Journal of Universal Computer Science, vol. 26, no. 3 (2020), 402-431
submitted: 25/9/19, accepted: 15/3/20, appeared: 28/3/20

Social Choice-based Explanations: An Approach to Enhancing Fairness and Consensus Aspects

Thi Ngoc Trang Tran, Muesluem Atas, Man Viet Le
Ralph Samer, Martin Stettinger
(Institute of Software Technology, Graz University of Technology
Inffeldgasse 16B/2, 8010 Graz, Austria
{ttrang;muesluem.atas;vietman.le;rsamer;martin.stettinger}@ist.tugraz.at)

Abstract: Explanations are integrated into recommender systems to give users an insight into the recommendation generation process. Compared to single-user recommender systems, explanations in group recommender systems have further goals. Examples thereof are *fairness*, which helps to take into account as much as possible group members' preferences and *consensus*, which persuades group members to agree on a decision. In this paper, we proposed different types of explanations and found the most effective ones in terms of increasing the *fairness perception, consensus perception* and *satisfaction* of group members with regard to group recommendations. We conducted a user study to evaluate the proposed explanations. The results show that explanations which consider the preferences of *all* or the *majority* of group members achieve the best results in terms of the mentioned dimensions. Besides, we discovered positive correlations among these aspects. In the context of repeated decisions, group members' satisfaction from previous decisions are helpful to improve the fairness perception of users concerning group recommendations and speed up the group decision-making process. Furthermore, we found out that gender diversity does influence the perception of users regarding the mentioned dimensions of the explanations. Although the proposed explanations were analyzed in group decision scenarios for non-configurable (no-attribute) items, there exist potential possibilities to apply them to explanations for *configurable items*.

Key Words: social choice theory, preference aggregation strategies, social choice-based explanations, group recommender systems, group decision making, social aspects, group composition
Category: H.1.2, H.4.0, H.5.2, J.4

1 Introduction

Recommender systems are helpful for users to choose items that fit their interests and desires. However, these systems often appear as "*black-boxes*", which prevent users from comprehending the underlying mechanisms of the recommendation generation process. Recently, explanations have been included in recommender systems to serve different goals, such as help users to understand the reasoning behind a recommendation (*transparency*), to make better decisions (*effectiveness*), to quickly decide which item is the best for them (*efficiency*), to make more accurate decisions (*satisfaction*), and to increase their trust and acceptance of recommended items (*trust* and *persuasiveness*) [Bilgic, 2005, Chen et al., 2013, Jannach et al., 2010, Tintarev and Masthoff, 2007, Tintarev and Masthoff, 2010, Tintarev and Masthoff, 2012].

While extensive research has been grown in this topic, most of them target at explanation approaches for single-user recommender systems [Chen and Pu, 2012, Gedikli et al., 2014, Herlocker et al., 2000]. Only a few existing studies give in-depth analyses of explanations in group recommender systems. Group recommender systems have been proven to be efficient for groups of users who jointly decide on a solution that satisfies every group member. Different from single-user recommender systems, group recommender systems should take into account not only the preferences of all group members, but also different social aspects in the group decision-making process such as *group consensus achieving* [Chiclana et al., 2007, Pérez et al., 2018], *conflicts resolving* [Felfernig et al., 2016], and *fairness fostering* among group members [Kacprzyk and Zadrozny, 2016]. These aspects make the group decision-making process more complex and last longer. Therefore, including explanations into these systems is crucial to facilitate and improve the quality of the group decision-making process. Compared to the explanations for single-users, those for groups could have further goals. One example thereof is *fairness*, which helps to increase *the fairness perception* of group members and motivates them to take into account as much as possible the preferences of others. Another goal is *consensus*, which helps to increase the *consensus perception* of group members and helps them to agree on the decision [Felfernig et al., 2018b].

The fairness and consensus aspects of explanations can be differently considered depending on the *decision type* (e.g., *repeated* or *non-repeated decisions*) [Felfernig et al., 2017]. For *non-repeated decisions* in which decisions are rarely repeated (e.g., *"selecting an apartment to buy for the whole family"*), these aspects should be taken into account right in the on-going decision. However, for *repeated decisions* which are periodically repeated by the same group (e.g., *"choosing a movie to watch with friends every weekend"*), these aspects can be considered in the on-going decision and in previous (or future) decisions as well. For instance, to foster *fairness*, group members whose preferences have not been considered in the on-going decision should have higher priorities in future decisions [Stettinger, 2014]. The *consensus* aspect can be defined as an acceptable solution, even if it is not the favorite of every group member. An explanation in such a context indicates a solution that makes the final decision more likely to be accepted by all group members [Hertzberg et al., 2013].

On the other hand, the current literature indicates that the performance of group decision making can be affected by group composition factors, such as *age* and *gender* of group members [Hannagan and Larimer, 2010]. Regarding the gender factor, the authors in [Sanz de Acedo Lizárrage et al., 2007] found out that males are likely to be more objective, realistic, and assertive. Whereas, females tend to be more intuitive, sensitive, and usually look for creative ideas. Because of gender diversity, males and females can have different behavioral patterns in group decision-making processes. Regarding the age factor, these authors claimed that when making the decision, young people are more likely to face emotional and social-aspect pressures compared to adults and retired people. Based on these findings, we assume the age diversity also influences the fairness and consensus perception of users concerning group recommendations.

In this context of explanations for groups, some open questions have been posted [Felfernig et al., 2018b]: (1) *"How to formulate explanations in such a way that helps to enhance the social aspects among group members?"*, (2) *"Which explanation can help to speed up the group decision-making process?"*, and (3) *"In which explanation, group composition triggers different perceptions of users concerning the social aspects in group decision making?"*. In this paper, we try to answer these questions and the contributions of the paper as in the following:

(1) We propose *different types of social choice-based explanations* by intuitively explaining the underlying mechanisms of preference aggregation strategies and taking group members' satisfaction into account.

(2) We investigate *the best social choice-based explanation* in terms of increasing the fairness perception, consensus perception, and satisfaction of users with regard to group recommendations.

(3) We discover *positive correlations* between the perceived fairness/perceived consensus of the explanations and the satisfaction of users with regard to group recommendations.

(4) In the context of *repeated decisions*, we find out that explanations that take into account group members' satisfaction from previous decisions can increase the fairness perception of users concerning group recommendations and speed up the group decision-making process.

(5) We investigate *the influence of age and gender* diversity of group members on fairness, consensus, and satisfaction dimensions of the explanations.

The remainder of the paper is organized as follows. In *Section 2*, we summarize the related work regarding explanations in group recommender systems. In *Section 3*, we introduce different approaches to generate social choice-based explanations. In *Section 4*, we define hypotheses with regard to the mentioned open questions and present the main steps of our user study. The results and discussions regarding the hypotheses are presented in *Section 5*. In *Section 6*, we discuss potential possibilities to apply the proposed explanations to clarify group recommendations for configurable items. Finally, we conclude the paper and discuss open issues for future work in *Section 7*.

2 Related Work

2.1 Explanations in Group Recommender Systems

Explanations for groups are usually generated by explaining the underlying recommendation mechanisms [Felfernig et al., 2018b]. [Ardissono et al., 2003] explained users a recommended tourism attraction by mentioning its positive aspects. For instance, *"attraction X has been recommended to the group since it is very eye-catching and requires low background knowledge"*. [Felfernig et al., 2018b] presented different approaches to explaining group recommendations in the context of *collaborative filtering, content-based filtering, constraint-based,* and *critiquing-based* recommendations. For the *collaborative filtering approach*, the preference of a group G for a specific item can be predicted based on the preferences of the nearest neighbors of group members in G. This way, the explanation of a recommended item in the group-based collaborative filtering scenarios can be formulated as follows: *"Movie X has been recommended to the group since the nearest neighbors of all group members like this movie"*. In the *content-based filtering* approach, explanations are tailored based on *item-related* content. For instance, *"groups that like item X also like item Y"* or *"since the group likes movie X, we also recommend movie Y from the same director"* [Bilgic, 2005, Felfernig et al., 2018b]. For *constraint-based* approach, explanations are used to answer *how*-questions which show the relationship between user requirements and the recommended items [Felfernig et al., 2007]. An example explanation of this approach is *"apartment X is recommended to the group since group members specified the upper monthly rate limit with €700, and all of you preferred staying in the city center"*. Besides, constraint-based group recommender systems are able to answer *why* and *why-not* questions. The explanations of *why* questions provide the reason why a certain question has to be answered [Felfernig et al., 2018b]. The explanations of *why-not* questions help users overcome the *"no solution could be found"* dilemma [Felfernig et al., 2009]. An example of such explanations can be: *"No camera could be found according to the requirements of group members. To resolve this, consider increasing the upper price limit or to decrease the camera resolution"*. For *critiquing-based* approach, recommendations are generated based on the similarity between the candidate and reference items. This way, the explanations of critiquing-based recommender systems describe the relationship between the currently shown reference item and candidate items [McCarthy et al., 2004]. Critiques specified by group members for the current reference item can be used to formulate the explanations [Felfernig et al., 2018b]. An example explanation thereof is *"as expected, apartment Y has an elevator and renting cost is lower than the upper limits specified by group members. Besides, it has a big kitchen and stays in the city center as all group members preferred"*.

