



Delft University of Technology

Generating Consensus Explanations for Group Recommendations

An exploratory study

Najafian, Shabnam; Tintarev, Nava

DOI

[10.1145/3213586.3225231](https://doi.org/10.1145/3213586.3225231)

Publication date

2018

Document Version

Accepted author manuscript

Published in

UMAP '18 Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization

Citation (APA)

Najafian, S., & Tintarev, N. (2018). Generating Consensus Explanations for Group Recommendations: An exploratory study. In *UMAP '18 Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization* (pp. 245-250). Association for Computer Machinery.
<https://doi.org/10.1145/3213586.3225231>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

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

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Generating Consensus Explanations for Group Recommendations

an exploratory study

Shabnam Najafian
Delft University of Technology
Delft, the Netherlands
s.najafian@tudelft.nl

Nava Tintarev
Delft University of Technology
Delft, the Netherlands
n.tintarev@tudelft.nl

ABSTRACT

In some scenarios, like music, people often consume items in groups. However, reaching a consensus is difficult, and often compromises need to be made. Such compromises can potentially help users expand their tastes. They can also lead to outright rejection of the recommended items. One way to avoid this is to explain recommendations that are surprising, or even expected to be disliked, by an individual user. This paper presents an approach for generating explanations for groups. We propose algorithms for selecting a *sequence of songs* for a group to consume. These algorithms consider *consensus* but have different trade-offs. Next, using these algorithms we generated explanations in a layered evaluation using synthetic data. We studied the influence of these explanations in structured interviews with users (n=16) on user satisfaction.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Human-centered computing** → **User studies**; **Natural language interfaces**; **Empirical studies in HCI**; *Laboratory experiments*;

KEYWORDS

Explanations; Group recommendation; Sequences

ACM Reference Format:

Shabnam Najafian and Nava Tintarev. 2019. Generating Consensus Explanations for Group Recommendations: an exploratory study. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Recommender systems suggest items for users to consume, try, or buy, by learning from their past interactions, inferring their interests, and making predictions. Recommendations are often presented as lists of top-N recommendations where the user selects one or only a handful of items from a list of N highest ranked items.

Making recommendations for *groups* is a particularly difficult and interesting challenge. Here, different members of the group may have highly diverging tastes. However, group recommendation scenarios also create possibilities for novel discovery for individual group members: Mary's favorite song may become John's new ear-worm!

This paper raises the question of what happens in domains, like music, where commonly several of the items are consumed

in sequence. A sequence affords the recommender system more chances to make accurate recommendations, as well as to mix safe prediction of preference, and more risky or unfamiliar items.

We propose that developing algorithms for aggregating the user preferences within a group can help improve user satisfaction. However to do this effectively, this may need the use of validated explanations. To address these issues, this paper presents the following contributions:

- *Two improved algorithms* for constructing sequences of recommendations to groups. These strategies build on existing work on generating sequences, and address scenarios where users have different preferences from each other.
- *Evaluating the satisfaction* of users receiving explanations resulting from two different algorithms. In structured interviews we evaluated how four (4) different explanations influenced user satisfaction for five (5) different scenarios.

2 RELATED WORK

A key requirement for the success and adoption of advice-giving systems, such as recommender systems, is that users must trust system choices or even fully automated decisions [6]. Good explanations could help inspire user trust and loyalty, increase satisfaction, make it quicker and easier for users to find what they want, and persuade them to try or purchase a recommended item [9]. In this study we focus on satisfaction oriented explanations.

In the next sections we describe related literature in constructing sequences of recommendations, and explanations for groups.

2.1 Sequences of Recommendations

Masthoff et al. [3] suggest several algorithms for generating sequences of recommendations. These have as input a set of predictions for all users in a group for a set of items. Following are some of the algorithms that are likely influence trade-off based scenarios proposed in Masthoff work:

Least Misery Strategy: Make a new list of ratings with the minimum of the individual user ratings per item. Items get selected based on their rating on that list, the higher the rating the earlier the item appears in the sequence.

