aim at exploring the effects of the two design choices, namely explanation focus and explanation level of detail on the perception of explainable recommendations. To this end, we conducted a user study where we investigated the dependencies between these two factors and their effects on the user perception of seven different explanation aims, namely transparency, scrutability, trust, effectiveness, persuasiveness, efficiency, and satisfaction. As a result, we derived some suggestions to be considered when designing explanations with different levels of detail.

To conduct this study, we developed a transparent Recommendation and Interest Modeling Application (RIMA) that provides on-demand personalized explanations of the recommendations (output) as well as the underlying interest models (input), both with three different levels of detail (basic, intermediate, advanced), in order to meet the needs and preferences of different users. The objective of the study was to answer the following research question: *How do explanation focus and explanation level of detail influence the perception of explanations in terms of seven explanation aims?* The results of our study show that the effects of the explanation level of detail on the perception of explainable recommendation depend on the explanation focus, thus providing evidence for a dependency relationship between explanation aim, explanation focus, and explanation level of detail.

The remainder of this paper is organized as follows. We first outline the background for this research (Section 2). We then present the different explanations used in RIMA application (Section 3). An empirical study is presented in (Section 4), followed by a discussion of the main findings (Section 5). Finally, we summarize the work and outline future research plans (Section 6).

2. Related work

In the following, we discuss related work on explainable recommendation related to two important design choices, namely explanation focus and explanation level of detail.

2.1. Explanation Focus

Explanations in RS can be classified based on the part of the recommendation they try to explain, namely the recommendation *input*, recommendation *process*, and recommendation *output* [14].

Explaining the input: The explainability of the recommendation input focuses on the user model which represents the user's interests and preferences. The rise of distrust and skepticism related to the collection and use of personal data, and privacy concerns in general has led to an increased interest in transparency of black-box user models, used to provide recommendations [15]. Explanations focusing on the input aim to open the user model by revealing the system's assumptions about the user's interests, preferences, or needs [2]. Graus et al. [4] stress the importance of enabling transparency by opening and explaining the black box user profiles, that serve as input for the RS. This can help users become aware of their interests used for the recommendations [16], facilitate users' self-actualization (i.e., developing, exploring, and understanding their unique personal tastes) [17], build a more accurate mental model of the system [8], detect wrong assumptions made by the system [16], and contribute to scrutability, allowing users to provide explicit feedback on their generated user profiles. Only few works followed this approach and provided explanations of the user model [2, 18, 19].

Explaining the process: Explanations that focus on the recommendation process attempt to expose (parts of) the underlying logic (i.e., explanation of algorithmic working) [7]. For example, 'SmallWorlds' [20] visualizes a complex network based on five layers to explain the connection between the active user and the recommended friends. Zhao et al. [12] reveals the inner logic of the RS by showing the exact algorithms used to compute similarities between users and predictions for recommendations. However, keeping in mind the complexity of the underlying algorithm, explaining the recommendation process is not a straightforward task, as in many cases the underlying complex algorithms can not be described in a human-interpretable manner [21].

Explaining the output: Explanations that focus on the recommendation output aim to provide a justification for why a particular recommendation was provided without revealing the inner logic of the system [7]. One example is the classic explanation "customers who are similar to you also like...", which can already be found in many commercial online services [12], especially in collaborative filtering RS. Another example is the music RS 'Moodplay' [22] that explains recommended artists by referring to the mood of the songs (e.g., joyful or sad) the user has previously listened to.

While the task of opening the black box of RS by explaining the recommendation output (i.e., why an item was recommended) or the recommendation process (i.e., how a recommendation was generated) is well researched in the explainable recommendation community, researchers have only recently begun exploring methods that support the exploration and understanding of the recommendation input (i.e., the user model) to provide transparency in RS [2]. Moreover, investigating different explanation foci (e.g., input and output) in parallel is lacking in the

explainable RS literature. RS explaining both the input and output allow the users to understand the relationship between their user model and the recommendations received, thus allowing them to interact with the system predictably and efficiently [23]. To fill this gap, we aim in this work to explain both the input and the output of the RS with varying level of details to address different explanation aims such as transparency, scrutability, and satisfaction. Further, we investigate the effects of the explanation focus on the perception of the explainable recommendation.

2.2. Explanation with varying level of details

In this work, the level of detail refers to the amount of information exposed in an explanation. Generally, in the explainable AI (XAI) domain, different users will have different goals in mind while using such systems. For example, Mohseni et al. [8] point out that while machine learning experts might prefer highly-detailed visual explanations of deep models to help them optimize and diagnose algorithms, systems with lay-users as target groups aim instead to enhance the user experience with the system through improving their trust and understanding. In the same direction, Miller [24] argue that providing the exact algorithm which generated the specific recommendation is not necessarily the best explanation. People tend not to judge the quality of explanations based on how they were generated, but instead around their usefulness. Aside from the goals of the users, another crucial aspect that will influence their understanding of explanations are their cognitive capabilities [12].