The mentioned approaches solely combine the preferences of group members, whereas other social aspects in the group seem to be ignored. Therefore, these approaches do not always guarantee a high satisfaction of every group member. Another approach to generating explanations for groups is to reveal *social choice-based preference aggregation strategies*. These strategies allow merging the preferences of individual group members into a model that represents the inferred preferences of the whole group [Felfernig et al., 2018b, Masthoff, 2011, Senot et al., 2010]. For instance, a textual explanation based on the *Least Misery* strategy can be shaped as follows: *"Item X has been recommended to the group since its group score of 2.7 supports the lowest rating given by user u_a"*. When using the *Majority* strategy, the corresponding explanation would be: *"Item X is recommended to the group since most group members like it"*. In this approach, the aggregation strategies consider some certain aspects in groups and can be selected depending on the item domains. For instance, in *high-involvement* item domains (i.e., domains with high decision efforts), the *Least Misery* strategy is usually chosen to find an item which avoids the misery within the group. In other words, the recommended item can be consumed by every group member without any issues [Felfernig et al., 2017].

In this line of research, [Ntoutsi et al., 2012] proposed some fundamental approaches to explaining aggregation strategies in group recommendation contexts. However, to some extent, the integration of social choice theories into group explanations is still an open issue. In this paper, we provide a more in-depth analysis of this approach by proposing textual explanations describing different types of preference aggregation strategies. Besides, we investigate which explanation performs the best in terms of fostering social aspects in the group decision-making process.

2.2 Social Aspects in Explanations for Groups

As mentioned in *Section 1*, group explanations should consider some social aspects in groups, such as the *fairness* and *consensus* perception of users. These two aspects could benefit group recommendations for achieving a high satisfaction of individual group members.

In this context, a paramount concern is that explanations should show how the preferences of group members are taken into account [Felfernig et al., 2018b]. Most of the existing studies seek to shed light on the influence of the mentioned aspects on group decision outcomes [Castro et al., 2015, Palomares et al., 2014, Quijano-Sanchez et al., 2013, Stettinger, 2014]. Unfortunately, the issue of how explanations could help to increase the fairness perception, consensus perception, and satisfaction levels of users with regard to group recommendations has not been adequately studied. In the current literature, these exist only a few research contributions which give an insight into the mentioned issue. For instance, [Kapcak et al., 2018] proposed an approach to generate group explanations in the tourism domain by explaining underlying preference aggregation strategies. After that, an *automated crowd-sourcing pipeline* utilizes the wisdom of crowds to improve the quality of the generated explanations and increase users' satisfaction with group recommendations. In another research, [Najafian and Tintarev, 2018] proposed explanations for groups using algorithms that combine different aggregation strategies. Besides, they investigated how the explanations can improve the *satisfaction* of users with recommended items.

The two approaches mentioned above can only be applied in the context of sequential recommendations and the social aspects (e.g., fairness and consensus) of the explanations were not discussed. To the best of our knowledge, up to now, *"which explanation generation approach performs the best in terms of fairness and consensus aspects"* is still an open issue. Moreover, there does not exist any research which provides an adequate analysis of the *efficiency* of group explanations. No research has found explanations that help to speed up the group decision-making process. On the other hand, how the group composition factors (such as *age* and *gender* of group members) influence the mentioned aspects of explanations has not been studied yet. Our contribution of this paper is to propose social choice-based explanations that help to bridge all these gaps.

3 Social Choice-based Explanations

3.1 Social Choice-based Preference Aggregation Strategies

In group recommender systems, recommendations are determined by aggregating the preferences of individual group members using preference aggregation strategies. These strategies are grouped into three categories: *consensus-based, majority-based,* and *borderline strategies* [Senot et al., 2010]. In this study, for each category, we only selected *some strategies* (but not all) which are feasible to tailor the explanations and easily comprehended by users.

Consensus-based aggregation strategies represent aggregation mechanisms which consider the preferences of *all group members* [Felfernig et al., 2018a, Masthoff, 2011, Senot et al., 2010]. In this study, we chose *Additive Utilitarian (ADD)* and *Fairness (FAI)* strategies as the representatives of this category. The *ADD* strategy recommends the item with the maximum sum of individual group members' evaluations. The *FAI* strategy is usually applied in repeated decisions in which the same group of users periodically repeats a decision. This strategy ranks items as if individuals are choosing them in turn [Masthoff, 2011].

Majority-based aggregation strategies represent aggregation mechanisms which focus on the most popular items [Masthoff, 2004, Senot et al., 2010]. In this paper, we chose the *Approval Voting (APP)* strategy as the representative of this category. This strategy recommends the item with the highest number of evaluations, which are greater than a *threshold*. The threshold can be predefined by the system or by the group.

Borderline strategies represent aggregation mechanisms that take into account only a subset of individual group members' preferences [Senot et al., 2010]. We chose *Least Misery (LMS), Majority (MAJ)*, and *Most Pleasure (MPL)* as the representatives of this category. The *LMS* strategy recommends the item with the highest of all lowest individual evaluations. The *MAJ* strategy recommends the item with the highest of all evaluations representing the majority of item-specific evaluations. The *MPL* strategy recommends the item with the highest of all individual evaluations.[1]

3.2 Social Choice-based Explanations

In this paper, we propose *three types* of textual explanations which are described in the following:

Type 1 - Based on Preference Aggregation Strategies: These explanations are generated to explain the underlying mechanisms of social choice-based preference aggregation strategies. Our purpose is to generate *textual explanations*, which are easily comprehended by users. As mentioned in *Section 3.1*, we chose *five* aggregation strategies (i.e., *ADD, APP, LMS, MAJ,* and *MPL*) which can be represented via textual explanations. The *FAI* strategy is not selected for this type since it can only be applied to repeated decisions and requires the information of group members' preferences from previous (or future) decisions. The templates of the explanations of *Type 1* are presented in Table 1.

Type 2 - Based on Preference Aggregation Strategies & Decision History: All selected aggregation strategies (mentioned in *Section 3.1*) are applied to tailor the explanations of *Type 2*. These explanations are proposed for scenarios in which decisions are repeated periodically by the same group (e.g., a group of family members decides on a movie to watch every month). Thereby, compared to the explanations of *Type 1*, those of *Type 2* additionally consider *decision history* indicating who was treated less favorably in previous decisions. The templates of these explanations are presented in Table 2.

Type 3 - Based on Preference Aggregation Strategies & Future Decision Plan: Similar to *Type 2*, the explanations of *Type 3* are also applied to repeated decisions. Each explanation additionally includes a *future decision plan* in which the preferences of disadvantaged group members from the on-going decision will have higher priorities in upcoming decisions. We use the strategies mentioned in *Type 2* to generate the explanations of *Type 3*. The templates of the explanations of *Type 3* are presented in Table 3.

[1] For a detailed discussion of the mentioned strategies, we refer the readers to [Felfernig et al., 2018a].

Table 1: *Type 1* - Explanations based on social choice-based preference aggregation strategies.

Strategies	Explanation templates
ADD-based	*"Item X has been recommended to the group since it achieves the highest total rating."*
APP-based	*"Item X has been recommended to the group since it achieves the highest number of ratings which are above a threshold th."*
LMS-based	*"Item X has been recommended to the group since no group member has a real problem with it."*
MAJ-based	*"Item X has been recommended to the group since most group members like it."*
MPL-based	*"Item X has been recommended to the group since it achieves the highest of all individual group members' ratings."*

Table 2: *Type 2* - Explanations based on social choice-based preference aggregation strategies and group members' satisfaction from previous decisions.

Strategies	Explanation templates
ADD-based	*"...gray the description of the strategy as shown in Type 1... This decision supports the preferences of users u_a, u_b, and u_c who were treated less favorably in the last n decisions."*
FAI-based	*"The preference of user u_a was not considered in the last n decisions. Therefore, in this decision, item X has been recommended to the group since he/she likes it the most."*
APP-based	*"...gray the description of the strategy as shown in Type 1... This decision supports the preferences of users u_a, u_b, and u_c who were treated less favorably in the last n decisions."*
LMS-based	*"...gray the description of the strategy as shown in Type 1...This decision supports the preferences of users u_a and u_b who were treated less favorably in the last n decisions."*
MAJ-based	*"...gray the description of the strategy as shown in Type 1...This decision supports the preferences of users u_a, u_b, and u_c who were treated less favorably in the last n decisions."*
MPL-based	*"...gray the description of the strategy as shown in Type 1... This decision supports the preference of user u_a who was treated less favorably in the last n decisions."*

Table 3: *Type 3* - Explanations based on social choice-based preference aggregation strategies and future decision plan.