Most Pleasure Strategy: Make a new list of ratings with the maximum of the individual user ratings per item. Items get selected based on their rating on that list, the higher the rating the earlier the item appears in the sequence.

Average Without Misery Strategy: Make a new list of ratings with the average of the individual ratings, but without items that score below a certain threshold (say 4) for individuals. Items get selected

based on their rating on that list, the higher the rating the earlier the item appears in the sequence.

Fairness Strategy: Individuals take turns to receive their preferred items. The idea behind this strategy is that it is not so bad to watch something you hate, as long as you get to watch the things you really love as well.

As can be observed, these algorithms represent different strategies. For example, Average Without Misery will make sure that no-one is unhappy, but it may miss (or place at a very low rank) items that are only loved by some users. Fairness on the other hand makes sure everyone gets the items they love, but only when it is their turn. In this paper, we expand on these strategies, by combining several of them. In the next sections we will combine these strategies to maximize the advantages of the strategies and mitigate their possible disadvantages.

2.2 Explaining to Groups

Although previous studies have already analyzed explanations in the context of single-user recommendations, designing and generating explanations for groups contains a couple of open research issues. We summarize the limited work to date, and then outline the gaps.

Ardissono et al. developed a handheld recommender system for sightseeing destinations and itineraries by taking into account the preferences of heterogeneous tourist groups [1]. This system developed user models for different user groups (e.g., children, impaired people etc), and suggested recommendations based on the weighted average preferences for all item characteristic (e.g., eye-catching, historical value etc). This system supplied explanations based on the properties of items.

Felfernig et al.[2] reviewed previous works on explanations in single-user recommendations and how they should be altered to be useful for group recommendations. Similar to single-user recommendations, explanations for groups are designed based on the underlying recommendation algorithm. For example the explanation "*users who purchased item x also purchased item y*" can be defined in a group context such as "*groups that like item x also like item y*" [2]. The authors presented some existing examples of current explanations in group recommender systems. The examples showed the chosen aggregation strategy approach has an impact on the explanation style.

Quijano-Sanchez et al.[8] took into account group’s social factors, for example, users’ personal relationships within a group, i.e., avoid explanations that might damage friendships. "*Although we have detected that your preference for this item is not very high, your close friend X (who you highly trust) thinks it is a very good choice.*" [8]. Explanations for groups can have further goals rather than for individuals as it should consider certain aspects of group decision making i.e. taking into account, as far as possible, the preferences of all group members [2].

In addition to that, they concluded although initial approaches have already been proposed, different ways to explain group recommendations depending on the applied aggregation algorithms(s) are an issue for future research [2].

Nguyen & Ricci combined user preferences generated by the interactions between group members and the users’ long-term

preferences to optimally adapt the recommendations. They showed the advantage of their proposed model in capturing correctly the changes of the users’ needs. Although they studied group decision making and consensus, they have not studied explanations [4].

In this work, we design explanation for groups of people that are directly based on the used aggregation strategy to construct the sequences. The novel contribution is that these strategies and explanations both aim to help users with different preferences reach an acceptable consensus.

3 PROPOSED ALGORITHMS

In the previous section, we saw that different algorithms reflect different strategies. Here we define new algorithms for sequence construction, and explain the motivation for them. Examples illustrate the ordered group list of sequences resulting from the proposed algorithms. These two algorithms represent two ways of resolving consensus when preferences are inconsistent, that may affect user satisfaction in terms of e.g., avoiding misery, maximizing pleasure, fairness; and system performance, e.g., completeness, a clear ordering.

3.1 Explaining Sequences

In this paper we propose two novel algorithms; in Table 1 we compare them in terms of their properties. We see that they have complementary strengths and weaknesses.

Table 1: Properties for the proposed algorithms

Property	A 1	A 2
For users		
No Misery	+	-
Least Misery	+	-
Most Pleasure	+	-
Fairness	-	+
For the algorithm		
Completeness	-	+
Clear Order	-	+

More specifically, the two above mentioned algorithms are:

- A 1: **Least Misery + Most Pleasure + Without Misery.** The plus signs imply chaining three strategies, applying one after the other.
- A 2: **Fairness -> Average.** The arrow implies applying a tie-breaking strategy, i.e., when several items receive an equal score using only Fairness.