Different levels of explanation detail would lead to different levels of RS transparency. Here, it is necessary to differentiate between objective transparency and user-perceived transparency. On the one hand, Objective transparency means that the RS reveals the underlying algorithm of the recommendations either by explaining it or justify it in case of high complexity of the algorithm. On the other hand, user-perceived transparency is thus based on the users' subjective opinion about how good the system is capable of explaining its recommendations [21]. In general, it can be assumed that a higher level of explanation detail increases the system's objective transparency but is also associated with a risk of reducing the user-perceived transparency, depending on the users' background knowledge.

Providing explanations with varying level of details remains rare in the literature on explainable recommendations. To the best of our knowledge, only Millecamp et al. [11] followed this approach while developing a music RS. The authors suggest that users should have the option to decide whether or not to see explanations, and explanation components should be able to present varying level of details to the users depending on their preferences. Consequently, their system allows users to choose whether or not to see the explanations by using a "Why?" button and also enables them to select the level of detail by clicking on a "More/Hide" button.

3. RIMA

We developed the transparent Recommendation and Interest Modeling Application (RIMA) with the goal of explaining the recommendations (output) as well as the underlying interest models (input). RIMA is a content-based RS that produces content-based explanations. It follows a user-driven personalized explanation approach by providing explanations with different levels

of detail and empowering users to steer the explanation process the way they see fit. The application provides on-demand explanations, that is, the users can decide whether or not to see the explanation and they can also choose which level of explanation detail they want to see [25]. In this work, we focus on recommending tweets and Twitter users and leveraging explanatory visualizations to provide insights into the recommendation process. The current design of the different levels of detail was mainly the result of brainstorming sessions involving the authors and was inspired by popular explanation visualizations used in the literature on explainable RS, such as word clouds and heatmaps.

3.1. Explaining the interest model

The aim of explaining the interest model in RIMA is to foster user’s *awareness* of the data that the RS uses as an input to generate recommendations, in order to increase *transparency* and improve user’s *trust* in the RS. Moreover, this may let users become aware of system errors and consequently help them give feedback and correction in order to improve future recommendations (*scrutability*). The application provides an on-demand explanation of the interest models (input) with three different levels of detail (basic, intermediate, and advanced). These interest models are generated from users’ publications and tweets [26, 27]. The inferred interest model is presented to the user in a tag cloud. The user can hover over an interest to see its source (i.e., publications or tweets) as a *basic explanation* (Figure 1a). When the user clicks on an interest in the tag cloud, s/he will get more information through a pop-up window highlighting the occurrence of the selected interest in the tweets or title/abstract of publications, which represents the *intermediate explanation* (Figure 1b). The next level of detail is provided in the *advanced explanation* which follows an explanation by example approach to show in detail the logic of the algorithm used to infer the interest model (Figure 1c).

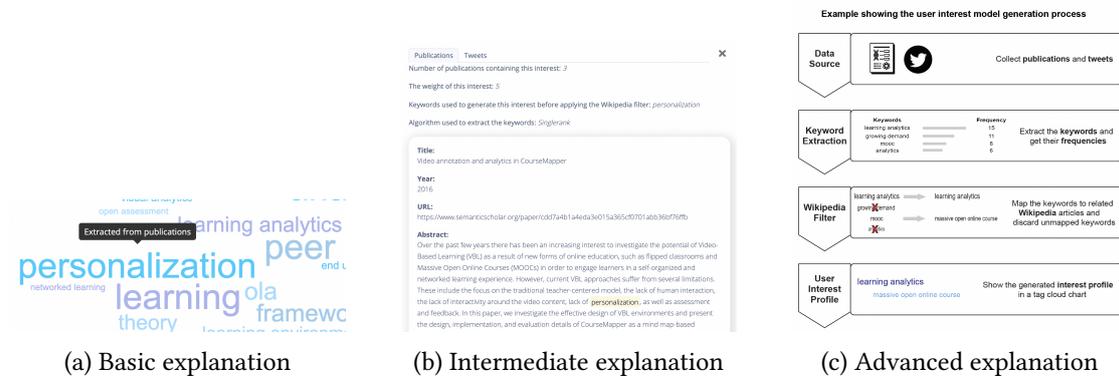


Figure 1: Explaining the interest model with three levels of details

3.2. Explaining the recommendation

The aim of explaining the recommendation in RIMA is to provide a *justification* on why a specific recommendation was presented and to help users’ *understanding* of how the recommendation

works. This can improve users' mental model of the underlying recommendation algorithm. Further, *transparency* of the RS can improve user experience through better understanding of the recommendation output, thus improving user interaction, *trust*, and *satisfaction* with the system.

The application provides an on-demand explanation of the recommendations (output) with three different levels of detail (basic, intermediate, and advanced). The *basic explanation* aims at explaining "why" a specific tweet was recommended in an abstract manner. The search box is initially populated with the user's top five interests, ordered by their weights as generated by the system. Users can also add new interests in the search box or remove existing ones. The system will use these interests as input for the recommendation process. The basic explanation is achieved using a color band to map the tweet to the related interest(s). Also, the interest will be highlighted in the text of the tweet to show that this tweet contains this specific word (interest). In addition to these two visual elements, we display the similarity score on the top right corner of the tweet to show the level of similarity between the user interests and the recommended tweet (Figure 2a).



