

An observational user study for group recommender systems in the tourism domain

Amra Delic¹  · Julia Neidhardt¹ · Thuy Ngoc Nguyen² · Francesco Ricci²

Received: 31 May 2017 / Revised: 15 December 2017 / Accepted: 12 February 2018 /

Published online: 19 February 2018

© The Author(s) 2018. This article is an open access publication

Abstract In this article we argue and give evidence that the research on group recommender systems must look more carefully at the dynamics of group decision-making in order to produce technologies that will be truly beneficial for groups. We illustrate the adopted research method and the results of a user study aimed at observing and measuring the evolution of user preferences and interaction in a tourism decision-making task: finding a destination to visit together as a group. We discuss the benefits and caveats of such an observational study method and we present the implications that the derived data and findings have on the design of interactive group recommender systems.

Keywords Group decision making · Group recommender systems · Observational study · Travel behavioral patterns

✉ Amra Delic
amra.delic@tuwien.ac.at

Julia Neidhardt
julia.neidhardt@tuwien.ac.at

Thuy Ngoc Nguyen
ngoc.nguyen@unibz.it

Francesco Ricci
fricci@unibz.it

¹ E-commerce Group, TU Wien, Favoritenstrasse 9-11/188, Vienna 1040, Austria

² Free University of Bozen-Bolzano, Bolzano, Italy

1 Introduction

Recommender systems for groups are becoming more and more important since many information needs arise in group and social activities such as listening to music, watching movies, traveling, attending social events, and many more. The importance of group recommender systems (GRSs) also has increased due to the social web, where users are not isolated but form interrelated groups of different sizes and compositions. A high number of papers on GRSs have been published (Masthoff 2015) but still, we believe, there is a gap between the current main focus of the research on GRSs and the information search and decision-making support needs of groups.

Research on GRSs often focuses on the core recommendation algorithms, which are based on a preference aggregation strategies. A preference aggregation strategy dictates how to combine individual preferences, which may be conflicting, into a group profile or in a set of recommendations. According to Arrow's theorem, a unique, optimal, aggregation strategy does not exist—and GRSs studies also confirmed that there is no ultimate winner. On a wider perspective, there are only a few studies that concentrate on the full problem of how to design decision/negotiation support functionalities in GRSs: *Travel Decision Forum* (Jameson 2004), *Trip@device* (Bekkerman et al. 2006), *Collaborative Advisory Travel System* (CATS) (McCarthy et al. 2006), *Choicla* (Stettinger et al. 2015). However, to our best knowledge, by now, no observational study of group decision processes in the context of GRSs, beside the one described in this paper, has been conducted. In fact, observational studies are usually conducted in the social disciplines. In Tindale and Kameda (2000) the importance of the discussion process, especially with respect to the information that is shared among group members is emphasized. An extensive overview of studies on group dynamics and the influence of several different aspects (e.g., group composition, group decision process structure, etc.) on the group choices is presented in Forsyth (2014).

The main motivation of this paper is to introduce a new type of studies to GRSs research: observing groups in naturalistic settings. In fact, we believe that the design of novel and more effective GRSs can be initiated if one better observes and understands groups in actions, measures their behaviors, and tries to identify concrete opportunities for computerized systems to become more useful to people. Therefore, in this paper we will illustrate the design, the outcome and the implications of an observational study where groups of people faced a concrete decision task—select a destination to visit as a group—and the researchers monitored the groups before, during and after the task. Moreover, to support our claims on the importance of such observational studies for GRSs, we present the results of several analyses of the collected data and we provide new insights into group decision-making and group preference construction. More precisely, our study has a wide range of motivations, that we list in the following.

- Supporting *decision-making process* is the ultimate motivation for a recommender system. This functionality is even more important in GRSs than for single-user recommenders, which can also be used for other reasons, such as,

expanding user knowledge or expressing oneself (Ricci et al. 2015). But, if group recommenders must effectively support decision-making process, we must understand how this task is executed in groups and how the decision issues, the group members and the contextual situation altogether impact on it.

