User Interfaces for Counteracting Decision Manipulation in Group Recommender Systems

Thi Ngoc Trang Tran
Graz University of Technology
Graz, Austria
ttrang@ist.tugraz.at

Alexander Felfernig
Graz University of Technology
Graz, Austria
alexander.felfernig@ist.tugraz.at

Viet Man Le
Hue University
Hue, Vietnam
lvman@hce.edu.vn

Müslüm Atas
Graz University of Technology
Graz, Austria
muesluem.atas@ist.tugraz.at

Martin Stettinger
Graz University of Technology
Graz, Austria
martin.stettinger@ist.tugraz.at

Ralph Samer
Graz University of Technology
Graz, Austria
rsamer@ist.tugraz.at

ABSTRACT

In group recommender systems, *decision manipulation* refers to an attack in which a group member makes attempts to push his/her favorite options. In this paper, we propose user interfaces to counteract decision manipulation in group recommender systems. The proposed user interfaces visualize information dimensions regarding rating adaptations of group members at different transparency levels. The results show that the user interface at the highest transparency level best helps to discourage users from decision manipulation. Besides, the ability of the user interfaces to counteract decision manipulation differs depending on the dimensions represented in the user interfaces. The information dimensions regarding "*item ratings*" and "*group recommendations*" have the strongest impacts on preventing users from decision manipulation.

CCS CONCEPTS

• **Information systems → Information systems applications**; **Recommender systems**; • **Human-centered computing → User studies**; **User interface design**.

KEYWORDS

Group Recommender Systems; Intelligent User Interfaces; Decision Manipulation; Decision Manipulation Counteraction

ACM Reference Format:

Thi Ngoc Trang Tran, Alexander Felfernig, Viet Man Le, Müslüm Atas, Martin Stettinger, and Ralph Samer. 2019. User Interfaces for Counteracting Decision Manipulation in Group Recommender Systems. In *27th Conference on User Modeling, Adaptation and Personalization Adjunct (UMAP'19 Adjunct), June 9–12, 2019, Larnaca, Cyprus.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3314183.3324977

UMAP'19 Adjunct, June 9–12, 2019, Larnaca, Cyprus

ACM ISBN 978-1-4503-6711-0/19/06…$15.00
https://doi.org/10.1145/3314183.3324977

1 INTRODUCTION

Recommender systems are useful tools that guide users to items they might love based on their own and other users' preferences [18]. However, recommender systems supporting explicit user feedback [4] might face biases triggered by "*insincere*" user preferences. In this scenario, users try to change their feedback to alter the recommendation to a preferred one [8, 12]. Such *decision manipulations* deteriorate the quality of recommendations [2]. In *recommender systems* for groups of users, manipulation issues can be triggered when user control mechanisms are implemented. These user controls can be: *(i)* to allow users to articulate their preferences for items, *(ii)* to enable users to see others' preferences, or *(iii)* to adapt their preferences for achieving a consensus solution among group members [17]. These mechanisms facilitate group members' rating adaptations which aim to push their favorite options.

Up to now, to some extent it is still unclear how to counteract decision manipulation in group recommender systems. Although there are some studies on manipulation resistance (e.g., [6, 14, 20]), most proposed solutions are for single-user recommender systems. To the best of our knowledge, in the current literature, there exist only a few research contributions with an in-depth analysis of decision manipulation issues [9, 10, 13]. In this paper, we propose user interfaces (UIs) for counteracting decision manipulation in group recommender systems. Our idea is to disclose the "*rating adaptation history*" of group members, which means *item rating changes on group decision tasks are saved and shown to all group members*. We assume that making group members' rating adaptations *transparent* can help to counteract decision manipulation.

The contribution of this paper is threefold: *First*, we propose UIs which visualize the rating adaptation history of group members at different transparency levels. *Second*, we identify UIs which are the most understandable and most effective for counteracting decision manipulation in group recommender systems. Third, we discuss information dimensions of rating adaptations which strongly help to prevent users from decision manipulation.

The remainder of the paper is organized as follows. In *Section 2*, we summarize the related work of decision manipulation issues in group recommender systems. In *Section 3*, we present different dimensions describing the rating adaptation history of group members and propose UIs on the basis of combining these dimensions. In *Section 4*, we define research questions and describe main steps done in our user study. Data analysis and results regarding the research questions are presented in *Section 5*. In *Section 6*, we conclude the paper and discuss open issues for future work.