Figure 2: Explaining the tweet recommendation with three levels of details

For more details, the user can choose the *intermediate explanation* level by clicking on "Why this tweet?" on the bottom right of the tweet. Similar to the basic level, the intermediate level also aims at answering the "why" question, but with more details. We used a Heatmap chart to show the semantic similarity between the user interest profile and the keywords extracted from the text of the tweet. The x-axis represents the keywords extracted from the tweet and the y-axis represents the user's interests used in the recommendation. The cells show the computed semantic similarity scores between each interest and keyword (Figure 2b).

To move to the *advanced explanation* level, the user has to click on the "more" button on the bottom right of the intermediate explanation window. The aim of the advanced explanation is to explain "how" the recommendation algorithm works. This is achieved by following an explanation by example approach to show in detail the logic of the algorithm used to semantically compare the keywords extracted from the recommended tweet and the user interests (Figure 2c).

4. Empirical Study

4.1. Participants

To obtain a diverse sample, the study included participants from different countries, educational levels, and study backgrounds. A total of 36 participants completed the study. We ensured the data quality through the examination of redundant answering patterns (e.g., consistent selection of only one answering option) and attention checks, accordingly, five participants were excluded. The final sample consisted of $N = 31$ participants (14 males, 17 females) with an average age of 32 years. Out of the 31 participants, 19 (61.3%) reported to live in Germany, where 12 (38.7%) were international users from eight different countries, and the highest reported education level by most participants was *master's degree* (61.3%).

4.2. Study Procedure

While the study was originally planned as a laboratory experiment, due to the COVID-19 pandemic and its restrictions, we decided to conduct an online study. Each session was accompanied by a research assistant for technical support. All participants gave informed consent to study participation. Participants were recruited via e-mail, word-of-mouth, and groups in social media networks and had to fulfill two participation requirements: they had to have at least one scientific publication and a Semantic Scholar ID.

Participants first answered a questionnaire in SosciSurvey¹ which asks for their Semantic Scholar ID and included general questions about their preferences and expertise. Next, participants were given a short demo video on how to use the RIMA application. Afterwards, participants were asked to (1) create an account using their Semantic Scholar ID, (2) explore the system and find matching recommendations to their interests, and (3) take a close look at each explanation provided by the system. After that, participants were asked to evaluate each of the six explanations in terms of seven explanation goals (transparency, scrutability, trust, effectiveness, persuasiveness, efficiency, and satisfaction [6]). All participants evaluated the

¹<https://www.socisurvey.de>

explanations in an iterative and randomized approach, by answering the same set of questions for each explanation. The order in which participants rated the explanations was randomized in order to avoid any order-related biases. They needed on average 48.09 minutes to complete the questionnaire ($SD = 9.40$, range = 24.08-65.23). At the end, they were debriefed and compensated with the possibility to win one of five Amazon vouchers.

4.3. Measurements

4.3.1. Explanation Aims

The measurements for the seven explanation aims were adopted from different previous works [28, 29, 16, 30, 31, 12]. The first six explanation aims were measured using a 5-point Likert-scale, while satisfaction was measured using a 7-point Likert-scale. An overview of used questionnaire items is shown in Table 1. Besides the quantitative measurement of the explanation aims, participants could also provide qualitative feedback on each explanation in open-ended questions.

Metric	Statement This explanation ...	Source
Transparency	helps me to understand what the recommendations are based on.	[28]
Scrutability	allows me to give feedback on how well my preferences have been understood.	[28]
Trust (Competence)	shows me that the system has the expertise to understand my needs and preferences.	[31]
Trust (Benevolence)	shows me that the system keeps my interests in mind.	[31]
Trust (Integrity)	shows me that the system is honest.	[31]
Effectiveness	helps me to determine how well the recommendations match my interests.	[16]
Persuasiveness	is convincing.	[13]
Efficiency	helps me to determine faster how well the recommendations match my interests.	[16]
Satisfaction	Question How good do you think this explanation is?	[29, 30]

Table 1

An overview of questionnaire items used for the evaluation of explanations.

4.3.2. Overall User Experience

In addition to the perception of the explanations, we included additional measurements in our study to capture the participants' perceptions towards the recommended tweets and the RIMA application as a whole. We adopted a number of questionnaire items from the "ResQue" evaluation framework by Pu et al. [32] and from the framework by Knijnenburg et al. [33]. In addition, we designed two questionnaire items to measure the participants' satisfaction with their interest model and the extent to which they had to adjust their interest model. In total, 14 additional questionnaire items were included in our study, which are shown in Table 2. Answers

were given on a 5-point Likert scale, ranging from 1 ("strongly disagree") to 5 ("strongly agree"). Finally, three open-ended questions were included to capture additional feedback on the most and least useful parts of the application and suggestions for improvements [34, 11].