- We also believe that the *application domain* is a crucial factor that must be considered in the design of a GRS. Recommending tourist attractions or destination for a group cannot follow the same interaction and recommendation model used for suggesting movies to watch (Werthner and Ricci 2004). In fact, the tourism product is more complex than other types of products (e.g., it is a bundle of products and services) and at the same time it is less tangible. Moreover, traveling is an emotional experience and explicit preference characterization is problematic especially in the early phase of the travel decision-making process. Finally, tourism products are typically experienced in groups. For these reasons, we have tried to generate a realistic decision task, i.e., destination selection, in which the study participants could easily imagine themselves. In this scenario, we made observations of users' characteristics and decision outcome that have emerged as important in tourism consumer behavior research (Delic et al. 2016c; Fernández-Tobía 2016; Werthner et al. 2015; Yiannakis and Gibson 1992).
- *Group recommendations techniques* have been influenced too strongly by social choice theory (Masthoff 2015) and not enough by group dynamics studies (Forsyth 2014). It is still unclear how a recommender can identify items to suggest in a group decision-making task, if the goal is not simply to aggregate the votes/preferences expressed by the group members. Hence, we believe that studies like the presented one can help to understand the key information that groups need in order to make decisions, which could not simply be the suggested outcome of the decision. We believe that the more general concept of information recommendation—which information to provide to the group next—rather than product recommendation, is important to implement Blanco and Ricci (2013).
- It is clear to us that the design of more effective GRSs requires a *multidisciplinary approach*. In that sense the study described in this paper brings together social and computer science disciplines. Observational studies are not part of the classical research repertoire of recommender systems research methods. However, we believe that these methods are now strictly required if we want to understand users in naturalistic settings and be able to generate fruitful conjectures about new and useful system functions to be added in a GRS.
- Another important motivation of this study is the desire to *collect data* about group decision-making process that can be exploited by several research groups. Hence, in some sense, an additional goal was to obtain raw data that could be used for different types of analyses, from different perspectives and with alternative motivations. We plan to make the data that we have collected, and that will also be collected in future implementations of the study, available to everyone for further analyses. This objective is of crucial importance for the research in GRSs, since one of the greatest obstacles for making advancements

in the field is the lack of datasets that comprise information about groups, their choices and behaviors.

- Finally, we believe that the research community on GRSs needs to discuss and build a *research agenda*. We must identify critical challenges and expected results. In this study we initiate this reflections by raising several issues, e.g., how to measure the collective behavior of a group, what properties of a group are more important in recommender systems and how they should be measured, how to define group satisfaction, how to compare and relate user preferences and group preferences.

Thus, the main result of this paper is the design of an experimental method for observing group decision-making process and for deriving observational data useful for the implementation of GRSs in the tourism domain. In order to demonstrate the importance of such a method and potential benefits for the further development of GRSs, we illustrate the results obtained by several different analyses of which some were previously published (Delic et al. 2016b, 2017). Moreover, we provide qualitative insights into the group decision-making processes adopted by the study participants. The paper is concluded with a broader reflection on the possible implications for the GRSs research.

We note that this paper is an updated and extended version of “Research Methods for Group Recommender Systems” Delic et al. (2016a) presented at the Workshop on Recommenders in Tourism (RecTour) 2016 held in conjunction with the RecSys 2016 conference (Fesenmaier et al. 2016).

The rest of this paper is structured as follow: Sect. 2 positions this work in the context of the research on GRSs; in Sect. 3 the study procedure is described in detail; Sect. 4 illustrates instruments used for the data collection; in Sect. 5 results of analyses are summarized; followed by Sect. 6 where implications for recommender systems are explained. Finally, in Sect. 7 we discuss limitations, challenges and possible variations of the study.

2 Background

The aim of this section is to position our study and to ease the understanding of its conclusions and implications. Therefore, in the first part, we give an overview of the GRSs research focus, related work and main challenges. In the second part of the section, we aim at clarifying the theoretical concepts used in different phases of the study. Thus, we describe the approach that was used to record the behavior of the participants during the group decision-making process, and we provide a theoretical background of the concepts used in the study questionnaires, i.e., personality model, travel types and social choice theory.

2.1 Group recommender systems state-of-the-art

Recommender systems help their users to find interesting content, for instance, in the overwhelming repository offered by the Web (Ricci et al. 2015). Actually,

recommender systems are employed in various domains, suggesting different types of items. Often these items involve activities that are experienced by groups of people, rather than by single users, e.g., movies, restaurants, travel destinations, etc. Thus, research on recommender systems is more and more dealing with systems that generate recommendations of items that are supposed to be consumed jointly by a group of people. A detailed overview of the state-of-the-art of GRSs is provided by Masthoff (2015). In order to offer a comprehensive overview of the current and previous research activities in GRSs, different research focuses are separately addressed.