Strategies	Explanation templates
ADD-based	*"...gray the description of the strategy as shown in Type 1... The preference of user u_a seems not to be considered in this decision. Therefore, all group members agreed that he/she will have a higher priority in the next decision."*
FAI-based	*"Item X has been recommended to the group since user u_a likes it the most. However, all group members agreed that the preferences of other group members will be taken into account in turn in the next decisions."*
APP-based	*"...gray the description of the strategy as shown in Type 1... The preference of user u_a seems not to be considered in this decision. Therefore, all group members agreed that he/she will have a higher priority in the next decision."*
LMS-based	*"...gray the description of the strategy as shown in Type 1... The preferences of users u_a and u_b seem not to be considered in this decision. Therefore, all group members agreed that these two users will have higher priorities in the next decisions."*
MAJ-based	*"...gray the description of the strategy as shown in Type 1... The preference of user u_a seems not to be considered in this decision. Therefore, all group members agreed that he/she will have a higher priority in the next decision."*
MPL-based	*"...gray the description of the strategy as shown in Type 1... The preferences of users u_a, u_b, and u_c seem not to be considered in this decision. Therefore, all group members agreed that these three users will have higher priorities in the next decisions."*

4 Hypotheses and User Study

4.1 Hypotheses

One of our goals is to discover explanations that effectively help to increase the *fairness perception, consensus perception,* and *satisfaction* of users with regard to group recommendations. We assume *ADD-based explanations* would perform the best since these explanations describe a group recommendation strategy considering the preferences of all group members. This leads us to the frist hypothesis H_1, which helps to find out the most effective explanation approach to boost the quality of group recommendations.

Hypothesis H_1: *"ADD-based explanations, which describe a group recommendation strategy taking into account the preferences of all group members,* *best help to increase the fairness perception, consensus perception, and satisfaction of users with regard to group recommendations".*

Another goal of our study is to find out whether *the perceived fairness/consensus of explanations relate to users' satisfaction with group recommendations*. We assume that the higher the fairness (or the consensus) levels, the higher the satisfaction levels. In other words, there could exist positive correlations among these dimensions. This assumption brings us to the second hypothesis H_2.

Hypothesis H_2: *"There exists a positive correlation between the perceived fairness (or the perceived consensus) of explanations and users' satisfaction with regard to group recommendations".*

In the context of *repeated decisions*, we try to analyze the influence of the information of previous decisions and future decision plans on the fairness and consensus perceptions of users concerning group recommendations. We assumed that adding the mentioned information to the explanations could increase the fairness and consensus perceptions of users and therefore convinces them to accept the recommended items. This motivates us to come up with another hypothesis H_3.

Hypothesis H_3: *"In the context of repeated decisions, the integration of group members' satisfaction from previous decisions and future decision plans into social choice-based explanations increases the fairness and consensus perceptions of users with regard to group recommendations".*

Regarding the *efficiency* of the explanations, we want to examine if the time of users' perception process differs depending on the explanation type. Therefore, we propose the next hypothesis H_4, in which we suppose *as the time of the users' perception process decreases, so does the time of group decision-making process.* The analysis of this hypothesis helps to find out which explanation significantly speeds up the group decision-making process.

Hypothesis H_4: *"The time that users spend to evaluate the fairness, consensus, and satisfaction dimensions of the explanations differ depending on the explanation type".*

Finally, we want to investigate the influence of *age* and *gender* diversities on users' perceptions in terms of the mentioned dimensions. Users with different age ranges and genders are assumed to have different evaluations for the explanations. To test this assumption, we propose the last hypothesis H_5. The analysis of this hypothesis helps us to find out which groups of users achieve the highest levels of the three mentioned dimensions.

Hypothesis H_5: *"The fairness perception, consensus perception, and satisfaction of users with regard to group recommendations differ depending on their age and gender".*

4.2 User Study Design

To examine the proposed hypotheses, we conducted a user study with staff members and students at two universities[2]. In total, there were *135* participants (*males*: 54.81%, *females*: 45.19%) from 20 to 51 years old. The participants were chosen using a *random sampling* method in which each participant had an adequate and independent opportunity of being chosen. Our user study was designed and conducted in the following steps:

Table 4: An example group decision task in the restaurant domain. Each group member explicitly evaluated restaurants using a 5-star rating scale (1 - the worst, 5 - the best).

	Rest A	*Rest B*	*Rest C*
Alex	2	1	5
Anna	2	4	1
Sam	4	4	1
Leo	4	4	1

Step 1 - Define a group decision scenario in the *restaurant* domain: "Assume, there is a group of four friends (Alex, Anna, Sam, and Leo). Every month, a decision is made by this group to decide on a restaurant to have dinner together. In this decision, each group member explicitly rated three restaurants (*Rest A, Rest B,* and *Rest C*) using a rating scale ranging from 1 (*the worst*) to 5 (*the best*). The ratings given by group members are shown in Table 4".

Step 2 - Explanation generation: The explanation templates proposed in *Section 3.2* were used to formulate the explanations. Some information in the explanation templates (e.g., *names of items, names of group members*, and *the number of previous decisions*) was accordingly adapted to make them appropriate for the proposed scenario. In total, we formulated *17* explanations for three explanation types (*Type 1*: *five*, *Type 2*: *six*, and *Type 3*: *six*). For the rest of the paper, we focus on analyzing these explanations.

Step 3 - Distribute explanations to the participants: We provided each participant with a sequence of *six* different explanations corresponding to six strategies (*ADD, FAI, APP, LMS, MAJ,* and *MPL*). Each explanation in the sequence could be either from *Type 1, Type 2*, or *Type 3*. This way, each participant received a different sequence of explanations. Besides, it also made the numbers of participants for each explanation in each type balanced. The explanations in each sequence were shown to the participants in random order to avoid possible biases. Moreover, at any given time, the participants

[2] Graz University of Technology - Austria and Hue University of Economics - Vietnam

read and evaluated "*only one*" explanation. The participants' evaluations for the explanations were *independent* of each other, which means the evaluations for one explanation did not rely on those for other explanations.[3]

Step 4 - Evaluate the explanations: Each participant had to evaluate the explanations according to three dimensions: *fairness, consensus,* and *satisfaction.* These dimensions were represented in the form of claims as follows:

- **Perceived fairness**: *"The explanation convinces you that the group recommendation is fair to group members."*

- **Perceived consensus**: *"The explanation helps group members agree on the group recommendation."*

- **Satisfaction**: *"The explanation helps to increase the satisfaction of group members with regard to the group recommendation."*

Each participant read the claims and provided his/her evaluations for the dimensions using a 5-point Likert scale ranging from 1 - *completely disagree* to 5 - *completely agree.* Besides, we measured the *total time* (in *second*) which each participant has spent to evaluate all dimensions of the explanation.

5 Data Analysis Results and Discussions

In this section, we provide data analysis methods[4], results, and discussions of the proposed hypotheses.

5.1 Hypothesis H_1

Method: To examine the hypothesis H_1, for each explanation in each type, we collected *three evaluation sets* corresponding to the *three dimensions.* These evaluations are *ordinal variables* ranging from 1 to 5. They are *independent* of each other since the evaluations of one explanation did not rely upon those of other explanations. Besides, they are *not normally distributed* (*Shapiro-Wilk tests*, $\alpha = .05$, and *p values* $< \alpha$). Based on the characteristics of the data, for each explanation type, we ran three *Kruskal-Wallis tests* ($\alpha = .05$) to examine whether there were statistically significant differences in the *fairness, consensus,* and *satisfaction levels* across different explanations. Additionally, we inspected the *mean ranks* received from these tests to identify the best explanation. *The best explanation achieves the highest mean rank.*

Besides, we performed follow-up *Mann-Whitney U tests* ($\alpha = .05$) between pairs of explanations to find out which of the explanations were significantly different from one another. We ran 10 different pairwise tests ($C_5^2 = 10$) for

[3] In this user study, the participant was not a group member of the mentioned group decision scenario. Instead, he/she played the role of a consultant who analyzed the group decision scenario and evaluated the explanations.

[4] All the tests in this paper were performed in the SPSS V.22.

Type 1 and *15* pairwise tests ($C_6^2 = 15$) for *Type 2* or *Type 3*. Running many Mann-Whitney U tests on the same evaluation sets could cause *Type I errors*[5]. To control these errors, before running the Mann-Whitney U tests, we applied a *Bonferroni adjustment* [Pallant, 2007] to adapt the significance level. The revised significance levels of the Mann-Whitney tests were $\alpha' = \alpha/10 = .005$ for *Type 1* and $\alpha' = \alpha/15 = .003$ for *Type 2* and *Type 3*.

Table 5: *p values* (*2-tailed*) of Kruskal-Wallis tests on *fairness, consensus,* and *satisfaction levels* across different explanations.

	fairness	*consensus*	*satisfaction*
Type 1	**.000**	**.000**	**.000**
Type 2	**.005**	**.000**	.056
Type 3	**.000**	**.000**	**.000**

Table 6: *p values* (*2-tailed*) of Mann-Whitney U tests between the *MPL-based/FAI-based* explanation and one of other explanations in *Type 1* and *Type 3*.