Metric	Statement	Source
Ease of initial learning	I became familiar with the recommender system very quickly.	[32]
Ease of preference elicitation	I found it easy to tell the recommender system about my preferences.	[32]
Ease of preference revision	I found it easy to alter the outcome of the recommended Tweets due to my preference changes.	[32]
Ease of decision making	Finding interesting Tweets with the help of the recommender system is easy.	[32]
Control	I feel in control of telling the recommender system what I want.	[32]
Usefulness	The recommendations effectively helped me find interesting Tweets	[32]
	I feel supported to find what I'm interested in with the help of the recommender system.	[32]
Interface adequacy	The layout of the recommender system is attractive and adequate	[32]
Overall satisfaction	Overall, I am satisfied with the recommender system.	[32]
Choice satisfaction	I like the Tweets I have chosen.	[33]
Recommendation quality	The provided recommended Tweets were interesting.	[33]
Recommendation variety	The list of recommended Tweets had a high variety.	[33]
Interest model accuracy	The recommender system knows my interests very well.	new item
Adjustment of interests	I had to adjust my interests to get suitable recommendations.	new item

Table 2

Overview of questionnaire items used for the evaluation of the overall user experience.

4.4. Results

4.4.1. Descriptive Data

As described earlier, the RIMA application explains the interest model (input) and recommendations (output), both with three different levels of detail (basic, intermediate, advanced). All participants rated the six explanations in terms of seven explanation aims (transparency, scrutability, trust, effectiveness, efficiency, persuasiveness, and satisfaction). We calculated the evaluation score for trust as the average of the individual values reported for the three trusting beliefs (i.e., competence, benevolence, and integrity).

4.4.2. Interaction Effects

To address our research question: *How do explanation focus and explanation level of detail influence the perception of explanations in terms of the seven explanation aims?*, we performed a set of seven repeated-measures ANOVA analyses to evaluate the simultaneous effects of the explanation focus and the explanation level of detail on the perception of explanations in terms of the seven explanation aims. Here, the evaluation scores of the explanation aims were included as measures, and *explanation focus* (input, output) and *explanation level of detail* (basic, intermediate, advanced) as factors. The results are summarized below.

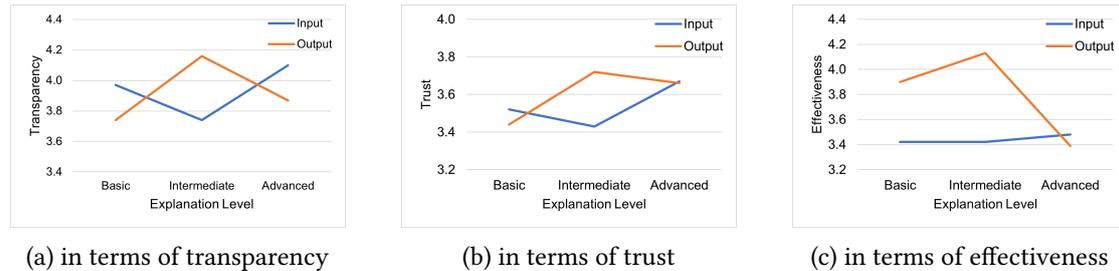


Figure 3: The interaction effects between explanation focus and explanation level of detail

Transparency: There were no main effects of explanation focus ($F(1,30) = 0.007, p = .934$) or explanation level of detail ($F(2,60) = 0.507, p = .605$) in terms of transparency. However, we found a significant interaction between *explanation focus* and *explanation level of detail* ($F(2,60) = 4.028, p = .023, f = .37$). The effect size corresponds to a moderate effect [35]. The interaction effect is depicted in Figure 3a. The simple slopes show that, for the input, the average rating of transparency was lower for the intermediate explanation and higher for the advanced explanation, while it was the other way around for the output.

Scrutability: No main effects of explanation focus ($F(1,30) = 1.752, p = .196$) or explanation level of detail ($F(2,60) = 1.348, p = .267$) in terms of scrutability were found, neither a significant interaction between explanation focus and explanation level of detail ($F(2,60) = 0.731, p = .485$).

Trust: There were no main effects of explanation focus ($F(1,30) = 0.362, p = .552$) or explanation level of detail ($F(2,60) = 1.680, p = .195$) in terms of trust. However, there was a significant interaction between *explanation focus* and *explanation level of detail* ($F(2,60) = 3.540, p = .035, f = .34$). The effect size corresponds to a moderate effect [35]. Figure 3b shows that the simple slopes look similar to the interaction effect in terms of transparency: for the input, the average rating of trust was lower for the intermediate explanation and higher for the advanced explanation, while it is the other way around for the output.

Effectiveness: We found a significant main effect of *explanation focus* in terms of effectiveness ($F(1,30) = 4.978, p = .033, f = .41$). The average rating of effectiveness was significantly higher for the output ($M = 3.81, SD = 0.13$) than for the input ($M = 3.44, SD = 0.14$). The effect size corresponds to a strong effect [35]. There was no main effect of explanation level of detail ($F(2,60) = 1.845, p = .167$). The interaction between *explanation focus* and *explanation level of detail* was significant ($F(2,60) = 3.929, p = .025, f = .38$). The effect size corresponds to a moderate effect [35]. Figure 3c shows that the basic and intermediate explanations of the input had higher

average ratings of effectiveness than the input. Further, the advanced explanations of both the input and output had equally lower ratings of effectiveness.