2 RELATED WORK

Decision manipulation has been experienced in one of the earliest group recommender systems, so-called MUSICFX, which automatically selects music genres to play in a fitness center [15]. In this system, some users were observed to intentionally indicate that they disliked the being-played music genre in order to force an immediate change of music genre. Another manipulation behavior was found in the TRAVEL DECISION FORUM [9] where group members can see the preferences of each other. This triggers rating adaptations of some group members for pushing their favorite options. To avoid this, group members' preferences should not be shown. However, it seems to be a sub-optimal solution since users might be able to guess others' preferences [11]. Moreover, in some cases, group members' preferences should be shown, e.g., for consensus making purpose. An alternative approach for counteracting decision manipulation is to have a group recommender system applying a preference aggregation strategy that is inherently *non-manipulatable*, such as "*median*" or "*random choice*" [9]. However, these strategies reveal the unacceptability of users with regard to group recommendations. In the line of *mechanism design* research, Conitzer and Sandholm [3] propose another approach that automatically generates aggregation functions so that desirable recommendations can be achieved for groups, even if group members rate items based on their self-interest. However, these approaches face the difficulty in providing understandable and adequate explanations of group recommendations. In this paper, our approach is to propose UIs *showing group members' rating adaptations in a group decision task* to the whole group. If a user as a manipulator knows that his/her rating adaptations can be seen by others, then he/she might not try to manipulate the decision. This idea originates from an observer effect, so-called the *Hawthorne effect*, in which *users modify an aspect of their behavior in response to their awareness of being observed* [19]. This effect indicates a psychological phenomenon in which users tend to do something positive or better if they are aware of being observed by others [1]. In the context of decision manipulation, this effect can be interpreted by the fact that users tend to avoid decision manipulation if they know that their rating adaptations are seen by others.

3 USER INTERFACES FOR COUNTERACTING DECISION MANIPULATION

3.1 Rating Adaptation Information

The rating adaptations of group members could be done for either "*positive*" purposes (e.g., making consensus within the group) or "*negative*" purposes (e.g., manipulating the decision). In this context, an important concern is that "*which information should be shown so that negative-purpose rating adaptations of group members can be detected*". We assume only showing the information regarding "*which of items whose ratings have been adapted by group members*" is insufficient to predict group members' decision manipulation behaviors. Additional information needs to be clarified, such as "*how the ratings of these items have been adapted*", "*when these ratings were adapted*", and "*how group recommendation has been changed after rating adaptations*". To address this, we propose the following dimensions:

- *Dimension 1* (*group member - GM*) indicating the *name* of a group member who has adapted the ratings of items;

- *Dimension 2* (*item - I*) indicating the item whose rating has been adapted;

- *Dimension 3* (*rating - R*) describing the rating of an item which can be *an original rating*, *an adapted rating*, or *a group rating* (i.e., the *rating* calculated by merging all individual group members' ratings using a preference aggregation strategy [5]);

- *Dimension 4* (*timeline - TL*) displaying group members' rating adaptations in a chronological order;

- *Dimension 5* (*tendency - TD*) describing the *direction* (increase (+) or decrease (-)) and the *magnitude* of a rating adaptation;

- *Dimension 6* (*group recommendation - GR*) revealing how the group recommendation has been changed according to group members' rating adaptations. The rating adaptation history shows the *group recommendation at the starting point* and *group recommendations generated after group members' rating adaptations*. It is more likely that manipulators' rating adaptations trigger changes with regard to group recommendations.

Table 1: The construction of the UIs of *Group 1* and *Group 2*. The highlighted UIs are the selected UIs for the user study.

group	trans. level	UI	*GM*	*I*	*R**	*TL*	*TD*	*GR*	visualization method
1	2	$UI1_{basis}$	✓	✓					*table*
	3	$UI1_{R}$	✓	✓	✓				*table*
		$UI1_{TL}$	✓	✓		✓			*table*
		$UI1_{TD}$	✓	✓			✓		*table*
		$UI1_{GR}$	✓	✓				✓	*table (Fig.1a)*
	4	$UI1_{R+TL}$	✓	✓	✓	✓			
		$UI1_{R+TD}$	✓	✓	✓		✓		*graph (Fig.1d)*
		$UI1_{R+GR}$	✓	✓	✓			✓	
		$UI1_{TL+TD}$	✓	✓		✓	✓		*graph*
		$UI1_{TL+GR}$	✓	✓		✓		✓	*table*
		$UI1_{TD+GR}$	✓	✓			✓	✓	*table*
	5	$UI1_{R+TL+TD}$	✓	✓	✓	✓	✓		*graph*
		$UI1_{R+TL+GR}$	✓	✓	✓	✓		✓	
		$UI1_{R+TD+GR}$	✓	✓	✓		✓	✓	*graph*
		$UI1_{TL+TD+GR}$	✓	✓		✓	✓	✓	*table (Fig.1b)*
	6	$UI1_{all}$	✓	✓	✓	✓	✓	✓	*graph (Fig.1e)*
2	3	$UI2_{R}$		✓	✓	✓			*table*
		$UI2_{TD}$		✓		✓	✓		*table*
		$UI2_{GR}$		✓		✓		✓	
	4	$UI2_{R+TD}$		✓	✓	✓	✓		
		$UI2_{R+GR}$		✓	✓	✓		✓	
		$UI2_{TD+GR}$		✓		✓	✓	✓	*table*
	5	$UI2_{all}$		✓	✓	✓	✓	✓	*table (Fig.1c)*