3.1.1 Algorithm 1: Least Misery + Most Pleasure + Without Misery. The strategy of this algorithm is to mitigate the original strategies’ disadvantages as much as possible. When using original Least Misery and Without Misery strategies on their own, items may be selected that nobody hates but also nobody really likes.

By applying Least Misery, Most Pleasure, and Without Misery strategies at the same time we try to ensure we avoid extreme low ratings, but support extreme high ratings. Sequences are then

ordering based on summing the predicted ratings of the lowest rating person (saddest) in the group and highest rating person (happiest) in the group regarding the recommended item. As a result we prioritize, and present first, items that maximize the rating of happiest person and at the same time minimize the unhappiness of the saddest person within the group.

Table 2: Least Misery (LM) + Most Pleasure (MP) + Without Misery

	A	B	C	D	E	F	G	H	I	J
John	10	4	3	6	10	9	6	8	10	8
Adam	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6
LM	-	4	-	6	7	8	5	6	-	6
MP	-	9	-	9	10	9	6	9	-	8
Sum	-	13	-	15	17	17	11	15	-	14

Group List: (E, F), (H, D), J, B, G (threshold 4)

The algorithm tries to keep the average satisfaction within the group high, by excluding low preference items by one or more people in the group. As it is shown in table 2, by applying Without Misery the songs that do not meet the certain threshold for individuals could be removed from the group list. Therefore we have no misery within the group. However, the sequence length could be shorter. To sum up, based on the properties mentioned on table 1 this proposed strategy could satisfy *no misery*, *least misery*, and *most pleasure* within the group. It does not satisfy completeness and clear order because of no misery part of the strategy. In addition to that, it does not satisfy the fairness property either because by applying Least Misery and Without Misery a minority opinion can dictate the group [3]: if everybody really wants to listen something, but one person does not like it, then it will never be played.

3.1.2 Algorithm 2: Fairness -> Average. We apply the base strategy Fairness, and for tie-resolving use the Average rating across all users in the group. By tie-resolving we mean that when the rating is the same for multiple items the one with higher average rating will be selected. In this strategy, one person chooses first, then another, until everyone has made one choice. The next rounds usually begin with the one who had to choose last in the previous round. However, if the user’s top two preferences have already been selected in that round we go on to the next person. It continues until all items are consumed [3]. According to Table 1 it satisfies *fairness*, *completeness*, and *clear order* as it does not exclude any item from the sequence.

We believe it will be interesting to compare these two algorithms. Table 1 represents trade-off between these two algorithms, in one having high average satisfaction by excluding the least preferred item(s) of one or more people or have a fair system that might recommend you your most hated item if it is a top item of one member (as long as you get to listen the songs you really love as well). Algorithm 2 considers the satisfaction of all the users, but includes the most hated item if it is a top item of one member. The turns change in each round. This algorithm in group settings can be characterized as a strategy without favoritism or discrimination

Table 3: Fairness -> Average

	A	B	C	D	E	F	G	H	I	J
John	10	4	3	6	10	9	6	8	10	8
Adam	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6
Sum	21	18	13	22	26	26	17	23	20	22

Group List: E, F, A, H, I, D, B, J, C, G

towards specific group members [2], compared to Algorithm 1 where one member could dictate her preferences (as seen in the example).

4 EXPLANATION STYLES

The choice of the used algorithm also influences the types of explanations we can generate. We give examples of how this might look for different explanation categories in Table 4.

To keep a group satisfied during the entire sequence of recommendations we need to consider the preferences of all the people in the group. This can be challenging when the preferences of individual group members diverge. An explanation in such contexts can indicate possible changes of requirements that help improve user satisfaction. In the context of group, such repair-related explanations help group members understand the constraints of other group members and decide in which way their own requirements or preferences should be adapted [2].