Efficiency: We found a significant main effect of *explanation level of detail* in terms of efficiency ($F(2,60) = 7.299, p = .002, f = .49$). Bonferroni-corrected pairwise comparisons revealed significant differences between the basic and advanced ($p = .013$) and between the intermediate and advanced explanations ($p = .023$), such that the average rating of efficiency was significantly higher for the basic explanations ($M = 3.73, SD = 0.15$) and the intermediate explanations ($M = 3.58, SD = 0.12$) than for the advanced explanations ($M = 3.11, SD = 0.17$). The effect size corresponds to a strong effect [35]. No main effect of explanation focus was found ($F(1,30) = 3.707, p = .064$), neither a significant interaction between explanation level of detail and explanation focus ($F(2,60) = 1.000, p = .374$).

Persuasiveness: No main effects of explanation focus ($F(1,30) = 3.306, p = .079$) and explanation level of detail ($F(2,60) = 0.355, p = .702$) in terms of persuasiveness were found, neither a significant interaction between explanation focus and explanation level of detail ($F(2,60) = 0.643, p = .529$).

Satisfaction: No main effects of explanation focus ($F(1,30) = 0.490, p = .489$) or explanation level of detail ($F(2,60) = 0.475, p = .624$) in terms of satisfaction were found, neither a significant interaction between explanation focus and explanation level of detail ($F(2,60) = 2.583, p = .084$).

4.5. Overall User Experience

In addition to the evaluation of the explanations, we included questionnaire items to evaluate the overall user experience of the RIMA application. Figure 4 shows the mean ratings of the different variables that were measured for this purpose, reported on a 5-point Likert-scale.



Figure 4: Overall user experience with the RIMA application

The average *overall satisfaction* with the RIMA application was near the mid-point ($M = 3.13, SD = 0.96$). We observed that the average rating of *ease of initial learning* was relatively high ($M = 3.77, SD = 1.12$), which indicates that participants became familiar with the RIMA

application quickly. The average rating of the *interface adequacy* was high ($M = 4.00$, $SD = 0.97$). This indicates that participants were satisfied with the general user interface design of the RIMA application. In addition, the average rating of *control* ($M = 3.65$, $SD = 1.05$) indicates that participants felt in control over their recommendations. The average rating of *ease of preference elicitation* ($M = 3.94$, $SD = 1.06$) also indicates that participants found it easy to tell the system about their preferences. The average rating of *ease of preference revision* was lower ($M = 3.48$, $SD = 1.18$), which indicates that participants found it more difficult to alter the outcome of their recommendations due to preference changes.

The average rating of *recommendation quality* was near the mid-point ($M = 3.06$, $SD = 1.06$). This result reflects the answers of participants to the open-ended questions, where almost half of the participants (14 out of 31) reported being dissatisfied with the quality of the recommendations. Out of these participants, the majority reported that the tweets were not related to their scientific interests. In addition, the reported issues with the interest extraction algorithm also reflect in the rating of *interest model accuracy*. Here, the average rating was below the mid-point ($M = 2.81$, $SD = 1.01$), which indicates that participants felt that the system did not know their interests very well. The average rating of *adjustment of interests* ($M = 4.19$, $SD = 1.01$) also indicates that participants had to adjust their interest model to get suitable recommendations. The average rating of *ease of decision making* was below the mid-point ($M = 2.77$, $SD = 1.12$), which indicates that participants found it difficult to find interesting tweets. The average rating of *recommendation variety* was higher ($M = 3.42$, $SD = 1.15$) than the rating of the recommendation quality.

The perceived *usefulness* of the RIMA application was near the mid-point ($M = 3.03$, $SD = 1.02$). This indicates that the ability of the RIMA application to help users find interesting tweets was perceived as relatively neutral. The average rating of *choice satisfaction* ($M = 3.35$, $SD = 1.08$) indicates that participants were on average neither satisfied nor dissatisfied with the tweets they selected as part of the task.

5. Discussion

In this section, we discuss the main findings of our study in relation to our research question: *How do explanation focus and explanation level of detail influence the perception of explanations in terms of the seven explanation aims?* and provide some suggestions for the effective design of explanations in RS.

Efficiency. Our analysis showed that the explanation level of detail influenced the perceived efficiency of explanations. In particular, the basic explanations were rated as most efficient, followed by the intermediate and advanced explanations, which indicates that increasing the explanation level of detail resulted in lowered perceptions of efficiency. This result is in line with the work of Lage et al. [36] who found that greater complexity of machine learning explanations resulted in longer user response times. This suggests that simple explanations are more suitable to increase the efficiency of an explanation facility. In contrast, explanations with a high level of detail reduce efficiency as users need more time and cognitive effort to interpret the provided information, which limits the ability of explanations to help users make decisions faster [16]. Overall, our result is in line with previous findings that some explanations help users determine

the quality of a recommendation more quickly than others [21]. Our finding also confirms the warnings of researchers that highly detailed information about the system's inner logic reduces efficiency [5, 37, 12] and that simple explanations are often better [24]. Therefore, we propose the following design suggestion for explainable RS:

Suggestion 1: *If an explanation facility should be optimized for efficiency, use explanations with a low level of detail*

At this point, we further note that, even if efficiency is an important aspect for users' decision-making, there may be other explanation aims that are more important. Gedikli et al. [21] found that efficiency is no important influencing factor for the overall user satisfaction. They argue that users are willing to invest time to interpret an explanation in order to make good decisions, especially if the recommended item is expensive or comes with risk. However, if the goal is to help users determine the quality of a recommendation faster and with less cognitive effort, we recommend using simple explanations.