The ratings represented in each UI of Group 1 are **original ratings and **adapted ratings**. The ratings represented in each UI of Group 2 are **group ratings***

3.2 User Interfaces for Counteracting Decision Manipulation

The UIs for counteracting decision manipulation were generated and visualized in the following steps:

Step 1 - Generate the UIs: We generated two groups of UIs (*Group 1* and *Group 2*) by combining the dimensions presented in Section 3.1. These UIs represent the rating adaptation history of group members at different transparency levels. The *transparency level* of a UI corresponds to *the number of dimensions* included in the UI. In *Group 1*, each UI shows at least the information of "*who has adapted the ratings of which items*". By this, each UI at least consists

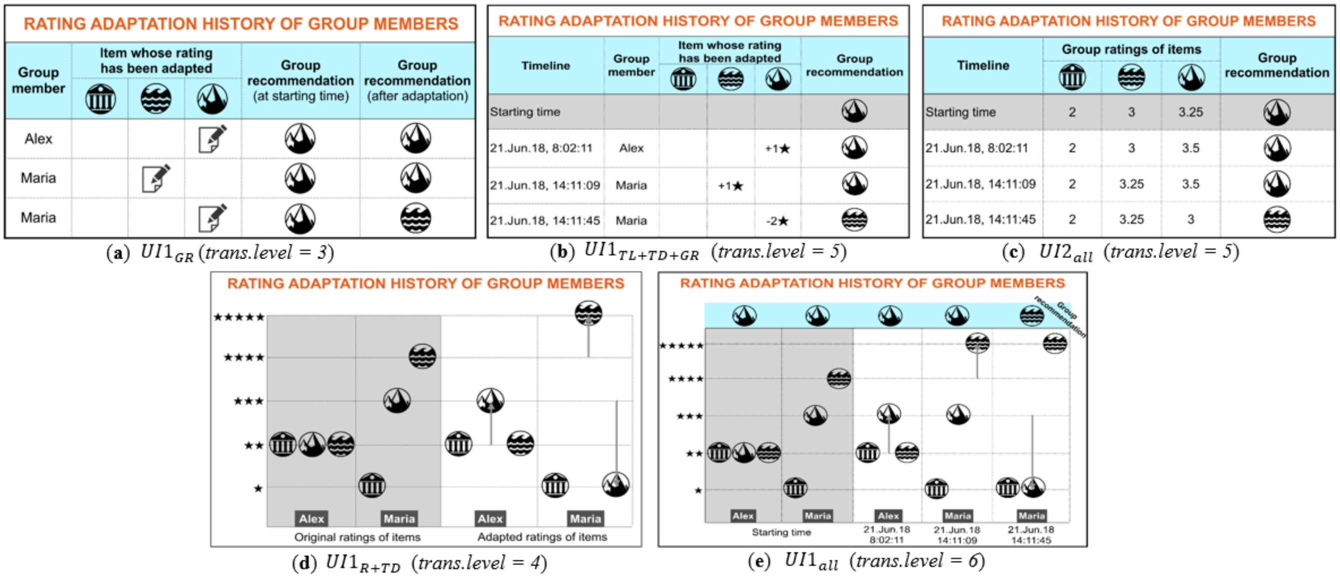

Figure 1: The visualizations of the UIs $UI1_{GR}$, $UI1_{R+TD}$, $UI1_{TL+TD+GR}$, $UI1_{all}$, and $UI2_{all}$.

of "*group member*" and "*item*" dimensions. A basic UI ($UI1_{basis}$) was tailored by these two dimensions with the transparency level of 2. In order to generate other UIs with higher transparency levels, we gradually added the remaining dimensions to the $UI1_{basis}$. The construction of the UIs of *Group 1* is described in Table 1.

The UIs of *Group 1* always show the "*group member*" dimension which could trigger *privacy* issues. Therefore, we additionally proposed *Group 2* whose UIs do not include this dimension. In this paper, we investigate which group of UIs better helps to counteract decision manipulation. The UIs of *Group 2* do not show the rating adaptations of group members. Instead, they reveal how the *group ratings* of items have been changed by the time. The UIs of *Group 2* are generated by gradually combining the "*item*" and "*timeline*" dimensions with one or more remaining dimensions. The construction of the UIs of *Group 2* is summarized in Table 1.