		MPL vs. ADD	*MPL vs. APP*	*MPL vs. LMS*	*MPL vs. MAJ*	*FAI vs. ADD*	*FAI vs. APP*	*FAI vs. LMS*	*FAI vs. MAJ*	*FAI vs. MPL*
Type 1 ($\alpha' = .005$)	*fairness*	**.000**	**.000**	**.000**	**.000**					
	consensus	**.000**	**.000**	**.000**	**.000**					
	satisfaction	**.000**	**.000**	**.000**	**.000**					
Type 3 ($\alpha' = .003$)	*fairness*	**.000**	**.000**	**.000**	**.000**	**.000**	**.000**	**.001**	**.000**	.022
	consensus	**.000**	**.000**	**.000**	**.000**	.016	.071	.085	.005	**.000**
	satisfaction	**.000**	.005	.097	**.000**	**.000**	**.003**	.116	**.000**	.703

Results: The Kruskal-Wallis tests show that in *Type 2*, there were no statistically significant differences in the participants' satisfaction levels across different explanations (see Table 5). In other words, the explanations of *Type 2* did not increase the participants' *satisfaction* with group recommendations.

In *Type 1* and *Type 3*, we found statistically significant differences in *fairness, consensus,* and *satisfaction* levels across different explanations (see Table 5). The follow-up Mann-Whitney U tests show that *MPL-based* and *FAI-based* explanations triggered these differences (see Table 6). By inspecting the mean ranks, we found out that *ADD-based* and *MAJ-based* explanations seemed to achieve the highest fairness, consensus, satisfaction levels in both *Type 1* levels in *Type 3* (see Table 7). Besides, the *APP-based* explanation also performed well and the mean ranks of this explanation are quite close to those of the *ADD-based* and

[5] In hypothesis testing, a *Type I error* involves rejecting the null hypothesis (e.g., there are no differences among the groups) when it is actually true [Pallant, 2007].

Table 7: *Mean ranks* generated in the Kruskal-Wallis tests for all explanations in *Type 1* and *Type 3*.

Explanations		*fairness*	*consensus*	*satisfaction*
Type 1	*ADD*	**143.61**	**143.51**	132.15
	APP	136.24	125.94	130.79
	MAJ	131.50	131.83	**132.47**
	LMS	114.85	118.68	117.72
	MPL	57.16	63.49	70.07
Type 3	*ADD*	182.90	174.67	**183.82**
	APP	184.68	165.54	167.87
	MAJ	**185.89**	**181.24**	182.08
	LMS	164.28	163.72	144.13
	MPL	83.24	80.92	115.32
	FAI	113.90	136.53	119.24

MAJ-based explanations (see Table 7). Based on the results of Mann-Whitney U tests, we found out that there were no significant differences in the participants' evaluations regarding the mentioned dimensions between pairs of these three explanations (see Table 8). That means, *ADD-based, APP-based*, and *MAJ-based* explanations achieved the best performance.

Table 8: *p values* (*2-tailed*) achieved from Mann-Whitney U tests between pairs of *ADD-based, APP-based*, and *MAJ-based* explanations in *Type 1* and *Type 3*.

Explanations		*fairness*	*consensus*	*satisfaction*
Type 1 ($\alpha' = .005$)	*ADD vs. APP*	.547	.134	.897
	ADD vs. MAJ	.329	.340	1.000
	APP vs. MAJ	.712	.638	.908
Type 3 ($\alpha' = .003$)	*ADD vs. APP*	.804	.615	.403
	ADD vs. MAJ	.736	.666	.870
	APP vs. MAJ	.954	.350	.497

Discussion: *ADD-based* explanations expose a group recommendation strategy taking into account as much as possible preferences of *all* group members. Thus, these explanations convince the participants that the group recommendation is a *fairness-oriented solution*. Besides, a recommendation considering the preferences of all individuals would create a *consensus* among group members [Senot et al., 2010]. This explains why these explanations achieved high consensus levels. Moreover, considering all group members' preferences is the premise to generate a more favorable recommendation that acquires higher satisfaction levels of the participants with regard to the group recommendation. Also, *APP-based* and *MAJ-based* explanations performed the best since they describe the *majority rule* which considers most group members' preferences [Hastie and Kameda, 2005]. This way, these explanations helped to increase the fairness and consensus perceptions of the participants. As fairness and consensus levels increase, so do satisfaction levels (see the result of Hypothesis H_2).

Main results: Hypothesis H_1 can be confirmed for the *ADD-based* explanations of *Type 1* and *Type 3*. These explanations describe a group recommendation strategy considering the preferences of all individual group members. Besides, the explanations describing the *majority rule* also effectively help to increase the fairness perception, consensus perception, and satisfaction of users with group recommendations.

5.2 Hypothesis H_2

Method: For this hypothesis, we examined two relationships: (1) *between fairness and satisfaction* and (2) *between consensus and satisfaction*. Before performing these tests, we collected three evaluation sets corresponding to three dimensions. Thereafter, we ran two *Spearman Rank Order Correlation* tests ($\alpha = .05$) on each explanation. These tests investigate the *direction* and the *strength* of a monotonic relationship based on a *correlation coefficient (r)*: $r > 0$ indicates a *positive correlation*, and $r < 0$ indicates a *negative correlation* [Pallant, 2007]. Concerning the *strength* of a relationship, a relationship is *"weak"* if $r \in$ [0.10 .. 0.29], *"moderate"* if $r \in$ [0.30 .. 0.49], and *"strong"* if $r \in$ [0.50 .. 1.0] (see [Cohen, 1988]).

Results and Discussion:

Between *fairness* and *satisfaction*: The Spearman Rank Order Correlations show that, in most explanations (except for *ADD-based* explanation of *Type 1*), there were *positive correlations* between fairness and satisfaction levels ($r > 0$). Furthermore, they revealed a *moderate* or *strong* correlation with each other.

Between *consensus* and *satisfaction*: In most explanations (except for *ADD-based* and *LMS-based* explanations of *Type 2*), Spearman Rank Order Correlation tests reveal *positive correlations* between consensus and satisfaction levels. Besides, an inspection of the correlation coefficients suggests that there were *moderate* or *strong* relationships between these two dimensions.

Main results: Hypothesis H_2 can be completely confirmed for the explanations of *Type 3*. This means, for the explanations consisting of future decision plans, *higher perceived fairness (or perceived consensus) levels of explanations are associated with higher satisfaction levels of users with regard to group recommendations.*

Table 9: Spearman Rank Order Correlations between *perceived fairness/consensus* and *satisfaction levels* of the explanations in *Type 1, Type 2,* and *Type 3*.

Explanations		*fairness vs. satisfaction*		*consensus vs. satisfaction*	
		r	*p(2-tailed)*	*r*	*p(2-tailed)*
Type 1	*ADD*	.290	.051	.466**	**.001**
	APP	.444**	**.002**	.741**	**.000**
	LMS	.562**	**.000**	.461**	**.001**
	MAJ	.581**	**.000**	.421**	**.004**
	MPL	.794**	**.000**	.748**	**.000**
Type 2	*ADD*	.538**	**.000**	.226	.126
	FAI	.390**	**.001**	.500**	**.000**
	APP	.342*	**.019**	.326*	**.026**
	LMS	.356*	**.014**	.242	.101
	MAJ	.616**	**.000**	.523**	**.000**
	MPL	.623**	**.000**	.440**	**.002**
Type 3	*ADD*	.570**	**.000**	.473**	**.001**
	FAI	.457**	**.000**	.373**	**.001**
	APP	.300*	**.046**	.510**	**.000**
	LMS	.815**	**.000**	.642**	**.000**
	MAJ	.441**	**.002**	.468**	**.001**
	MPL	.515**	**.000**	.716**	**.000**

* *Correlation is significant at the .05 level (2-tailed)*
** *Correlation is significant at the .01 level (2-tailed)*

5.3 Hypothesis H_3

Method: For this hypothesis, we examined whether the explanations of *Type 2* and *Type 3* worked better than those of *Type 1* in terms of *perceived fairness* and *perceived consensus*. To examine this, for each explanation in each type, we collected two sets of evaluations corresponding to two dimensions. For each dimension of a specific explanation, we ran two Mann-Whitney U tests: (1) between *Type 1* and *Type 2* and (2) between *Type 1* and *Type 3*. To control the *Type I errors*, we revised the significance level as follows: $\alpha' = \alpha/2 = .025$.

We further inspected the mean ranks generated in the Mann-Whitney U tests to find out if the explanations of *Type 2* and *Type 3* outperform those of *Type 1*. Besides, when testing this hypothesis, we eliminated *FAI-based* explanations since it solely exists in *Type 2* and *Type 3*, not in *Type 1*.

Results: *Between Type 1 and Type 2*: Mann-Whitney U tests show that there were no statistically significant differences in the perceived fairness and perceived consensus levels across *APP-based*, *LMS-based*, and *MAJ-based* explanations (see Table 10). In contrast, in the *ADD-based* explanation, we found a statistically

Table 10: Mann-Whitney U tests ($\alpha' = .025$) for all explanations between *Type 1* and *Type 2*.