Main challenges and aggregation strategies Four major challenges for GRSs were identified and elaborated in Jameson (2004).

1. Elicitation of the group members' individual preferences.
2. Aggregation of the group members' individual preferences to a group model.
3. Representation and explanations of group recommendations.
4. Supporting group members to reach their final group decision.

In fact, current research is mostly focused on the second challenge, i.e., how individuals' preferences should be aggregated into a group model. Three types of aggregation approaches are defined (Jameson and Smyth 2007). In the first approach, the recommender system first generates recommendations for each group member separately and then, in order to produce a group recommendation, it aggregates the individuals' recommendations. In the second approach, the recommender system first predicts the ratings of group members, and then aggregates predicted ratings into a group rating in order to generate group recommendations. Finally, in the third approach the system generates recommendations by using a group preference model that is derived by using existing information about group members.

Commonly used aggregation strategies, i.e., methods to aggregate either *individuals' recommendations* into a group recommendation or *individuals' ratings* into a group rating, are derived from Social Choice Theory (Masthoff 2015). Some of the most popular aggregation strategies are listed below:

- *Plurality voting* Each group member votes for a preferred option and the one with the largest number of votes wins.
- *Borda count* Each group member creates a ranked list of options according to his/her preferences; points are assigned to options, separately for each individual, based on the position of an option in a list (i.e., the last option gets zero points, the second last receives one point, etc.); a group score for an option is calculated as the sum of the individually assigned points; the option with the highest score wins.
- *Copeland rule* Firstly, the pairwise comparison of options is applied, and for each option the number of wins and losses against all other options is counted (i.e., we count how many times an option was rated/ranked higher by group members in comparison to other options). To obtain group scores, number of losses is deducted from the number of wins; the highest score wins.

- *Additive* Individuals' ratings are summed up, the option with the highest score wins. Possible implementations of the additive strategy are to calculate the mean value (i.e., average strategy) or the median value (i.e., median strategy) of the individuals' ratings.
- *Multiplicative* Individuals' ratings are multiplied, and the option with the highest score wins.
- *Least misery* A group score is the minimum of individuals' ratings; the strategy assumes that a group is as satisfied as its least satisfied member.
- *Most pleasure* A group score is the maximum of individuals' ratings; the strategy assumes that a group is as satisfied as its most satisfied member.
- *Weighted average* Based on certain metrics, weights are assigned to group members, and thus, a group score is a weighted average of individuals' ratings; the strategy assumes that in certain cases the wishes of some group members should be valued more than those of other group members.

Research has clearly demonstrated that there is no strategy that outperforms all the other aggregation strategies in any given situation.

Influence and roles in group recommender systems A very important section of the research on GRSs is dedicated to defining and identifying (1) the influence that a group member can have on determining the final choice of a group, and (2) the role that a group member plays in a group. The first researchers that tackled this issue were Masthoff and Gatt (2006). They defined two types of influence: (a) emotional contagion and (b) conformity. In the same paper, the authors also introduced several satisfaction functions that account for the influence in groups. Later on, contributions from other researchers arose and different types of role-based and influence-base group recommendation approaches were introduced. For example, a very simple role-based approach was introduced and evaluated in Ali and Kim (2015). The authors defined different group members' roles and accordingly assigned them weights in the aggregation strategy based on the group member's activity in the system, i.e., the more item-ratings a group member provided the greater the weight would be. However, it is noteworthy that the group context was disregarded and only individually provided item-ratings were considered. In the work of Berkovsky and Freyne (2010), three role-based models were introduced, and all three took the similar approach as in the previous case. The main difference is the integration of the group context in the models. The third approach, introduced in Gartrell et al. (2010), defined weights based on the number of item-ratings, but only considering a pre-selected set of movies. A considerably different approach was introduced by Quijano-Sanchez et al. (2013). The authors defined influential group members, and accordingly delivered group recommendations, based on (a) group members' personality strength, i.e., the more assertive a group member is the greater influence of that group member is assumed; and (b) social relationships between group members. Finally, in Quintarelli et al. (2016), the authors defined influence based on the match/mismatch between users' individual choice and the group choice in which a user has participated. For example, if a user was a member of six different groups and her preferred option was selected as the group choice in three out of six cases, then her weight in the influence-based model is $3/6 = 0.5$.