Step 2 - Select the UIs: In this step, we inspected the generated UIs and removed some UIs which are not understandable or represent similar information with other UIs. In *Group 1*, $UI1_{R+TL}$ and $UI1_{R+TL+TD}$ represent nearly the same information. In fact, compared to the $UI1_{R+TL+TD}$, although the $UI1_{R+TL}$ does not include the *"tendency"* dimension, this information can be figured out based on the *"rating"* dimension. In this case, we retained the $UI1_{R+TL+TD}$ since with the "*tendency*" dimension, this UI seems to be more understandable to users. For the same reasons, we removed $UI1_{R+GR}$, $UI1_{R+TL+GR}$, $UI2_{GR}$, $UI2_{R+TD}$, and $UI2_{R+GR}$. Consequently, we retained *17 UIs* (see the highlighted UIs in Table 1) and focused on evaluating and analyzing these UIs.

Step 3 - Visualize the UIs: Before visualizing the UIs, we defined a decision manipulation scenario in group recommender systems. In our study, this scenario was described in the context of group decisions where *group members know each other quite well* (e.g., friends or family members) and jointly decide on *a small given set of items* (e.g., 3-5 items). Since group members are familiar with each other, some group member's attempts to push his/her favorite options could result in negative impacts on the group decision, such as *fairness validation* and *cohesion damage* among group members. Therefore, avoiding decision manipulation in this context is crucial to conserve decision quality as well as group cohesion. Our decision manipulation scenario was described as follows: "Suppose that a group of four friends (Alex, Maria, Rosie, and Thomas) used a group recommender system to decide on a tourism destination type for the upcoming holiday. The group members explicitly rated three destination types (*museum, sea,* and *mountain*) using a 5-star rating scale (see Table 2). After articulating their preferences for destination types, each member can see others' preferences. The system recommends to the group a destination type with the maximum average of individuals' ratings (i.e., the "*mountain*" option). Alex does not like any destination type. However, he recognizes that most group members seem to be interested in "*mountain*". Therefore, he increases his rating for this option *from 2 to 3 stars* to fasten the consensus achieving process within the group. The rating adaptation of Alex does not change the group recommendation. Maria likes "*sea*" and wants this option will be chosen by the system. Therefore, she increases the rating of this option (*from 4 up to 5 stars*) and decreases the rating of the "*mountain*" (*from 3 down to 1 star*). The rating adaptations of Maria push the group recommendation to the "*sea*" option. The behavior of Maria in this context is a so-called *decision manipulation* and Maria is a *manipulator*."

Table 2: Ratings of group members for the destination types.

group member	museum	sea	mountain
Rosie	3	3	4
Alex	2	2	2
Maria	1	4	3
Thomas	2	3	4
group rating	2	3	**3.25**

Based on the proposed scenario, we visualized the UIs. We inspected existing visualization methods in recommender systems [7] and selected *tables* and *graphs* methods to visualize the proposed UIs. The reason was that, compared to other visualization methods, *tables* and *graphs* are quite appropriate to intuitively visualize different dimensions of rating adaptation history in a UI. After visualizing the UIs, we conducted a pilot user study with eight *experts* in our institute[1] who are working in the fields of Software Engineering and UI design to get feedback regarding the visualization methods and understandability of the UIs. Thereafter, we collected the experts' feedback and improved the UIs. The UIs' visualizations are briefly presented as follows:[2]

- *At the transparency level of 2,* the $UI1_{basis}$ is visualized using a *2-column table* on which the first column shows *group members' names*, the second one represents *items*, and each row indicates that *a group member has adapted the rating of an item*.

- *At the transparency level of 3*, all UIs are visualized using *multi-column tables*. Each table consists of columns for showing *group members' names* and *items*. The remaining columns represent additional dimensions. For instance, in $UI1_{GR}$ (see Figure 1a), the two last columns show *group recommendations*.

- *At the transparency level of 4*, the $UI1_{R+TD}$ (see Figure 1d) is visualized using a graph on which the Y axis shows *rating values* and the X axis shows *group members' names*. The space between these two axes shows *items* and *rating adaptations' tendencies*. These tendencies are visualized by arrows with different directions and lengths which indicate how much the ratings of items have been increased/decreased. The $UI1_{TL+TD}$ is also visualized using a graph where the Y axis shows *items*, the X axis shows a *timeline*, and the space between these axes shows the *tendencies* of rating adaptations. The remaining UIs are visualzed using *tables* whose structures are quite similar to the UIs at the transparency level of 3.

- *At the transparency level of 5*, the $UI1_{R+TL+TD}$ is visualized using the graph of the $UI1_{R+TD}$ (see Figure 1d) on which the X axis additionally represents a *timeline*. The $UI1_{R+TD+GR}$ is visualized by inserting into the graph of the $UI1_{R+TD}$ an additional axis that shows group recommendations. $UI1_{TL+TD+GR}$ and $UI2_{all}$ are visualized using *tables* as shown in Figure 1b and Figure 1c.