Explanations	*fairness*			*consensus*		
	p (2-tailed)	*mean rank* Type 1	*mean rank* Type 2	*p* (2-tailed)	*mean rank* Type 1	*mean rank* Type 2
ADD	.143	50.97	43.12	**.015**	53.49	40.65
APP	.155	51.37	43.63	.811	48.14	46.86
LMS	.713	45.99	47.99	.968	46.89	47.11
MAJ	.810	47.65	46.36	.306	49.77	44.29
MPL	**.013**	40.79	54.21	.103	43.05	51.95

Table 11: Mann-Whitney U tests ($\alpha' = .025$) for all explanations between *Type 1* and *Type 3*.

Explanations	*fairness*			*consensus*		
	p (2-tailed)	*mean rank* Type 1	*mean rank* Type 3	*p* (2-tailed)	*mean rank* Type 1	*mean rank* Type 3
ADD	.301	49.25	43.75	.460	48.46	44.54
APP	.919	46.77	46.22	.912	46.21	46.80
LMS	.797	45.80	47.20	.731	45.58	47.42
MAJ	.732	45.59	47.41	.534	44.85	48.15
MPL	.900	46.82	46.17	.779	45.77	47.27

significant difference regarding the perceived consensus ($p = .015 < \alpha'$). However, the mean ranks in Table 10 show that the perceived consensus levels of this explanation in *Type 2* were lower than those in *Type 1*. This means, the information of disadvantaged users from previous decisions did not improve the perceived consensus of the *ADD-based* explanation. However, in the *MPL-based* explanation, such information significantly improved the participants' fairness perception. The mean ranks obviously show that the participants provided *higher fairness-related evaluations* for this explanation of *Type 2* (see Table 10).

Between Type 1 and Type 3: On both fairness and consensus dimensions, the Mann-Whitney U tests do not reveal any statistically significant differences across the explanations between *Type 1* and *Type 3* (see Table 11). That means future decision plans did not help to increase the participants' fairness and consensus perceptions concerning group recommendations.

Discussion: The *MPL-based* explanation describes a preference aggregation strategy that only supports group members who provided the highest ratings for items. In other words, this strategy recommends an item based on the preferences of *a subset of group members*. This causes the dissatisfaction of some group members who provided lower ratings for the recommended item.

In the proposed group decision scenario (see Table 4), the *MPL* strategy suggests *Rest C* to the group since it achieves the highest of all group members' ratings. This decision only supports the preference of Alex and triggers the dissatisfaction of Anna, Sam, and Leo. In such a situation, if a *MPL-based* explanation of *Type 2* is provided (e.g., *"... this decision supports the preference of Alex, who was treated less favorably in the last three decisions"*), then other group members would be aware of the fairness of the group recommendation. Psychologically, the participant might apply the *Equity Theory* [Tanner, 2018] to analyze the fairness aspect of the explanations. According to this theory, a situation is *equitable* when users who invested similar efforts should receive similar rewards. However, this seems not to be the case in the mentioned scenario. Besides, a user who is aware of an *inequitable treatment* will be emotionally motivated to gain equity [Tanner, 2018]. Thanks to the *MPL-based* explanation, the participants perceived the inequity inside the group, and therefore they gave higher fairness levels for this explanation.

Main results: In the context of repeated decisions, hypothesis H_3 can only be confirmed for the *perceived fairness* of the *MPL-based* explanation of *Type 2*. In other words, only the information about less-favored group members from previous decisions helps to significantly increase users' fairness perception concerning group recommendations. Such information is especially helpful for the explanations which describe a group recommendation strategy taking into account the preferences of a subset of group members.

5.4 Hypothesis H_4

Method: This hypothesis was examined in levels: (1) *explanation level* and (2) *explanation type level*. In the *explanation level*, we examined the difference in the evaluation time for the explanations in the same explanation type. In the *explanation type level*, we examined the difference in the evaluation time for the explanations across different explanation types. In the second level, we only compared the explanations between *Type 2* and *Type 3*. The explanations of *Type 1* were not analyzed since this could be a bias. The explanations of *Type 2* and *Type 3* include more information than those of *Type 1* (see Tables 1-3). Obviously, the participants had to spend more time to evaluate these explanations.

Before performing the tests, in each explanation type, we collected the time values (in *seconds*) that the participants spent on each explanation. These values can vary in a broad range. Therefore, we had to re-scale them to the range of [0..1] using the *max-min normalization* (see Formula 1). These values violate the *normality* (Shapiro-Wilk tests, *p values* $< \alpha$) and the *homogeneity of variances* (*sig.* $< \alpha$). Therefore, we had to use *non-parametric* tests to analyze the data. To test the first level, in each explanation type, we ran a Kruskal-Wallis test across different explanations and checked the mean rank to figure out which explanation is the most efficient. To test the second level, for each explanation, we ran a Mann-Whitney U test between *Type 2* and *Type 3* to see in which explanation type, the participants spend less time to evaluate the explanation.

$$t_{norm} = \frac{t - min(t)}{max(t) - min(t)} \tag{1}$$

Table 12: Kruskal-Wallis tests on the explanations in three explanation types. The lower the mean rank, the higher the efficiency of the explanations.

Explanations	*p(2-tailed)*	*Mean rank*					
		ADD	*APP*	*LMS*	*MAJ*	*MPL*	*FAI*
Type 1	.188	102.03	109.97	128.96	129.70	112.09	-
Type 2	.000	152.30	119.78	197.47	125.36	**106.54**	182.65
Type 3	.000	149.07	139.05	**99.84**	147.89	145.51	179.49

Table 13: Mann-Whitney U tests for the explanations between *Type 2* and *Type 3*. The lower the mean rank, the higher the efficiency of the explanations.

Explanations	*p(2-tailed)*	*Mean rank*	
		Type 2	*Type 3*
ADD	0.087	42.27	51.84
FAI	0.243	61.13	68.82
APP	**0.003**	37.91	54.64
LMS	0.073	56.83	35.71
MAJ	**0.001**	38.12	56.08
MPL	**0.000**	35.55	56.88

Results:

In the explanation level: The Kruskal-Wallis tests reveal an insignificant time difference across the explanations of *Type 1*, whereas significant time differences were detected across the explanations of *Type 2* and *Type 3* (see *p values* in Table 12). In *Type 2*, the *MPL-based* explanation received the lowest amount of time, whereas in *Type 3*, the *LMS-based* explanation worked the most efficiently. Besides, in both types, the participants spent quite long time on evaluating the *FAI-based* explanations (see Table 12).

In the explanation type level: The Mann-Whitney U tests show that compared to the explanations of *Type 3*, the participants spent shorter time on evaluating the explanations of *Type 2*, especially for *APP-based, MAJ-based,* and *MPL-based* explanations (see Table 13).

Discussion:

In the explanation level, the results confirm that the participants spent different amounts of time on evaluating the explanations from different explanation types. Additionally, considering group members' satisfaction from previous/future decisions triggered these differences. For instance, for the *MPL-based* explanation, taking into account the satisfaction of group members from the past decisions helped to speed up the participants' perception process. Similarly, the inclusion of future decision plans in the *LMS-based* explanation helped the participants to shorten the perception process. Besides, we found out that the *FAI-based* explanation always triggered a lengthy perception process. The reason could come from the nature of the group recommendation strategy. The *FAI strategy* considers the preferences of group members in turn. In a specific decision, the preference of *only one* group member is considered, and the preferences of others will be considered in future decisions. As a result, the group recommendation generated by this strategy does not satisfy all group members. Therefore, when evaluating the *FAI-based* explanations, the participants seemed to spend more time on considering the fairness and consensus aspects of the recommended items. Consequently, these explanations required a longer perception process compared to other explanations.

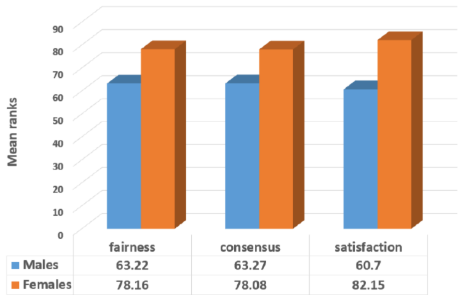
In the explanation type level, *APP-based, MAJ-based,* and *MPL-based* explanations showed significant differences in the spending time between *Type 2* and *Type 3*. In *Type 2*, these explanations include the satisfaction of group members from previous decisions, which can convince the less-favored group members in the on-going decision to agree on the recommended items. In other words, these explanations helped the participants more easily evaluate the fairness and the consensus aspects of the recommended item. The result is compatible with the finding of hypothesis H_3 that the *MPL-based explanation* of *Type 2* helps to increase the fairness perception of users concerning group recommendations. This gives an additional clarification why the *MPL-based explanation* can accelerate the participants' perception process and therefore, can speed up the group decision-making process.

Main results: Hypothesis H_4 can be confirmed for the *MPL-based* explanation of *Type 2* and the *LMS-based* explanation of *Type 3*. These explanations work more efficiently than others. Besides, the inclusion of the satisfaction of group members from previous decisions in the explanations can speed up the group decision-making process.

Table 14: Mann-Whitney U tests ($\alpha = .05$) for the explanations between *males* and *females*.