Group recommender systems in the travel and tourism domain Various research activities were dedicated to develop and evaluate GRSs to support group decision-making process in the tourism domain:

- *Intrigue* Ardissono et al. (2003) assists tour guides to plan touristic tours for heterogeneous groups with somewhat homogeneous subgroups (e.g., children, elderly). The system generates personalized recommendations by matching the attributes of tourism attractions to the explicitly given preferences of subgroups, and it uses the weighted average strategy to build a group preference model. The weights applied in the aggregation strategy are adjusted to the subgroup importance.
- *Travel decision forum* Jameson (2004) is a system that allows its users, i.e., group members, to decide on preferred attributes of a joint holiday. The main idea of the system is to simulate a face-to-face, asynchronous discussion, by allowing group members to use animated characters. In order to build the group model and to aggregate individuals' preferences, the system uses the additive and median strategies.
- *Trip@dvice* Venturini and Ricci (2006) is a case-based reasoning recommender system with a cooperative negotiation methodology approach. The system uses automated negotiation agents as mediators of a cooperative negotiation. The case-based reasoning module generates individuals' recommendations, which are then used as group members' proposal items for the group. To generate group recommendations, the negotiation agents apply one of the available negotiation strategies (e.g., maximizing the utility of the least happy group member) and chooses one of the previously generated proposal items as an agreement for the group. Based on different aggregation approaches, the system generates several more suggestions for the group.
- *Collaborative Advisory Travel System* (CATS) McCarthy et al. (2006) allows group members to express their opinion about each others preferences and preferred options by employing the critiquing approach. Critiquing-based techniques allow users to comment, i.e., critique, a specific item or item-attribute, e.g., "I would prefer a destination that is not that distant", meaning that a user is critiquing the distance attribute of the destination. The system adapts the next set of recommendations accordingly. In CATS, this specific approach was used to support the negotiation process, i.e., group members can comment on each-others' item-attribute preferences and the group model is built as the average of individuals' preferences.
- *Where2eat* is a mobile app for restaurant recommendation that implements "interactive multi-party critiquing", i.e., an extension of the critiquing concept to a computer-mediated conversation between two individuals (Guzzi et al. 2011). The system allows group members to generate proposals and counter-proposals until the agreement is reached.

As we mentioned in the introduction, only recently, the research on GRSs has started to acknowledge the importance of the group decision-making process and the dynamic of group members' preferences through the decision-making process

(Nguyen and Ricci 2016, 2017a, b, c; Nguyen 2017). In this works, the authors aim at generating group recommendations not only based on individual and independent preferences of group members, gathered outside of the group context, but also based on the preferences that evolve during the group decision-making process. The authors propose a group model that combines group members' individuals and independent preferences with the preferences constructed within the group decision-making process.

2.2 Group decision-making and observational studies

A very small fraction of the research on GRSs is dedicated to understanding how groups make choices and, therefore, how the group decision-making process can be supported (Chen et al. 2013). An example of a group recommendation study that can be described as an “indirect” observational study of group decision-making processes was conducted by Masthoff (2004). The participants were asked to create an item-sequence, i.e., a ranked set of recommendations, for a given, fictional group of people, based on their individual, independent item-ratings. The objective for the study participants was to maximize the satisfaction of group members with the generated item-sequence. The author aimed at understanding if participants would use certain aggregation strategies when deciding the best item-sequence for a given, fictional group, and how would they explain the goodness of fit of the generated item-sequence. Moreover, in the same study, the author designed a second experiment where participants were asked to imagine themselves in a group of three, they received item-ratings of each group member, including themselves, and asked how satisfied they and the rest of the group would be if the system recommended certain item-sequences.

In social science disciplines numerous observational studies have been conducted and a considerable amount of literature about group decision-making processes exists. For example, Tindale and Kameda (2000) discuss the importance of the so called “social sharedness”, i.e., the extent to which preferences, information or anything related to a group-decision making process, is exchanged and shared between the group members. The authors found evidence of “social sharedness” being one of the key elements in understanding group decision-making outcomes. Moreover, researchers who study the functional theory of group decision-making observed that groups that reach their decisions in a more structured fashion, actually, are more likely to make better decisions. In Forsyth (2014) an approach to structure a decision-making process is proposed. The approach suggests that four phases should be adopted:

1. *Orientation phase* The group defines important aspects and goals of the decision-making process:
 - The problem that needs to be solved.
 - Goals that should be achieved.
 - Strategy and procedures that should be used in the process.