- *At the transparency level of 6*, the $UI1_{all}$ is visualized using a graph as shown in Figure 1e.

4 RESEARCH QUESTIONS AND USER STUDY

4.1 Research Questions

Our goal in this paper is to answer the following research questions:

- RQ_1: *Which transparency level of a rating adaptation history best helps to counteract decision manipulation?*

- RQ_2: *At a specific transparency level, which UI performs the best in terms of decision manipulation counteraction?*

- RQ_3: *Which UI is the most understandable one?*

- RQ_4: *Which dimension in the rating adaptation history best helps to prevent users from decision manipulation?*

4.2 User Study

To address the research questions, we conducted a user study with staff members and students from two universities[3]. In total, there were 120 participants (*males*: 45.83%, *females*: 54.17%). Our user study was performed in the two following steps:

[1]Institute of Software Technology - Graz University of Technology, Graz, Austria
[2]For the detailed of the UIs' visualization, we refer the reader to the following link: http://www.ist.tugraz.at/trang/ManipulationCounteractionUIs/
[3]Hue University of Economics - Vietnam and Graz University of Technology - Austria

Table 3: Kruskal-Wallis test across different transparency levels of the UIs in *Group 1* (α = .05, p = .000).

trans. level	2	3	4	5	6
mean rank	211.66	239.46	256.04	278.80	**288.10**

Step 1 - Distribute the UIs to participants: Each user study participant was provided with *a scenario description* (see Section 3.2) and *a sequence of five UIs* at five different transparency levels. To avoid possible biases, the UIs in each sequence were shown to the participant in a random order. Additionally, the UIs were distributed to the participants using a *between-subjects* method (i.e., each participant received different UI sequence). At a certain time, the participant observed and evaluated *only one UI*. The evaluation for this UI did not rely upon the evaluations of other UIs. In addition, the UIs were distributed so that the total number of participants for each transparency level is equal.

Step 2 - Define criteria to evaluate the UIs: Each participant was asked to evaluate the *understandability* of the UIs using a 5-point Likert scale (1 - *completely not understandable*, 5 - *completely understandable*). Thereafter, the participant had to answer the question ("*Assume you were Maria in the mentioned scenario. If you had known that your rating adaptations would be shown to all group members as in the UI, then what would you have done?*") using the scale of [1..3] which measures the participant's *preparedness level* with regard to decision manipulation (*1 - "manipulate"*, *2 - "not sure/hesitate"*, and *3 - "not manipulate"*). In addition, the participant was asked to give a brief *explanation* of his/her answer. In case the participant decided *"not to manipulate"* the decision, he/she had to additionally specify the *influence level* of each dimension on his/her intention of decision manipulation. The *influence level* of a dimension was measured by a 5-point Likert scale ranging from 1 ("*the dimension did not change the participant's mind regarding decision manipulation*") to 5 ("*the dimension made the participant give up manipulating the decision*").

5 DATA ANALYSIS AND RESULTS

5.1 Data Analysis

To address research questions, we collected the evaluations of the participants for the UIs as follows:

- RQ_1: We separately collected the participants' answers for the UIs of *Group 1* and *Group 2*. In each group, at a specific transparency level, we gathered the participants' answers for all UIs.

- RQ_2: For each transparency level, we collected the participants' answers for all UIs from both *Group 1* and *Group 2*. In total, we had five sets of answers corresponding to five transparency levels.

- RQ_3: We gathered the participants' evaluations regarding the *understandability* of all UIs from two groups.

- RQ_4: We filtered out the participants who decided "*not to manipulate*" the decision. Thereafter, in each UI, we gathered the participants' evaluations concerning the *influence level* of the dimensions.

The collected evaluations share the same characteristics: (i) *independent* (i.e., the evaluation of a UI was independent to the evaluations of other UIs), (ii) *ordinal* (in the range of [1..3] or [1..5]), and (iii) *not normally distributed* (Shapiro-Wilk tests, significance level α = .05, $p < \alpha$). Because of that, we used non-parametric tests

Table 4: Kruskal-Wallis tests for all UIs at the transparency level of 3, 4, and 5. The numbers shown in the table are mean ranks.

trans.level = 3	$UI1_R$	$UI1_{TL}$	$UI1_{TD}$	$UI1_{GR}$	$UI2_R$	$UI2_{TD}$
	61.69	61.66	64.93	**79.45**	41.29	52.80
trans.level = 4	$UI1_{R+TD}$	$UI1_{TL+TD}$	$UI1_{TL+GR}$	$UI1_{TD+GR}$	$UI2_{TD+GR}$	
	75.29	62.76	53.68	63.57	45.93	
trans.level = 5	$UI1_{R+TL+TD}$	$UI1_{R+TD+GR}$	$UI1_{TL+TD+GR}$	$UI2_{all}$		
	68.63	62.21	63.54	46.78		

Table 5: Kruskal-Wallis test ($\alpha = .05$, $p = .008$) in the understandability levels across different UIs.