Effectiveness. Similar to efficiency, we found that the explanation focus influenced the perceived effectiveness of explanations. In particular, explanations that focus on the recommendation output (i.e., recommended items) were perceived as more effective than explanations that focus on the recommendation input (i.e., interest model). This indicates that the explainability of the output is more effective in helping users make good decisions, as these explanations directly focus on the recommendations by specifying how well a specific item matches their interests. In contrast, the explainability of the input aims to open the underlying user model, thus it may be less helpful for determining the quality of a specific recommendation. Our second design suggestion is therefore:

Suggestion 2: *To achieve higher effectiveness of explanations, focus directly on explaining the recommended items.*

However, we want to note that, in the RIMA application, the explanations of the interest model were shown on a different page than the recommended items. This visual separation may lowered the ratings of effectiveness for the explanations of the interest model. Nevertheless, we believe that explanations that focus on the output are more suitable to increase effectiveness, as users need to accurately estimate the quality of a recommendation in order to make good decisions [38].

In addition, we found an interaction effect between the explanation level of detail and the explanation focus in terms of effectiveness. As depicted in Figure 3c, the effect of the explanation level of detail on the perceived effectiveness depends on whether the explanations focus on the input or the output. The interaction plot shows that, for the input, the explanation level of detail had no great impact as all three explanations had equally low ratings of effectiveness. However, for the output, users perceived the intermediate explanation to be most effective, whereas the advanced explanation was perceived as least effective. As the intermediate explanation consisted of a heatmap that shows the computed similarities between the user's interest keywords and the keywords extracted from a tweet, it seems that users could leverage this information to determine how well a recommended tweet matches their interests. The basic explanation of the output was also perceived as relatively effective, however, the similarity score alone may not be enough to make an informed decision, and users need more information about the relevance of a specific recommendation. On the other hand, the advanced explanations of both the input and the output had lower ratings of effectiveness. The answers to the open-ended questions suggest

one main reason for this finding: as the advanced explanations revealed the system's inner logic via example and were not directly linked to the users' data, they were less effective in showing users how well a recommendation matches their actual interests. This is in line with researchers who suggest that a good explanation should reflect the users' actual preferences to support them in correctly determining the quality of a recommendation [38, 30]. Explanations with poor effectiveness could negatively impact user satisfaction to the extent that the user ceases to use the system [30]. Gedikli et al. [21] also argue that effective explanations are important for the success and user satisfaction with the RS in the long run. Therefore, we derive a further design suggestion for effective explanations:

Suggestion 3: *Boost effectiveness through highlighting the match between a recommended item and the user's actual interests.*

Transparency. Our analysis revealed an interaction effect between the explanation focus and the explanation level of detail in terms of perceived transparency. The interaction plot in Figure 3a shows that the explanation level of detail had different effects on transparency, depending on whether the explanations focus on the input or the output. In particular, for the input, the intermediate explanation led to lower and the advanced explanation led to higher perceptions of transparency. A possible explanation for this could be that the information in the basic explanation (i.e., source of interest keyword) was sufficient for users to understand that the system extracts their interests from their publications in order to generate recommendations, whereas the additional information about their publication in the intermediate explanation could not further increase transparency. However, as the advanced explanation differs in that it provided detailed information about the keyword extraction algorithm, it might improved users' understanding of the system.

For the output, the effect of the explanation level of detail on transparency looks exactly the opposite: the intermediate explanation led to higher and the advanced explanation led to lower perceptions of transparency. We believe that the basic explanation only created a general understanding of what the recommendations are based on (i.e., similarity score), whereas the heatmap in the intermediate explanation helped users further understand that the recommendations are based on a matching-process between their interest keywords and the keywords extracted from a tweet. Contradictory to the input, the advanced explanation of the output could not further improve users' understanding of the system, and it also had a lower rating of transparency than the advanced explanation of the input.

The answers to the open-ended questions indicated that participants could not fully read the advanced explanation of the output as the example values in the flowchart were too small, while the other flowchart of the input was more compact and fully readable. We also observed that participants found the advanced explanation of the input less overwhelming and easier to process. Thus, we believe that design issues limited the ability of the advanced explanation of the output in helping users understand what their recommendations are based on. Therefore, we suggest:

Suggestion 4: *When providing visual explanations with a high level of detail to increase transparency, ensure that they are fully readable and not overwhelming.*

Trust. The third interaction effect we found between the explanation level of detail and the explanation focus was in terms of trust. When looking at the interaction plot in Figure 3b, we observe that the simple slopes appear similar to the interaction effect for transparency: for the

input, the intermediate explanation led to lower and the advanced explanation led to higher trust, while it is the other way around for the output. Moreover, the intermediate explanation of the output had the highest ratings of trust. As the heatmap in the intermediate explanation showed users how their actual interest keywords relate to a specific tweet recommendation, we believe that the explanation created beliefs that the system keeps the users' interests in mind when generating recommendations, thus it increased users' perceived trustworthiness. In contrast, the advanced explanation of the output was designed as an explanation via example and did not differ from tweet to tweet, which might negatively influenced the perceived trustworthiness of the system. Therefore, we suggest:

Suggestion 5: *To increase trust in the system, provide explanations that address the users' actual interests.*

Scrutability. We could not find any main or interaction effects in terms of scrutability. This contradicts other researchers' assumptions that the explainability of the input enhances scrutability as opening the user model should allow users to give feedback on how well their interests have been understood [16]. The answers to the open-ended questions also contradict this result as a number of participants reported that the heatmap helped them identify errors which led to unwanted tweet recommendations. Therefore, further investigations are needed to find out which types of explanations are more suitable to enhance scrutability.