2. *Discussion phase* In this phase, a “communication peak” should be reached. Group members exchange collected information, opinions, agreements and disagreements. The main tasks of the phase are:
 - Gathering relevant information.
 - Exchanging information.
 - Discussion about possible alternatives.
3. *Decision phase* Based on the previous phases a group makes a decision using a decision scheme, e.g., voting, consensus reaching, etc. If a decision cannot be reached a group can return to any of the previous phases.
4. *Implementation/evaluation phase* A decision is implemented and evaluated.

While we believe that structured decision-making approaches should be considered when developing a GRS, as a matter of fact, current GRSs, as we mentioned already, focus on the generation of suggestions for a group, based on individuals’ preferences, hence only marginally attack the issue of how to better support the full decision-making process. Thus, in this study, we aim at understanding how to truly facilitate groups in their decision-making process with the GRSs.

Many different approaches to perform an observational study and record interactions within small groups exist. In our study we use that proposed by Bales, i.e., the interaction process analysis (IPA) (Bales 1950; Forsyth 2014). IPA is a coding method for observing group interactions and it is widely used as it increases the objectivity of observations. The approach requires from an observer to identify a “unit” of interaction for each group member. Bales defines a “unit” of interaction as a single simple sentence or its equivalent. Therefore, complex sentences containing an independent clause and at least one dependent clause, or compound sentences joined by “and”, “but”, “or”, should be broken down into a single expression “unit”. For example, if a group member states “How about voting, but I think we still might not get the winner.”, the observer should break down the sentence into two “units”: (1) “How about voting”, and (2) “I think we still might not get the winner.”. Furthermore, in addition to speech, a “unit” of interaction includes also facial expressions, gestures, body attitudes, emotional signs, etc. Then, for each group member, the observer categorizes each “unit” of interaction into one among twelve behavior categories:

1. Show solidarity/“Friendly” (e.g., expressing gratitude or appreciation; apologizing, or smiling directly at another; offering assistance, time, energy, money; etc.).
2. Show tension release (e.g., showing cheerfulness, satisfaction, enjoyment, relish, pleasure, etc.).
3. Agree (e.g., agreement reflected through verbal or nonverbal expressions).
4. Give suggestion (e.g., mentioning a problem to be discussed: “I want to call your attention to the budget issue”).
5. Give opinion (e.g., stating judgment or inference: “I believe that Amsterdam is the most beautiful place to visit in spring”).

6. Give information (e.g., reporting factual, verifiable observations or experiences: “The weather in Amsterdam at this time is not good”).
7. Ask for suggestion (e.g., requesting guidance in problem-solving process).
8. Ask for opinion (e.g., questions seeking value judgment, beliefs or attitudes).
9. Ask for information (e.g., questions requesting a simple factual, descriptive, objective type of answer).
10. Disagree (e.g., rejecting another person’s statement).
11. Show tension (e.g., appearing startled, blushing, showing embarrassment).
12. Show antagonism (e.g., attempting to override the other in conversation, interrupting the other, making fun of others, criticizing, ill-treating, tricking, deceiving, etc.).

These categories are split in order to capture (a) *relationship interactions* (i.e., categories from 1 to 3, and 10 to 12) and (b) *task interactions* (i.e., categories from 4 to 9). The categories are grounded on Bales’s long-term work on group interactions observations. The IPA system enables qualitative analysis as the behavior of each group member is classified and quantified in a clear manner. In our best knowledge, no studies have tried to relate observations recorded with the IPA system with the theoretical concepts used in this study, i.e., the Big Five factor model and travel types.

2.3 Theoretical concepts of the study

The research on groups and their performance in particular tasks, such as the decision-making task, has shown that inter-subject relations (i.e., the group dynamics, group identity, etc.), emotions, personality, group similarity of interests, opinions, preferences, etc., play an important role in the final outcomes of those tasks (Forsyth 2014). However, those aspects are often neglected in the research of GRSs. To this end, in our study, besides individual explicit preferences of group members, we covered additional aspects that we believe might have an impact on the final outcomes of the group decision-making process.