Table 4 demonstrates the proposed explanation categories, which we also explain in relation to our paper below:

Repairing versus reassuring. In this study, we proposed to generate explanations for Algorithm 1 based on repairing inconsistency category with *pleasure* as a basis, and for Algorithm 2 based on the same category but *fairness* as a basis. Both describe group disagreement situations. We call the explanations of this category repair explanations. Here is an example of fairness basis: *"The system detected you might not like song 1 but it is the song Mary prefers most. You made your choice in the previous round, now it's Mary's turn"*.

For comparison, we also study the situation where all group members agree on the selected item. In this paper, we call these *reassuring explanations*, which are similar to the positive explanations which have been discussed in Quijano-Sanchez et al.[8] work. For instance, *"The system detected that you all will enjoy this song. Moreover, you and Adam will love it"*.

In our study, we put persuasiveness (as defined by Quijano-Sanchez et al.) under the repair inconsistency category.

Complete and vital. The privacy preserving category is used when the underlying recommendations are aggregated models instead of aggregating recommendations for individual users, this approach constructs a group preference model (group profile) that is then used for determining recommendations. The advantage of applying group preference models is that the privacy concerns of users can be diminished [2].

In this paper, we represent this as complete explanations and explanations with only vital information. With *complete information* we describe the ratings of everyone in the group, however with *vital*

Table 4: Explanation categories and examples.

Categories	Example
Privacy Preserving	<i>A majority thinks that it is a good choice. Some group members think that it is an excellent choice [2].</i>
Repairing Inconsistency (Persuasiveness)	<i>Although your preference for this item is not very high, your close friend X (who you highly trust) thinks it is a very good choice [8].</i>
Repairing Inconsistency (Fairness)	<i>The interest dimensions favored by user u1 has been given more consideration since u1 was at a disadvantage in previous decisions [2].</i>
Repairing Inconsistency (Pleasure)	<i>Item y is recommended because nobody hates it in the group due to the lowest rating determined for user a and support the the highest rating determined for user b.</i>
Reassuring	<i>Additionally, Jaime, who you trust the most, would really love this movie, so why not give it a try [8].</i>

information we only report partial information. More specifically, for the least misery part of the strategy, we report the member of the group with the minimum personal value score for the item, i.e., the member that is responsible for this selection. Similarly, for the most pleasure part of the algorithm, we report the member of the group with the maximum personal value score for the item. Finally, for the fairness strategy, we report with each item the member of the group whose turn it is, i.e., the member direct towards this selection [5]. Following examples are represented as complete and vital information respectively, "You, Mary and Adam have rated song 5 with values 4, 10, 5 respectively. Song 5 is recommended because it avoids dissatisfaction within the group due to the lowest rating determined for you and supports the highest rating determined for Mary." and "Song 5 is recommended because it avoids dissatisfaction within the group due to the lowest rating determined for you and supports the highest rating determined for Mary".

5 STRUCTURED INTERVIEWS

In the previous section, we introduced two algorithms for generating sequences of recommendations for groups of users. These naturally influence the explanations that are generated. In addition, whether there is a disagreement in preference will also influence the resulting explanation; in this paper, we study *repair* (the group disagrees) and *reassuring* (the group agrees) explanations.

This is a formative and exploratory evaluation with an aim to study how explanations should be designed to maximize satisfaction even when no consensus exists.

We used the layered evaluation proposed by Paramythis et al.[7], which suggest that for effective adaptation, the process needs to be decomposed and evaluated in layers. This ensured accurate input to the explanation presentation layer. To create a controlled experiment we used synthetic ratings for individual users. The ratings

could be potentially the output prediction of any recommendation algorithm, such as Collaborative Filtering, Content-based filtering and so on. If we chose any particular algorithm, the quality of the prediction would affect the quality of the sequence and would affect the quality of the explanation.

5.1 Study Design

In a structured interview¹ participants were asked to assume that they would be listening to a playlist with two of their friends during their travel sitting in a car. Each participant conducted the particular individually (with the interviewer).