Persuasiveness. Further, there were no main or interaction effects in terms of persuasiveness. This might indicate that the level of detail does not impact the persuasiveness of explanations, and other design choices such as the explanation style or format may be more important. For instance, Kouki et al. [13] found that the persuasiveness of explanations depends to a large extent on the explanation style, for example, item-based explanations were more convincing than user-based explanations. In addition, they found that textual explanations were more convincing than visual explanations. Thus, when it comes to persuasion, we believe that the content or quality of an explanation play a more important role than the quantity of provided information. Miller [24] also argues that, if the goal of an explanation is persuasion, then it is more important that the content of the explanation convinces the explainee that the decision of the explainer is correct, instead of actually providing the most likely cause of an event (i.e., the true reason behind a recommendation).

Satisfaction. Finally, we found no main or interaction effects in terms of satisfaction. The answers to the open-ended questions also indicated that there were no "best" or "worst" explanations in the RIMA application, but satisfaction with the different explanations rather seemed to be equally divided across participants. To put in other words, the explanations did not differ significantly in their ratings of satisfaction, and the effect of explanation level of detail on user satisfaction can only be observed when taking individual differences into account.

6. Conclusion and Future Work

In this paper, we investigated the effects of explanation focus and explanation level of detail on the perception of explanations in a recommender system (RS) in terms of seven explanation aims. To this end, we developed and evaluated a transparent Recommendation and Interest Modeling Application (RIMA) that explains both the user interest model (input) and recommendations

(output) with three different levels of detail (basic, intermediate, advanced). The results of our study demonstrated that the explanation focus affects to different degrees the perception of explainable recommendation with varying level of details. From our findings, we provided some suggestions to be considered when designing explanation interfaces in RS. In future work we will explore other possible visualizations to provide explanations at the three levels of detail. The current design of the different levels of detail was mainly the result of brainstorming sessions involving the authors. In the future, we are planning to follow a human-centered design approach to come up with more systematic designs. Moreover, we plan to enlarge the sample size and improve our analysis. Further, we will investigate the interaction effects of personal characteristics and explanation level of detail on the perception of explainable RS.

References

- [1] J. L. Herlocker, J. A. Konstan, J. Riedl, Explaining collaborative filtering recommendations, in: Proceedings of the 2000 ACM conference on Computer supported cooperative work, 2000, pp. 241–250.
- [2] K. Balog, F. Radlinski, S. Arakelyan, 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, 2019, pp. 265–274.
- [3] M. Guesmi, M. A. Chatti, A. Muslim, A review of explanatory visualizations in recommender systems, in: Companion Proceedings 10th International Conference on Learning Analytics and Knowledge (LAK20), 2020, pp. 480–491.
- [4] D. Graus, M. Sappelli, D. Manh Chu, "let me tell you who you are" - explaining recommender systems by opening black box user profiles, in: Proceedings of the FATREC Workshop on Responsible Recommendation, 2018.
- [5] I. Nunes, D. Jannach, A systematic review and taxonomy of explanations in decision support and recommender systems, *User Modeling and User-Adapted Interaction* 27 (2017) 393–444.
- [6] N. Tintarev, J. Masthoff, Explaining recommendations: Design and evaluation, in: *Recommender systems handbook*, Springer, 2015, pp. 353–382.
- [7] Y. Zhang, X. Chen, Explainable recommendation: A survey and new perspectives, arXiv preprint arXiv:1804.11192 (2018).
- [8] S. Mohseni, N. Zarei, E. D. Ragan, A multidisciplinary survey and framework for design and evaluation of explainable ai systems, arXiv (2018) arXiv–1811.
- [9] R. F. Kizilcec, How much information? effects of transparency on trust in an algorithmic interface, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2016, pp. 2390–2395.
- [10] T. Kulesza, S. Stumpf, M. Burnett, S. Yang, I. Kwan, W.-K. Wong, Too much, too little, or just right? ways explanations impact end users' mental models, in: 2013 IEEE Symposium on visual languages and human centric computing, IEEE, 2013, pp. 3–10.
- [11] M. Millecamp, N. N. Htun, C. Conati, K. Verbert, 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, 2019, pp. 397–407.