The Big Five factor model In psychology research, many models have been developed to capture individuals’ characteristics and to explain their overall behavioral patterns. One of the most widely used models, in this sense, is the five-factor model of personality, also known as the Big Five (McCrae and Costa 1987). It breaks down the personality into five orthogonal dimensions: (1) *openness to new experiences*, i.e., the extent to which someone is prone towards experiencing new and unusual things; (2) *conscientiousness*, i.e., the extent to which one is precise, careful and reliable, or rather sloppy, careless, and undependable; (3) *extraversion*, i.e., the extent to which people are outgoing, cheerful, warm, or rather quiet, timid, and withdrawn; (4) *agreeableness*, i.e., the extent to which someone is altruistic, caring, and emotionally supportive, or rather indifferent, self-centered and hostile; (5) *neuroticism*, i.e., the extent to which someone experiences distress or rather is calm and even-tempered (McCrae and John 1992). The five-factor model of personality has been converted in many bigger and smaller measures, i.e., with more

and less dimensions (Donnellan et al. 2006), and is used in a wide range of application domains, including tourism (Neidhardt et al. 2014).

Travel types Specific for the tourism domain, there is an important line of research that is concerned with the relationship between individual characteristics, psychological needs and personal expectations on the one hand, and travel-related attitudes on the other. A well-established classification of tourist preferences is offered by the framework introduced in Gibson and Yiannakis (2002), which distinguishes, as authors named them, 17 Tourist Roles. Even though these Tourist Roles represent short-term characteristics, if compared to the long-term Big Five factors, evidence exists for associations between these two constructs (Delic et al. 2016c). Factor analyses on the 17 Tourist Roles and the Big Five yielded seven basic travel types, i.e., *Sun and Chill-out*, *Knowledge and Travel*, *Independence and History*, *Culture and Indulgence*, *Social and Sport*, *Action and Fun* and *Nature and Recreation* (Neidhardt et al. 2014).

Social identity theory Social psychology is a branch of psychology that deals with relations of individuals' circumstantial and social characteristics with individuals' attitudes and behavior in the context of social groups. It analyses the influence of social groups on personal processes, close relationships, intergroup and societal phenomena (Fiske et al. 2010). Social identity theory emerged as an extension to a wide-spread research on small groups in social psychology, trying to account for another set of dimensions related to the, so called, social identity (Tajfel 2010).

Social identity is defined in terms of how one perceives himself/herself in relation to a social environment together with one's sentiment of belonging to that particular social environment, i.e., it is the "*individuals* self-concept which derives from their knowledge of their membership to a social group (or groups) together with the value and emotional significance attached to the membership" (Tajfel 2010). However, social identity theory does not define the general concept of identity or the "self-concept", but it rather claims that an important part of the overall "self-concept" is a result of one's association to a certain social group or category. Therefore, the social identity theory explores the role of social identity in relation with how groups of people are formed and how members relate to each other in those groups.

In our study, we focus on the strength of participants' identification with the others in their group (further referred as the group identification). In that sense, strong group identification means: (a) a member feels a *high level of belonging* to a particular group; (b) a member is *willing to participate* in a group activity; and (c) a group member *wants to belong* to a particular group. Strong group identification occurs even when preferences related to some specific topic are not shared among group members, but they perceive similarity on a more comprehensive level, i.e., the social identity level.

3 Procedure

The study was initiated in a cooperation with the International Federation for Information Technologies in Travel and Tourism (IFITT) and 11 universities worldwide. The first implementations of the study took place at the Delft University of Technology (TU Delft), the University of Klagenfurt (UNI Klagenfurt) and the University of Leiden (UNI Leiden), while an extended study was carried out at the Vienna University of Technology (TU Wien). Each implementation was conducted as a part of a regular lecture and followed a three-phases structure: a pre-questionnaire phase, groups meetings/discussions phase and a post-questionnaire phase (see Fig. 1).

Prior to the first study phase, an introductory presentation containing the general instructions for the participants was arranged. The first task for all participant was to form groups. At TU Delft, UNI Klagenfurt and UNI Leiden, students were free to form their groups and decide the size, but they were requested not to exceed the size of five members in a group. At TU Wien students were instructed to form groups of six members and to select two students (further referred to as *observers*) whose task was to observe and record activities of their group in the second study phase. All the other group members took part in the decision-making process (further referred to as *decision-makers*). It is important to note that the detailed recordings of the decision-makers behavior was part of the TU Wien study implementation only.