UI	UI1 basis	UI1 R	UI1 TL	UI1 TD	UI1 GR	UI1 R+TD	UI1 TL+TD	UI1 TL+GR	UI1 TD+GR	UI1 R+TL+TD	UI1 R+TD+GR	UI1 TL+TD+GR	UI1 all	UI2 R	UI2 TD	UI2 TD+GR	UI2 all
mean rank	301.05	363.17	340.66	343.40	339.85	297.35	242.68	263.00	342.14	305.83	206.19	**380.95**	273.34	294.53	309.50	305.15	325.28

(Kruskal-Wallis, $\alpha = .05$) to analyze these evaluations. Besides, for RQ_1 and RQ_2, we additionally ran follow-up Mann-Whitney U tests ($\alpha = .05$) on the same sets of evaluations to further consider the participants' evaluations between pairs of different UIs and this could trigger *Type I errors*[4]. To control this, we applied a *Bonferroni adjustment* [16] to adapt the significance level. The *revised significance level* of each test was $\alpha' = \frac{\alpha}{N}$, N is the number of tests.

5.2 Results

5.2.1 ***Research Question 1*** *- RQ_1.* In *Group 2*, we found out that there were no statistically significant differences in the participants' preparedness levels regarding decision manipulation across different transparency levels (Kruskal-Wallis, $p = .564 > \alpha$). This means, in *Group 2*, it was unclear which transparency level best helps to counteract decision manipulation. In *Group 1*, there existed statistically significant differences in the participants' preparedness levels regarding decision manipulation across different transparency levels (Kruskal-Wallis, $p = .000 < \alpha$). The mean ranks in Table 3 shows that at the *transparency level of 6*, the $UI1_{all}$ achieved the lowest preparedness levels, whereas the $UI1_{basic}$ (*transparency level = 2*) reported the highest. Besides, by performing 10 follow-up Mann-Whitney U tests ($\alpha' = \alpha/10 = .005$) between pairs of *five* transparency levels, we found out that there was a statistically significant difference in the participants' preparedness levels between the UIs at the transparency levels of 2 and 6 (*p=.000 <α'*).

The results can be explained as follows: At the transparency level of 2, the $UI1_{basic}$ only reveals the information of "*who has adapted the ratings of which items*" which is too abstract and deficient to detect who is actually a manipulator. In contrast, at the transparency level of 6, the $UI1_{all}$ (see Figure 1e) makes the rating adaptations of group members *completely transparent.* Therefore, this UI effectively helped to discourage the participants from decision manipulation. Indeed, 66.67% of the participants who observed this UI decided *not to manipulate the decision.* They mentioned that the rating adaptations were so obvious and they might be recognized as manipulators by other group members.

5.2.2 ***Research Question 2*** *- RQ_2. At the transparency level of 3*, there was a significant difference in the participants' preparedness levels regarding decision manipulation across different UIs (Kruskal-Wallis, $p = .009 < \alpha$). An inspection of the mean ranks in Table 4 suggests that the $UI1_{GR}$ (see Figure 1a) best helped to counteract decision manipulation. In fact, the $UI1_{GR}$ (with the "*group recommendation*" dimension) helped the participants to make sure that the second rating adaptation of Maria has actually pushed the group recommendation to the "*sea*" option. This could explain as to why *85%* of the participants were hesitant or gave up manipulating the decision when observing the $UI1_{GR}$.

At the transparency level of 4, we found a significant difference in the participants' preparedness levels regarding decision manipulation across different UIs (Kruskal-Wallis, $p = .018 < \alpha$). The mean ranks in Table 4 show that the $UI1_{R+TD}$ best helped to counteract decision manipulation. In fact, compared to other UIs at the same transparency level, only the $UI1_{R+TD}$ (see Figure 1d) explicitly shows the ratings of items. Moreover, each rating adaptation is additionally represented by a directed arrow showing how item rating has been changed. This could explain as to why 76% of the participants who observed the $UI1_{R+TD}$ were hesitant or decided not to manipulate the decision.