- [12] R. Zhao, I. Benbasat, H. Cavusoglu, Do users always want to know more? investigating the relationship between system transparency and users' trust in advice-giving systems (2019).
- [13] P. Kouki, J. Schaffer, J. Pujara, J. O'Donovan, L. Getoor, Personalized explanations for hybrid recommender systems, in: Proceedings of the 24th International Conference on Intelligent User Interfaces, 2019, pp. 379–390.
- [14] R. Zhao, I. Benbasat, H. Cavusoglu, Transparency in advice-giving systems: A framework and a research model for transparency provision., in: IUI Workshops, 2019.
- [15] E. Sullivan, D. Bountouridis, J. Harambam, S. Najafian, F. Loecherbach, M. Makhortykh, D. Kelen, D. Wilkinson, D. Graus, N. Tintarev, Reading news with a purpose: Explaining user profiles for self-actualization, in: Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, 2019, pp. 241–245.
- [16] N. Tintarev, J. Masthoff, Designing and evaluating explanations for recommender systems, in: Recommender systems handbook, Springer, 2011, pp. 479–510.
- [17] B. P. Knijnenburg, S. Sivakumar, D. Wilkinson, Recommender systems for self-actualization, in: Proceedings of the 10th acm conference on recommender systems, 2016, pp. 11–14.
- [18] H. Badenes, M. N. Bengualid, J. Chen, L. Gou, E. Haber, J. Mahmud, J. W. Nichols, A. Pal, J. Schoudt, B. A. Smith, et al., System u: automatically deriving personality traits from social media for people recommendation, in: Proceedings of the 8th ACM Conference on Recommender Systems, 2014, pp. 373–374.
- [19] J. Barria Pineda, P. Brusilovsky, Making educational recommendations transparent through a fine-grained open learner model, in: Proceedings of Workshop on Intelligent User Interfaces for Algorithmic Transparency in Emerging Technologies at the 24th ACM Conference on Intelligent User Interfaces, IUI 2019, Los Angeles, USA, March 20, 2019, volume 2327, 2019.
- [20] B. Gretarsson, J. O'Donovan, S. Bostandjiev, C. Hall, T. Höllerer, Smallworlds: visualizing social recommendations, in: Computer graphics forum, volume 29, Wiley Online Library, 2010, pp. 833–842.
- [21] F. Gedikli, D. Jannach, M. Ge, How should i explain? a comparison of different explanation types for recommender systems, International Journal of Human-Computer Studies 72 (2014) 367–382.
- [22] I. Andjelkovic, D. Parra, J. O'Donovan, Moodplay: Interactive mood-based music discovery and recommendation, in: Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization, 2016, pp. 275–279.
- [23] P. Pu, L. Chen, R. Hu, Evaluating recommender systems from the user's perspective: survey of the state of the art, User Modeling and User-Adapted Interaction 22 (2012) 317–355.
- [24] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, Artificial Intelligence 267 (2019) 1–38.
- [25] M. Guesmi, M. A. Chatti, L. Vorgerd, S. Joarder, S. Zumor, Y. Sun, F. Ji, A. Muslim, On-Demand Personalized Explanation for Transparent Recommendation, Association for Computing Machinery, New York, NY, USA, 2021, p. 246–252. URL: <https://doi.org/10.1145/3450614.3464479>.
- [26] M. Guesmi, M. A. Chatti, Y. Sun, S. Zumor, F. Ji, A. Muslim, L. Vorgerd, S. A. Joarder, Open, scrutable and explainable interest models for transparent recommendation, in: IUI

Workshops, 2021.

- [27] M. A. Chatti, F. Ji, M. Guesmi, A. Muslim, R. K. Singh, S. A. Joarder, SIMT: A Semantic Interest Modeling Toolkit, Association for Computing Machinery, New York, NY, USA, 2021, p. 75–78. URL: <https://doi.org/10.1145/3450614.3461676>.
- [28] K. Balog, F. Radlinski, 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, 2020, pp. 329–338.
- [29] N. Tintarev, J. Masthoff, The effectiveness of personalized movie explanations: An experiment using commercial meta-data, in: International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Springer, 2008, pp. 204–213.
- [30] N. Tintarev, J. Masthoff, Evaluating the effectiveness of explanations for recommender systems, *User Modeling and User-Adapted Interaction* 22 (2012) 399–439.
- [31] W. Wang, I. Benbasat, Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs, *Journal of Management Information Systems* 23 (2007) 217–246.
- [32] P. Pu, L. Chen, R. Hu, A user-centric evaluation framework for recommender systems, in: Proceedings of the fifth ACM conference on Recommender systems, 2011, pp. 157–164.
- [33] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, C. Newell, Explaining the user experience of recommender systems, *User Modeling and User-Adapted Interaction* 22 (2012) 441–504.
- [34] M. Millecamp, N. N. Htun, Y. Jin, K. Verbert, Controlling spotify recommendations: effects of personal characteristics on music recommender user interfaces, in: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, 2018, pp. 101–109.
- [35] J. Cohen, *Statistical power analysis for the behavioral sciences*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1988.
- [36] I. Lage, E. Chen, J. He, M. Narayanan, B. Kim, S. Gershman, F. Doshi-Velez, An evaluation of the human-interpretability of explanation, *arXiv preprint arXiv:1902.00006* (2019).
- [37] N. Tintarev, J. Masthoff, A survey of explanations in recommender systems, in: 2007 IEEE 23rd international conference on data engineering workshop, IEEE, 2007, pp. 801–810.
- [38] M. Bilgic, R. J. Mooney, Explaining recommendations: Satisfaction vs. promotion, in: *Beyond Personalization Workshop, IUI*, volume 5, 2005, p. 153.