They were given a sample of individual ratings (based on synthetic data) for 10 songs just for themselves, not for their friends. They were told that the system has selected a sequence of songs for them and has provided an explanation for the selected sequence.

Next, they were asked how satisfied they are with the presented explanation, what can be made better, or what they liked about that explanation as well as how it affects their satisfaction for the recommended song.

The sequences resulting from the two proposed aggregation strategies were at most 10 songs or less because in some cases strategy resulted in a short sequence. We used "you" when referring to the participant. We explained that their real names would be replaced with their names in the real explanation, and that their real friend's names would be used in the place of "Adam" and "Mary".

5.2 Procedure

The main user task was to "report her satisfaction degree regarding the proposed explanations in different scenarios". In addition to that she gave her feedback on what can be made better or what she liked about that explanation.

The *independent variables* manipulated in this interview were: **Explanation style:** repair or reassuring explanations (2) * only vital information versus complete information (2).

Scenarios: two for each algorithm, & one where all users agree (5).

These were studied in a within-subjects design with each participant seeing all versions. To control for order effects, the scenarios and explanation styles were counterbalanced across participants.

5.3 Explanation Category

We presented four types of explanations:

- (1) Repair-related explanation with vital information
- (2) Repair-related explanation with complete information
- (3) Reassuring explanation with vital information
- (4) Reassuring explanation with complete information

5.4 Scenarios

Users were asked to imagine that they were listening to the playlist with two friends in a car during a roadtrip. The different scenarios studied were:

Sc 1: A song that the user hate has been selected resulting from *Algorithm 2*.

Sc 2: The song(s) that the user really likes has not been selected at all resulting from *Algorithm 1*.

¹<https://goo.gl/DA7Kmf>, retrieved April 2018

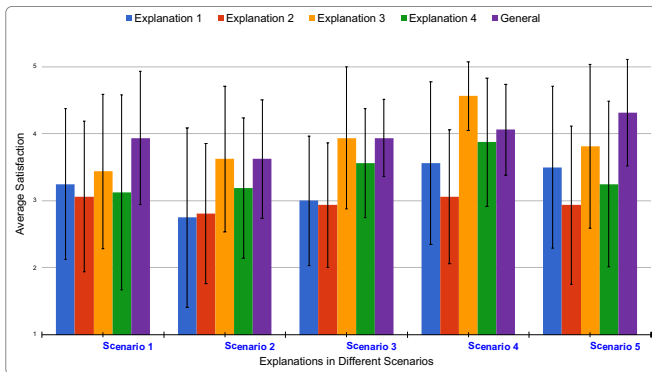


Figure 1: Average satisfaction, by different explanation types and scenarios. Whisker plot depicts 1 SD.

Sc 3: The song(s) that the user really likes has not been selected yet resulting from *Algorithm 1*.

Sc 4: All group members agree on the selected song (Baseline).

Sc 5: It is the user turn to pick and her favorite song has been selected resulting from *Algorithm 2*.

6 RESULTS

Designing explanations that improved user satisfaction was the goal in this study. We have proposed different types of explanations based on different sequence constructing algorithms, and we investigated user impressions of these explanations in different scenarios. Figure 1 summarizes the results by explanation and scenario. The vertical axis shows average satisfaction for each explanation per scenario. Moreover, the error bars indicate the standard deviation (SD) of these results. Due to the small sample size and that this study is exploratory, we have not performed statistical analysis.

6.1 Participants

Sixteen participants from the staff and student population of Delft University of Technology participated voluntarily in the experiment. They were at least 18 years of age, and 20% female.