Explanations		*p (2-tailed)*		
		fairness	*consensus*	*satisfaction*
Type 1	*consensus-based*	0.605	0.589	0.636
	majority-based	0.982	0.532	0.351
	borderline	0.524	0.435	0.789
Type 2	*consensus-based*	0.422	0.871	0.795
	majority-based	0.808	**0.028**	0.145
	borderline	0.965	0.864	0.162
Type 3	*consensus-based*	0.798	0.742	0.855
	majority-based	0.217	0.163	**0.022**
	borderline	**0.028**	**0.03**	**0.03**

Figure 1: Mean ranks generated by the Mann-Whitney U tests on the *borderline* explanations of *Type 3* between males and females.



5.5 Hypothesis H_5

Method: For this hypothesis, we performed two tests: for *gender* and for *age range*. Regarding the gender, we grouped the participants into two groups: *males* and *females*. For the age range, we categorized the participants into two groups: [18..25] and [26..51] since the participants of our user study are *students* (from 18 to 25 years old) and *employees* (from 26 to 51 years old).

In each explanation type, we first collected the participants' evaluations for the explanations on three mentioned dimensions according to the gender and age range. These evaluations were grouped according to the explanation category: *consensus-based*, *majority-based*, and *borderline*. For the gender test, in each category, we ran *three* Mann-Whitney U tests on three dimensions between males and females. Similarly, for the age-range test, we also ran three Man-Whitney U tests on three dimensions between [18..25] and [26..51] age ranges.

Results:

Regarding gender: The Mann-Whitney U tests show that only *borderline* explanations of *Type 3* was highly sensitive to gender diversity, whereas the remaining explanations did not show any significant differences (see Table 14). In the *borderline* explanations, females were likely to give higher evaluations for the mentioned dimensions compared to males (see Figure 1).

Regarding age range: The Mann-Whitney U tests show that for most explanations, there were not any significant differences in the *fairness, consensus,* and *satisfaction* levels between [18..25] and [26..51] age ranges. This leads to the conclusion that age-range diversity did not influence the participants' perception concerning fairness, consensus, and satisfaction with regard to group recommendations.

Table 15: Mann-Whitney U tests ($\alpha = .05$) for the explanations between *students* (from 18 to 25 years old) and *employees* (from 26 to 51 years old).

Explanations		*p (2-tailed)*		
		fairness	*consensus*	*satisfaction*
Type 1	*consensus-based*	0.876	0.729	0.82
	majority-based	0.674	0.878	0.328
	borderline	0.886	0.56	0.238
Type 2	*consensus-based*	0.145	0.594	0.818
	majority-based	0.444	0.575	0.152
	borderline	0.345	0.727	**0.039**
Type 3	*consensus-based*	0.522	0.563	0.117
	majority-based	0.057	0.562	0.827
	borderline	0.895	0.793	0.227

Discussion: The results mentioned above show that males and females adapted their perception levels according to the explanation type. For the explanations of *Type 1* and *Type 2*, they gave similar *fairness, consensus,* and *satisfaction* levels. Whereas, for the explanations of *Type 3*, males and females behaved differently. For the *borderline* explanations of *Type 3*, females gave higher levels of the mentioned dimensions compared to males. These differences could be explained by the nature of the aggregation strategies conveyed in the explanations. Borderline strategies consider the preferences of a subset of group members, which could trigger further discussions in the group. In such scenarios, males and females show different ways to achieve consent in the group. Females were more likely to employ *cooperative strategies*, whereas males took *winner-and-loser* approaches [Kelly et al., 1991, Rosenthal, 2000]. This way, females could reach a consensus faster than males. These behaviors explain why females gave higher consensus levels for these explanations. Moreover, females were proven to be more sympathized than males [Hannagan and Larimer, 2010]. Therefore, they can accept a solution which was not the favorite of every group member. Females found this solution quite fair since the preferences of less-favored users at least will not be ignored and they will be considered soon in the near future decisions. This could be the reason why females gave higher fairness levels compared to males. Moreover, higher fairness and consensus levels lead to higher satisfaction levels of users concerning group recommendations (this can be confirmed in hypothesis H_2).

Main results: Hypothesis H_5 can be confirmed for gender diversity. Explanations including *future decision plans* trigger significant differences in fairness, consensus, and satisfaction levels between males and females. The sensitivity of gender diversity is clearly shown when the preferences of only a *subset* of group members are considered in the group recommendation generation process.

6 Possibilities to Develop Explanations of Group Recommendations for Configurable Items

The explanation approach mentioned in the previous sections were analyzed in the context of *non-configurable* or *no-attribute* items. However, there exist some possibilities to make this approach applicable to the explanations of group recommendations for *configurable* (or *multi-attribute*) items.

Configurable items are composed from a set of predefined attributes. For instance, a tourism package can be considered as a configurable item which consists of many attributes: *destination* (i.e., *"where to go"*), *hotel* (i.e., *"where to stay"*), *food* (i.e., *"what to eat"*), *attraction* (i.e., *"what to do"*), *cost* (i.e., *"how much to pay"*), and *transportation* (i.e., *"how to get around"*). Making decisions on the configuration items is referred to as a *group-based configuration task* which can be represented based on a *Constraint Satisfaction Problem* (CSP) [Tsang, 1993].

Group-based configuration task. A group-based configuration task can be defined as a CSP (V,D,C) where V is a set of variables, D represents the corresponding domain definitions, and $C = PREF \cup CKB$ represents a set of constraints. In this context, $PREF = \bigcup PREF_i$ is the union of group members' preferences $PREF_i$ and CKB represents a configuration knowledge base [Felfernig et al., 2016].

We exemplify a group-based configuration task in the *tourism domain* with components defined as in the following:

- $V = \{$*hotel#stars, food, dest, att, cost, trans*$\}$
- $D = \{$*dom(hotel#stars)* = [1..5], *dom(food)* = [*African, Asian, European, American, Australian*], *dom(dest)* = [*Africa, Asia, America, Europe*],

dom(att) = [*monument, museum, palace, beach, mountain, river*], *dom(cost)* = [0..5000], *dom(trans)* = [*bus, airplane, train*]}

- *CKB* = {c_1: *dest* = *Asia* ⇒ *food* = *Asian*, c_2: (*trans* = *airplane*) ∧ (*hotel#stars* ≥ *3*) ⇒ *cost* ≥ *1500*, c_3: *dest* = *Asia* ⇒ *trans* = *airplane*, c_4: *cost* ≤ *500* ⇒ *(dest* = *Europe)* ∧ *(trans* = *bus)*, c_5: *dest* = *America* ⇒ *cost* ≥ *2000*}
- *user* u_1: $PREF_1$ = {*trans* = *airplane*, *1500* ≤ *cost* ≤ *2500*, *hotel#stars* = *3*, *att* = *beach*},
- *user* u_2: $PREF_2$ = {*food* = *Asian*, *1000* ≤ *cost* ≤ *1500*, *att* = *beach*},
- *user* u_3: $PREF_3$ = {*dest* = *Asia*, *1000* ≤ *cost* ≤ *1500*, *hotel#stars* = *3*}

Explaining a configuration (solution) of group-based configuration task. *Constraint-based recommender systems* usually make recommendations for configurable items that are built upon deep knowledge about items and their corresponding recommendation rules (constraints). This information serves as a basis for explaining item recommendations by analyzing reasoning steps that led to the derivation of solutions [Friedrich and Zanker, 2011]. Such explanations follow the tradition of *AI-based expert systems* [Buchanan and Shortliffe, 1985, Friedrich, 2004]. On the one hand, explanations are used to answer *how*-questions, i.e., questions related to the reasons behind a recommendation. How questions are answered in terms of showing the relationship between the defined user preferences $PREF_i$ and the recommended items. For instance, on the basis of the presented example about group-based configuration task, a corresponding solution is specified by a configuration solver: X = {*trans* = *airplane*, *food* = *Asian*, *dest* = *Asia*, *att* = *beach*, *cost* = *1500*, *hotel#stars* = *3*}. An example explanation of such a solution is *"X is recommended to the group since you like Asian beaches, and the upper-cost limit is not higher than \$1500"*.

Besides, constraint-based recommender systems are able to answer *why* and *why not* questions. Explanations for the first type are used to provide users with insights into why specific questions have to be answered, whereas explanations for *why not* questions help users escape from the *"no solution could be found"* dilemma [Felfernig et al., 2009]. These explanations help to increase use trust in the recommender applications. Moreover, explanations related to *why not* questions can increase the user perception of item domain knowledge [Felfernig et al., 2007].

Consider fairness and consensus aspects in group decisions. Situations can occur where the preferences of group members are conflict with each other or with the knowledge base [Felfernig et al., 2012, Felfernig et al., 2016, Mahyar et al., 2017]. In group recommendation scenarios, a *consensus* is defined

Table 16: An example about the impact of different diagnoses on the preferences of users. For example, if the minimal diagnosis set (r_{11}, r_{22}) is chosen then $user_1$ has to adjust one of all his/her preferences.