At the transparency level of 5, the Kruskal-Wallis ($p = .038 < \alpha$) test reveals a statistically significant difference in the participants' preparedness levels regarding decision manipulation across different UIs. The $UI1_{R+TL+TD}$ of *Group 1* seems to be the best UI in term of decision manipulation counteraction (see the mean ranks shown in Table 4). However, we also recognized that both $UI1_{R+TD+GR}$ and $UI1_{TL+TD+GR}$ achieved similar mean ranks to this UI. Using Mann-Whitney U tests ($\alpha' = .05/3 = .017$), we proved that there were no significant differences among these three UIs. That means, at the transparency level of 5, the UIs of *Group 1* have the same ability to counteract decision manipulation. When observing these UIs, more than 60% of the participants avoided manipulating the decision since these UIs reveal quite "*clear picture*" of how group members have adapted the ratings of items.

In addition, when answering the RQ_2, we found out that at every transparency level, the UIs of *Group 1* performed better than those of *Group 2* (see the mean ranks in Table 4). Follow-up Mann-Whitney U tests also show that the UIs of *Group 2* had *significantly higher preparedness levels*. For instance, between $UI1_{GR}$ and $UI2_R$ ($\alpha' = .003, p = .000$, *mean rank($UI1_{GR}$)=25.63*, *mean rank($UI2_R$)=14.08*) and between $UI1_{R+TD}$ and $UI2_{TD+GR}$ ($\alpha' = .005, p = .002$, *mean rank($UI1_{R+TD}$)=30.29*, *mean rank($UI2_{TD+GR}$)*=19.02). This could be explained by the missing of *group member* information in the

[4]In hypothesis testing, a *Type I error* involves rejecting the null hypothesis (i.e., "*there are no differences among evaluation sets*") when it is actually true [16].

Table 6: Kruskal-Wallis test in the influence levels across different dimensions of the $UI1_{TL+TD+GR}$ and $UI1_{all}$.

$UI1_{TL+TD+GR}$	*GM*	*I*	*TL*	*TD*	*GR*	
mean rank	44.53	45.72	31.06	46.78	**59.42**	
$UI1_{all}$	*GM*	*I*	*R*	*TL*	*TD*	*GR*
mean rank	262.60	242.69	**272.73**	198.57	225.40	258.99

UIs of *Group 2*. One common explanation of the participants for the UIs of *Group 2* was that: "*It was so hard to track who has adapted the ratings of items*".

5.2.3 ***Research Question 3*** *- RQ_3.* The Kruskal-Wallis ($p = .008 < \alpha$) reveals a significant difference in the understandability levels across different UIs. Besides, the $UI1_{TL+TD+GR}$ (see Figure 1b) was detected as the most understandable one (see Table 5). 78% of the participants who observed this UI found it *understandable* or *compleletely understandable*. The average score for the understandability of this UI was 4.2/5. Typical comments about the understandability of this UI were "*it is very intuitive and understandable*", "*it obviously shows how the ratings of items have been adapted*", or "*it reveals possible attempts of decision manipulations*".

5.2.4 ***Research Question 4*** *- RQ_4.* The Kruskal-Wallis tests show that the dimensions represented in the UIs at the transparency levels of 2, 3, and 4 have equal impacts on the participants' decision manipulation behaviors. However, two exceptions were found in $UI1_{TL+TD+GR}$ (see Figure 1b) and $UI1_{all}$ (see Figure 1e). For these UIs, there were significant differences in the influence levels across different dimensions (Kruskal-Wallis, $p(UI1_{TL+TD+GR}) = .024 < \alpha$ and $p(UI1_{all}) = .006 < \alpha$). Besides, the mean ranks in Table 6 show that: For the $UI1_{TL+TD+GR}$, the "*group recommendation*" dimension had the strongest influence on the participants' mind regarding decision manipulation. For the $UI1_{all}$, the "*rating*" dimension best helped to prevent the participants from decision manipulation. In contrast, in both UIs, the "*timeline*" dimension seemed to have the lowest impact on changing decision manipulation behavior of users.

6 CONCLUSION AND FUTURE WORK

In this paper, we proposed different UIs and investigated which of them are the most understandable and most effective for counteracting decision manipulation in group recommender systems. The outcomes of our study further confirm the *Hawthorne effect* in the context of decision manipulation, which means if group members know that their rating adaptations could be seen by others, then they tend to avoid decision manipulation. This work also provides practitioners with a hint to design UIs which help to avoid decision manipulation issues in group recommender systems.

To the best of our knowledge, up to now, there does not exist any research which proposes UI-driven solutions to counteract decision manipulation in group recommender systems. Therefore, in our work we faced the difficulty in specifying *a baseline* to evaluate our proposed UIs. Another limitation of our research lies in decision manipulation context. In this work, we discussed a manipulation issue in the context of group decisions in which users *are familiar with each other* and make decisions on *small sets of items*. Therefore, within the scope of future work, we will expand our study by proposing UIs counteracting decision manipulation in other contexts (e.g., group members have no relationship and make a decision on a large set of items). Alternatively, the UIs could also be evaluated in *other decision manipulation scenarios* (e.g., a group member alters item ratings to make the favorite options of his/her opponents never be chosen by the system). Besides, the evaluation process will be done with *a bigger set of observations* (in this study, the current sample data set of 120 observations for 17 UIs is quite limited) and with an engagement of *group dynamics aspects* (i.e., *age, gender, education background*, and *personality*).