6.2 Which Explanation Performed Better

Comparing between the aforementioned four types of explanations, explanation 3 (reassuring with vital information) performed better in terms of satisfaction *regardless of the scenario* in which it was presented. The average satisfaction for explanation 3 are ($m=3.4$, $SD=1.15$), ($m=3.6$, $SD=1.09$), ($m=3.9$, $SD=1.06$), ($m=4.6$, $SD=0.51$), ($m=3.8$, $SD=1.2$), in scenarios 1, 2, 3, 4, and 5 respectively. In particular, explanation 3 in scenario 4 has the highest average satisfaction ($m=4.6$, $SD=0.51$). In addition to scoring the explanations, we asked participants why they liked that specific explanation and why not. Some reasons that they liked explanation 3 are as follows: "The explanation is easy to understand", "The encouraging tone.", "Nice, friendly, clear and short", "The explanation is short and concise". The traits were mostly mentioned by participants include brevity, simplicity, friendly tone, as well as clear and understandable content.

6.3 Influence of Explanation Category

We compared *vital information* (explanations 1,3) vs *complete information* (explanations 2,4). We found that for *all scenarios, except scenario 2* vital information led to more satisfied participants.

In contrast, the satisfaction for scenario 2 is slightly higher for explanation 2 (compared to explanation 1), with the complete repair explanation. In the case of scenario 2, the increased complexity of the complete information may help users to deal with missing a song they really like. However, both explanations 1 and 2 have low scores. Note that explanation 3 (reassuring-vital) still outperforms explanation 4 (reassuring-complete) also for scenario 2. I.e., this is similar to the other scenarios.

Next, we compared *repair-related explanation* vs *reassuring explanation*, and found that reassuring explanation performed better rather than repair-related explanation. According to the users' feedback we can infer that they preferred to receive positive and encouraging explanations rather than receiving explanations showing misery or dissatisfaction of any of the group members.

6.4 Influence of Scenarios

Overall, participants were more satisfied with the explanations in *scenario 4* (Assume all group members like the selected song) and *scenario 5* (Assume it is your turn and you got your favourite song). This can be expected as these are positive scenarios for the users. At the end of each scenario we asked participants "how the explanation influenced their satisfaction regarding the selected song in general", results are demonstrated as General in figure 1. Scenario 5 has higher general (with the song) satisfaction than scenario 4 when comparing across all explanation styles ($m=4.31$, $SD=0.8$). This suggest that users care more about their own preferences than global satisfaction in the group.

In scenario 1 (Assume a song that you really hate is now playing.) the difference between average satisfaction of explanation 1 and 3 is small with values ($m=3.25$, $SD=1.12$) and ($m=3.4$, $SD=1.15$) respectively. Some comments for explanation 1 in this scenario are "Sad result, but the explanation makes it a bit better.", "It provides proper reasoning as to why the song was selected.", "It acknowledges I don't like the song." or "Seems fair! I'm willing to let them enjoy." and feedback for the explanation 3 include: "It doesn't acknowledge my dislike, but I like the part that my friends will like it." or "It's short and informal."

Although we only asked users about their impressions of the explanations, they also gave feedback regarding the applied algorithm. This was mostly for scenario 1 where we applied algorithm 2: "It feels strange that the song is chosen only because it's Mary's favorite song. I would expect a solution where none of the extreme valued songs are chosen to keep the overall satisfaction of both of us higher.", "The songs that anyone likes a little should be kept to the last, even if it is someone's favourite.". The users' feedback illustrate that algorithm 2 was found to be less satisfying than algorithm 1.

6.5 Influence of Wording

The results suggest that explanation type 1 in scenario 2 has the lowest average satisfaction but with the highest SD ($m=2.75$, $SD=1.34$). This is the explanation: "Song 5 is recommended because it avoids

Table 5: Possible explanation purposes for groups

Aim	Definition
Transparency	Explain how the system works [9]
Trust	Increase users' confidence in the system [9]
Privacy-Preserving	Preserving users' confidential data, like their preferences
Effectiveness	Help users make good decisions [9]
Persuasiveness	Convince users to try or buy [9]
Group Satisfaction	Increase the average ease of use or enjoyment of all group members
Individual Satisfaction	Increase the ease of use or enjoyment of each member of the group
Single Item Satisfaction	Increase the average ease of use or enjoyment of all group members for each single item
Several Items Satisfaction	Increase the average ease of use or enjoyment of all group members for several items

misery within the group due to the lowest rating determined for you and support the highest rating determined for Mary."