Users	(r_{11}, r_{22})	(r_{23}, r_{22})	(r_{23}, r_{31})
$user_1$	1	0	0
$user_2$	1	2	1
$user_3$	0	0	1
Least Misery	**1**	**2**	**1**

in terms of *inconsistencies/disagreement* between individual group members regarding item evaluations (ratings) [Amer-Yahia et al., 2009]. In order to provide a basis for establishing consensus, such situations have to be explained and visualized [Jameson, 2004, Mahyar et al., 2017]. In this context, *diagnosis methods* can help to recover the consistency [Felfernig et al., 2016]. This method helps to identify corresponding minimal diagnoses which have to be deleted/adapted from group members' preferences so that a solution can be found. These repair actions propose changes to the current set of preferences such that a group recommendation can be identified. Such repairs can take into account the individual preferences of group members [Felfernig et al., 2016]. In this context, *social choice-based preference aggregation strategies* presented in *Section 3.1* can be applied to figure out diagnoses *acceptable* for the whole group. For instance, the *Least Misery* strategy enables to choose a diagnosis in which the number of adaptations of group members is lowest. This strategy helps to foster the fairness among group members [Atas et al., 2019].

Let's have a look at again the mentioned example and assume group members have specified their preferences as follows:
$PREF_1$ = {r_{11}: *trans = airplane*, r_{12}: *1000 ≤ cost ≤ 1500*, r_{13}: *hotel#stars = 4*}
$PREF_2$ = {r_{21}: *food = Asian*, r_{22}: *dest = Asia*, r_{23}: *trans = train*}
$PREF_3$ = {r_{31}: *dest = Australia*, r_{32}: *hotel#stars* = 4, r_{33}: *att = museum*}

In this example, three minimal conflict sets are determined as follows: (r_{11}, r_{23}), (r_{22}, r_{31}) and (r_{22}, r_{23}). Corresponding minimal diagnoses extracted from these minimal conflict sets are: (r_{11}, r_{22}), (r_{23}, r_{22}), and (r_{23}, r_{31}). *Table 16* depicts the influence of different diagnoses on the current preferences of users. In this example, we use the *Least Misery* strategy and show how this strategy affects the selection of diagnoses in the group scenario. This strategy prefers alternatives minimizing the misery of individual group members (i.e., less adaptations is better). Using this ranking criterion, a list of diagnoses shown in the explanation will be in the following order: *1*-(r_{11}, r_{22}), *2*-(r_{23}, r_{31}) and *3*-(r_{23}, r_{22}).

Another approach to take into account consensus in group decision making is to indicate possible changes in the preferences of users that help to restore consistency. In group-based settings, such *repair-related explanations* help group members understand the constraints of other group members and decide in which way their preferences should be adapted. An example of repair-related explanations in the tourism domain could be *"there is no solution that could be found according to your specified preferences. Maybe consider to increase the cost a little bit or to change means of transportation from airplane to train or bus"*.

In situations where some group members have to adapt their preferences much more than others, information regarding *the satisfaction of group members from previous decisions* might be helpful to increase their acceptance of the adaptations. An example explanation could be as follows: *"Alex does not have to adapt his preferences many times since he was treated less favorably in the last two decisions and he should have a higher priority in this decision"*.

On the other hand, in the context of configurable items, it is quite challenging to find a solution on which the preferences of group members for *"all"* attributes are fulfilled. In case the preferences of group members for some specific attributes are not satisfied, their preferences for other attributes should be immediately considered, especially for group decisions in the non-repeated item domains such as *apartments/houses, digital-camera,* and *cars.* An example explanation of a recommended item in the apartment domain could be formulated as follows: *"Apartment X is recommended to the group since it is in the city center and the renting cost is lower than your preferred price. Although users u_2 and u_4 have to accept minor drawbacks (no balcony or big kitchen), there is a fitness room in the ground floor that meets the requirements of u_4, and it is just some steps far from the working place of u_2"*.

7 Conclusion and Future Work

7.1 Conclusion

In this paper, we have proposed a novel approach to integrating explanations into group recommender systems for the sake of fostering the social aspects within groups, such as fairness perception, consensus perception, and satisfaction with regard to recommended items. In particular, explanations describing recommendation strategies considering the preferences of *all* or *a majority* of group members should be implemented and sent to users to increase the mentioned aspects of users. For decisions periodically repeated by the same group, information regarding *group members' satisfaction* from previous decisions should be included in the recommendations to increase users' fairness and consensus perception, as well as speed up the group decision-making process.

Regarding the influence of group composition on the group decision-making process, we have shown that explanations with future decision plans are highly sensitive to gender diversity. Males and females show different levels of fairness and consensus perceptions, especially in decisions that only take into account the preferences of a subset of group members in the group recommendation-generation process.

On the other hand, we have found some possibilities to make our approach applicable to the explanations of configurable products/services. The mentioned approach can be adapted to explain *why* or *why-not* a solution (configuration) of a group-based configuration task *can* or *cannot* be generated. Moreover, this approach can also be applied to consider fairness and consensus aspects in group decisions in which there exist conflicts between group members' preferences. The proposed explanations can help groups achieve resolving solutions that avoid misery among group members.

7.2 Future Work

One limitation of our study lies in the distribution of explanations to the participants. At any given time, each participant could observe and evaluate *only one* explanation, and the evaluation for one explanation was independent of the evaluations for other explanations. However, since the break time between two different evaluations was not long enough, this could trigger potential biases in the evaluation process. As a result, the participants' evaluations for the explanations were not wholly independent of each other. Therefore, within the scope of future work, we will run our data with *multi-level models* to achieve more precise results. Besides, we will further investigate the social aspects of the explanations based on the following features:

Item domains: We will investigate social aspects in the context of different decision domains. An example explanation in such a context could be *"the preferences of user X were not taken into account in the last two decisions - restaurants and movies. Therefore, he should have higher priorities in future decisions concerning restaurants and movies"*. Besides, the social aspects of group decision making could be further considered based on the item domain. For instance, for high-involvement item domains (i.e., domains which require high decision efforts), the social aspects should be strictly taken into account in the on-going decision. An explanation in this item domain could be *"although the apartment X is not the favorite option of all members, no group member has a real problem with it. Therefore, this apartment seems to be the most appropriate solution for all group members to stay together in the next two years"*. This explanation describes the *Least Misery* strategy, which suggests a fair solution to all group members and avoids the misery within the group.

Group dynamics: We will integrate some features of group dynamics (e.g., group members' *personalities* and *individual situations* into the explanations to further investigate their impacts on the social aspects in groups. For instance, regarding group members' personalities, an explanation could be shaped as follows: *"A group member with a strong personality who was treated less favorably last time should be immediately compensated in the next decision"*. Regarding the individual situations of group members, an example explanation could be *"restaurant A has been chosen to the group since user X is a vegan and only this restaurant serves additional dishes for vegan"*.

References

[Amer-Yahia et al., 2009] Amer-Yahia, S., Roy, S. B., Chawlat, A., Das, G., and Yu, C.: "Group recommendation: Semantics and efficiency"; Proc. VLDB Endow. (2009), 2(1):754–765.

[Ardissono et al., 2003] Ardissono, L., Goy, A., Petrone, G., Segnan, M., and Torasso, P.: "INTRIGUE: Personalized recommendation of tourist attractions for desktop and hand held devices"; Applied Artificial Intelligence (2003), 17(8):687–714.

[Atas et al., 2019] Atas, M., Tran, T. N. T., Felfernig, A., Erdeniz, S. P., Samer, R., and Stettinger, M.: "Towards similarity-aware constraint-based recommendation"; In Wotawa, F., Friedrich, G., Pill, I., Koitz-Hristov, R., and Ali, M., editors, Advances and Trends in Artificial Intelligence. From Theory to Practice, Cham. Springer International Publishing (2019), 287–299.

[Bilgic, 2005] Bilgic, M.: "Explaining recommendations: Satisfaction vs. promotion. In Proceedings of Beyond Personalization 2005"; the Workshop on the Next Stage of Recommender Systems Research (IUI'05) (2005), 13–18.

[Buchanan and Shortliffe, 1985] Buchanan, B. and Shortliffe, E.: "Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project"; Addison-Wesley (1985).

[Castro et al., 2015] Castro, J., Quesada, F. J., Palomares, I., and Martínez, L.: "A consensus-driven group recommender system"; Int. J. Intell. Syst. (2015), 30(8):887–906.

[Chen et al., 2013] Chen, L., de Gemmis, M., Felfernig, A., Lops, P., Ricci, F., and Semeraro, G.: "Human decision making and recommender systems"; ACM Trans. Interact. Intell. Syst. (2013), 3(3):1–7.

[Chen and Pu, 2012] Chen, L. and Pu, P.: "Critiquing-based recommenders: Survey and emerging trends"; User Modeling and User-Adapted Interaction (2012), 22(1):125–150.

[Chiclana et al., 2007] Chiclana, F., Mata, F., Alonso, S., Herrera-Viedma, E., and Martínez, L.: "Group decision making: From consistency to consensus"; In Torra, V., Narukawa, Y., and Yoshida, Y., editors, Modeling Decisions for Artificial Intelligence, Berlin, Heidelberg. Springer Berlin Heidelberg (2007), 80–91.