ACKNOWLEDGMENTS

The work presented in this paper has been conducted within the scope of the Horizon 2020 project OPENREQ (732463).

REFERENCES

[1] 2018. Hawthorne effect. https://en.wikipedia.org/wiki/Hawthorne_effect
[2] Paul-Alexandru Chirita, Wolfgang Nejdl, and Cristian Zamfir. 2005. Preventing Shilling Attacks in Online Recommender Systems. In *Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management (WIDM '05)*. ACM, New York, NY, USA, 67–74.
[3] Vincent Conitzer and Makoto Yokoo. 2010. Using Mechanism Design to Prevent False-Name Manipulations. *AI Magazine* 31, 4 (2010), 65–78.
[4] Michael D. Ekstrand, John T. Riedl, and Joseph A. Konstan. 2011. Collaborative Filtering Recommender Systems. *Found. Trends Hum.-Comput. Interact.* 4, 2 (Feb. 2011), 81–173.
[5] Alexander Felfernig, Müslüm Atas, Denis Helic, Trang T.N Tran, Martin Stettinger, and Ralph Samer. 2018. *Group Recommender Systems: An Introduction*. Springer, Chapter Algorithms for Group Recommendation, 27–58.
[6] Ihsan Gunes, Cihan Kaleli, Alper Bilge, and Huseyin Polat. 2014. Shilling Attacks Against Recommender Systems: A Comprehensive Survey. *Artif. Intell. Rev.* 42, 4 (Dec. 2014), 767–799.
[7] Chen He, Denis Parra, and Katrien Verbert. 2016. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Syst. Appl.* 56 (2016), 9–27.
[8] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. Evaluating Collaborative Filtering Recommender Systems. *ACM Trans. Inf. Syst.* 22, 1 (Jan 2004), 5–53.
[9] Anthony Jameson. 2004. More Than the Sum of Its Members: Challenges for Group Recommender Systems. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '04)*. ACM, New York, NY, USA, 48–54.
[10] Anthony Jameson, Christopher Hackl, Thomas Kleinbauer, and Holm Presented. 2003. Evaluation of Automatically Designed Mechanisms.
[11] Anthony Jameson and Barry Smyth. 2007. *Recommendation to Groups*. Springer Berlin Heidelberg, Berlin, Heidelberg, 596–627.
[12] Gawesh Jawaheer, Peter Weller, and Patty Kostkova. 2014. Modeling User Preferences in Recommender Systems: A Classification Framework for Explicit and Implicit User Feedback. *ACM Trans. Interact. Intell. Syst.* 4, 2, Article 8 (June 2014), 26 pages.
[13] Juan Lang, Matt Spear, and S. Felix Wu. 2010. Social Manipulation of Online Recommender Systems. In *Proceedings of the Second International Conference on Social Informatics (SocInfo'10)*. Springer-Verlag, Berlin, Heidelberg, 125–139.
[14] Cong Li and Zhigang Luo. [n. d.]. Detection of shilling attacks in collaborative filtering recommender systems. In *Proceedings of the International Conference of Soft Computing and Pattern Recognition (SoCPaR) (SoCPaR'11)*. 190–193.
[15] Joseph F. McCarthy and Theodore D. Anagnost. 1998. MUSICFX: An Arbiter of Group Preferences for Computer Supported Collaborative Workouts.
[16] Julie Pallant. 2007. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows Version 15* (3rd ed.). Open University Press, Milton Keynes, UK, USA.
[17] Iván Palomares, Luis Martínez, and Francisco Herrera. 2014. A Consensus Model to Detect and Manage Non-cooperative Behaviors in Large-Scale Group Decision Making. *IEEE Transactions on Fuzzy Systems* 22, 3 (June 2014), 516 – 530.
[18] Paul Resnick and Rahul Sami. 2008. The Information Cost of Manipulation-resistance in Recommender Systems. In *Proceedings of the 2008 ACM Conference on Recommender Systems (RecSys'08)*. ACM, New York, NY, USA, 147–154.
[19] Philip Sedgwick and Nan Greenwood. 2015. Understanding the Hawthorne effect. *BMJ* 351 (2015).
[20] W. Zhou, J. Wen, Y. S. Koh, Q. Xiong, M. Gao, G. Dobbie, and S. Alam. 2015. Shilling Attacks Detection in Recommender Systems Based on Target Item Analysis. *PLoS ONE* 10 (July 2015), e0130968.