This result suggests that although the average satisfaction is low for this explanation participants' opinion vary about that. Some reasons that participants mentioned are *"The word "misery" is too strong."* or *"The explanation sounds a bit complicated. I have to read it twice to understand..* Therefore it was mainly due to words we used like 'misery'. In addition to that, positive feedback were also given, such as: *"It shows me that it knows that it's not my favorite song but also tries to minimize misery."*, or *"At least it explains the reasoning."*

It can be interpreted as people are prefer to receive more friendly and light explanations rather than explanations with complicated words to describe the algorithm behind the sequence generation.

In addition to that, explanation 2 in the same scenario (scenario 2) performed slightly better ($m=2.81$, $SD=1.05$), as this more complete explanation contains ratings which helped users understand the explanation better.

6.6 Other Comments

Additional comments related to individual participants are as follows: *"It depends on my personality and mood. Example: if I am in my car with friends in summer and there is sunshine I could be happily let others favorite songs play even if I hate that. But if it's winter and I'm sad, I can't accept it easily."* Other comments about complete explanations vary between participants e.g., some have comments like *"Good to know about the ratings."* but on the other hand for the same explanations others have comments like *"My friends rating is not so interesting, it's sort of privacy violation."* or *"I would not be comfortable with the system giving out my rating."*

7 CONCLUSION AND FUTURE WORK

In this paper, we proposed two improved algorithms for constructing sequences of recommendations. We then suggested different explanations styles that could be *reassuring* or *repairing*. Participants preferred short, simple, informal and friendly and encouraging explanations rather than long, complex and negative explanations. However, when maximal misery (not getting their liked item at all) was expected, a more complicated explanation was acceptable. While intuitive, our results give an empirical basis for requirements for explanations of sequences.

Our next steps will be to study the behavior of groups in a joint setting, with all group members present. We will evaluate whether the simpler explanations are effective only for the "active user", and what the effect is on their group members. This is reflected in the evaluation criteria proposed in Table 5. We are also working on automatically generating these explanations from ratings so that we can evaluate their effect in a more systematic way.

REFERENCES

- [1] Liliana Ardissono, Anna Goy, Giovanna Petrone, Marino Segnan, and Pietro Torasso. 2003. Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices. *Applied artificial intelligence* 17, 8-9 (2003), 687–714.
- [2] Alexander Felfernig, Ludovico Boratto, Martin Stettinger, and Marko Tkalčić. 2018. Explanations for Groups. In *Group Recommender Systems*. Springer, 105–126.
- [3] Judith Masthoff. 2004. Group modeling: Selecting a sequence of television items to suit a group of viewers. In *Personalized digital television*. Springer, 93–141.
- [4] Thuy Ngoc Nguyen and Francesco Ricci. 2018. Situation-Dependent Combination of Long-Term and Session-Based Preferences in Group Recommendations: An Experimental Analysis. In *SAC*.
- [5] Eirini Ntoutsi, Kostas Stefanidis, Kjetil Nørvg, and Hans-Peter Kriegel. 2012. Fast group recommendations by applying user clustering. In *International Conference on Conceptual Modeling*. Springer, 126–140.
- [6] Ingrid Nunes and Dietmar Jannach. 2017. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction* 27, 3-5 (2017), 393–444.
- [7] Alexandros Paramythis, Stephan Weibelzahl, and Judith Masthoff. 2010. Layered evaluation of interactive adaptive systems: framework and formative methods. *User Modeling and User-Adapted Interaction* 20, 5 (2010), 383–453.
- [8] Lara Quijano-Sanchez, Christian Sauer, Juan A Recio-Garcia, and Belen Diaz-Agudo. 2017. Make it personal: a social explanation system applied to group recommendations. *Expert Systems with Applications* 76 (2017), 36–48.
- [9] Nava Tintarev and Judith Masthoff. 2007. Effective explanations of recommendations: user-centered design. In *Proceedings of the 2007 ACM conference on Recommender systems*. ACM, 153–156.