[Cohen, 1988] Cohen, J.: "Statistical Power Analysis for The Behavioral Sciences"; Routledge (1988).

[Felfernig et al., 2007] Felfernig, A., Gula, B., and Teppan, E.: "Knowledgebased recommender technologies for marketing and sales"; International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI) (2007), 21(2):333–354.

[Felfernig et al., 2009] Felfernig, A., Friedrich, G., Schubert, M., Mandl, M., Mairitsch, M., and Teppan, E.: "Plausible repairs for inconsistent requirements" (2009), 791–796.

[Felfernig et al., 2012] Felfernig, A., Schubert, M., and Zehentner, C.: "An efficient diagnosis algorithm for inconsistent constraint sets"; AI EDAM (2012), 26(1):53–62.

[Felfernig et al., 2016] Felfernig, A., Atas, M., and Stettinger, M.: "Towards group-based configuration"; In International Workshop on Configuration (2016), 69–72.

[Felfernig et al., 2017] Felfernig, A., Atas, M., Tran, T. T., Stettinger, M., Erdeniz, S. P., and Leitner, G.: "An analysis of group recommendation heuristics for high- and low-involvement items"; In IEA/AIE, volume 10350 of Lecture Notes in Computer Science, Springer (2017), 335–344.

[Felfernig et al., 2018a] Felfernig, A., Atas, M., Helic, D., Tran, T. T., Stettinger, M., and Samer, R.: "Group Recommender Systems: An Introduction, chapter Algorithms for Group Recommendation"; Springer (2018), 27–58.

[Felfernig et al., 2018b] Felfernig, A., Tintarev, N., Tran, T. T., and Stettinger, M.: "Group Recommender Systems: An Introduction, chapter Explanations for Groups"; Springer (2018), 105–126.

[Friedrich, 2004] Friedrich, G.: "Elimination of spurious explanations"; In ECAI (2004), 813–817.

[Friedrich and Zanker, 2011] Friedrich, G. and Zanker, M.: "A taxonomy for generating explanations in recommender systems"; AI Magazine (2011), 90–98.

[Gedikli et al., 2014] Gedikli, F., Jannach, D., and Ge, M.: "How should i explain? a comparison of different explanation types for recommender systems"; Int. J.Hum.-Comput. Stud. (2014), 72(4):367–382.

[Hannagan and Larimer, 2010] Hannagan, R. J. and Larimer, C. W.: "Does gender composition affect group decision outcomes? Evidence from a laboratory experiment"; Political Behavior (2010), 32(1):51–67.

[Hastie and Kameda, 2005] Hastie, R. and Kameda, T.: "The robust beauty of majority rules in group decisions"; American Psychological Association (2005), 112(2):494–508.

[Herlocker et al., 2000] Herlocker, J. L., Konstan, J. A., and Riedl, J.: "Explaining collaborative filtering recommendations"; In Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work - CSCW, New York, NY, USA. ACM. (2000), 241–250.

[Hertzberg et al., 2013] Hertzberg, M., Smith, R., and Westphal, R.: "A Consensus Handbook: Co-operative decision-making for activists, co-ops and communities. Seeds for Change Lancaster Co-operative"; Ltd - 1st edition (2013).

[Jameson, 2004] Jameson, A.: "More than the sum of its members: Challenges for group recommender systems"; In Proceedings of the Working Conference on Advanced Visual Interfaces, AVI, New York, NY, USA. ACM. (2004), 48–54.

[Jannach et al., 2010] Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G.: "Recommender Systems: An Introduction"; Cambridge University Press, New York, NY, USA, 1st edition (2010).

[Kacprzyk and Zadrozny, 2016] Kacprzyk, J. and Zadrozny, S.: "Towards a Fairness-Oriented Approach to Consensus Reaching Support Under Fuzzy Preferences and a Fuzzy Majority via Linguistic Summaries"; Springer Berlin Heidelberg, Berlin, Heidelberg (2016), 189–211.

[Kapcak et al., 2018] Kapcak, O., Spagnoli, S., Robbemond, V., Vadali, S., Najafian, S., and Tintarev, N.: "Tourexplain: A crowdsourcing pipeline for generating explanations for groups of tourists"; In ACM RecSys Workshop on Recommenders in Tourism (2018).

[Kelly et al., 1991] Kelly, R. M., Saint-Germain, M. A., and Horn, J. D.: "Female public officials: A different voice"; The Annals of the American Academy of Political and Social Science (1991), 515:77–87.

[Mahyar et al., 2017] Mahyar, N., Liu, W., Xiao, S., Browne, J., Yang, M., and Dow, S. P.: "Consesnsus: Visualizing points of disagreement for multi-criteria collaborative decision making"; In Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW Companion, New York, NY, USA. ACM (2017), 17–20.

[Masthoff, 2004] Masthoff, J.: "Group modeling: Selecting a sequence of television items to suit a group of viewers"; User Modelling and User-Adapted Interaction (2004), 14(1):37–85.

[Masthoff, 2011] Masthoff, J.: "Group recommender systems: Combining individual models"; In Ricci, F., Rokach, L., Shapira, B., and Kantor, P. editors, Recommender Systems Handbook, Springer (2011), 677–702.

[McCarthy et al., 2004] McCarthy, K., Reilly, J., McGinty, L., and Smyth, B.: "On the dynamic generation of compound critiques in conversational recommender systems"; In De Bra, P. M. E. and Nejdl, W., editors, Adaptive Hypermedia and Adaptive Web-Based Systems, Berlin, Heidelberg. Springer (2004), 176–184.

[Najafian and Tintarev, 2018] Najafian, S. and Tintarev, N.: "Generating consensus explanations for group recommendations: An exploratory study"; In Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization, UMAP '18, New York, NY, USA. ACM. (2018), 245–250.

[Ntoutsi et al., 2012] Ntoutsi, E., Stefanidis, K., Nørvåg, K., and Kriegel, H.-P.: "Fast group recommendations by applying user clustering"; In Proceedings of the 31st International Conference on Conceptual Modeling, ER'12, Berlin, Heidelberg. Springer-Verlag (2012), 126–140.

[Pallant, 2007] Pallant, J.: "SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows Version 15"; Open University Press, Milton Keynes, UK, USA, 3rd edition (2017).

[Palomares et al., 2014] Palomares, I., Martínez-López, L., and Herrera, F.: "A consensus model to detect and manage non-cooperative behaviors in large-scale group decision making"; IEEE Trans. Fuzzy Systems (2014), 22(3):516–530.

[Pérez et al., 2018] Pérez, I., Cabrerizo, F., Alonso, S., Dong, Y., Chiclana, F., and Herrera-Viedma, E.: "On dynamic consensus processes in group decision making problems"; Information Sciences (2018), 459:20–35.

[Quijano-Sanchez et al., 2013] Quijano-Sanchez, L., Recio-Garcia, J. A., Diaz-Agudo, B., and Jimenez-Diaz, G.: "Social factors in group recommender systems"; ACM Trans. Intell. Syst. Technol. (2013), 4(1):8:1–8:30.

[Rosenthal, 2000] Rosenthal, C. S.: "Gender styles in state legislative committees: Raising their voices in resolving conflict"; Women and Politics (2000), 21:21–45.

[Sanz de Acedo Lizárrage et al., 2007] Sanz de Acedo Lizárrage, M. L., Sanz de Acedo Baquedano, M. T., and Cardelle-Elawar, M.: "Factors that affect decision making: Gender and age differences"; International Journal of Psychology Psychological Therapy (2007), 7(3):381–391.

[Senot et al., 2010] Senot, C., Kostadinov, D., Bouzid, M., Picault, J., Aghasaryan, A., and Bernier, C.: "Analysis of strategies for building group profiles"; In De Bra, P., Kobsa, A., and Chin, D., editors, User Modeling, Adaptation, and Personalization, Berlin, Heidelberg. Springer Berlin Heidelberg (2010), 40–51.

[Stettinger, 2014] Stettinger, M.: "Choicla: Towards domain-independent decision support for groups of users"; In Proceedings of the 8th ACM Conference on Recommender Systems, RecSys '14, New York, NY, USA. ACM (2014), 425–428.

[Tanner, 2018] Tanner, R.: "Equity theory - why employee perceptions about fairness do matter" (2018).

[Tintarev and Masthoff, 2007] Tintarev, N. and Masthoff, J.: "A survey of explanations in recommender systems"; In Proceedings of the 2007 IEEE 23rd International Conference on Data EngineeringWorkshop, ICDEW '07, Washington DC, USA. IEEE Computer Society (2007), 801–810.

[Tintarev and Masthoff, 2010] Tintarev, N. and Masthoff, J.: "Designing and Evaluating Explanations for Recommender Systems"; chapter 15, Springer US (2010), 479–510.

[Tintarev and Masthoff, 2012] Tintarev, N. and Masthoff, J.: "Evaluating the effectiveness of explanations for recommender systems"; User Modeling and User-Adapted Interaction (2012), 22(4-5):399–439.

[Tsang, 1993] Tsang, E. P. K.: "Foundations of Constraint Satisfaction"; Computation in Cognitive Science, Academic Press (1993).