In the first study phase, the task for the decision-makers was to fill out the online, pre-questionnaire that captured their individual profiles, preferences and dislikes (for details see Sect. 4). Also, in this phase at TU Wien, a short training for observers was organized. The purpose was to introduce the observers with the details of the second and third study phases, and to instruct them on how to perform and document the observations of group behavior. A report template for documenting the group behavior, i.e., actions of the decision-makers, designed based on Bales's IPA (Bales 1950), was clarified and distributed to the observers. Moreover, the observers received detailed written explanations on how to perform observations and a continuous contact with them was maintained until the end of the study.

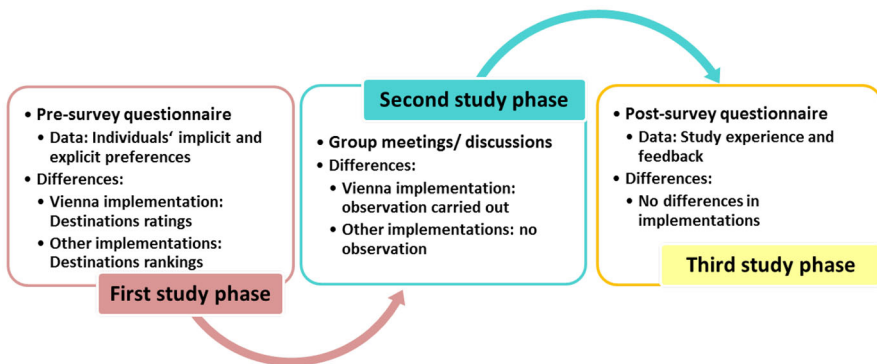


Fig. 1 Overall structure of the study and differences between implementations

In the second study phase, the groups meetings and discussions took place. To this end, the decision-makers received written instructions with the following structure:

1. Ten predefined destination options together with informational Wiki pages.
2. Description of the decision task scenario: “Imagine that you are working on a research paper together with the other group members. Interestingly, your university offers you the opportunity to submit this paper to a conference in Europe. If the paper gets accepted, the university will pay to each group member the trip to the conference. In addition, you will be able to spend the weekend after the conference at the conference destination. Ten conferences will take place in European capitals around the same summer period”.
3. Decision task: “Decide to which conference (destination) you will submit your paper, and what would be your second choice (in case the first choice would not be feasible for some unexpected reason)”.

Groups were not instructed on how to perform the decision-making task and whether they should necessarily check the informational Wiki pages or not. This specific design was chosen due to its simplicity. Usually, when a group is planning a trip a number of different trip aspects have to be considered, e.g., timing, budget, destination, accommodation, transport, etc. A proper discussion on all these issues would be almost impossible to simulate in a controlled environment. Thus, we concentrated on a simple aspect, i.e., the selection of a destination, to analyze the basis of group interactions and dynamics in this specific context. At TU Wien, observers were included in the group work. They audio recorded and documented the group decision-making process using the Bales’s IPA report template (for details see Sect. 4).

In the third phase, the decision-makers filled out an online, post-questionnaire inquiring about the previous phase and the overall experience. During this phase, interviews with the observers were arranged in Vienna: for each group one meeting with the two observers and one of the authors of this paper. At the interviews, firstly, we asked the observers to explain different sections of their report template and behavior categories in order to evaluate their understanding of the task they were given. Secondly, the two observers elaborated their own submissions and compared them, if the recordings differentiated to a great extent, the observers were asked to come to an agreement and revise their reports.

At each university the study implementation followed the described structure. However, some minor differences existed, they are explained in Sect. 7. After the first implementation round, considering all the locations where the study was conducted, the size of the collected data sample comprised 78 decision-makers in 24 groups of 2, 3 and 4 members, plus 16 observers, two for each of the eight groups at TU Wien. At TU Delft, after a first implementation round (referred to as *TU Delft*), a second one with the same configuration took place (referred to as *TU Delft2*). It introduced 122 new decision-makers in 31 groups. Thus, at the end the data sample comprised 200 decision-makers in 55 groups of 2–5 group members (see Tables 1, 2) plus 16 observers.

Table 1 Groups sizes per university

Group size	2	3	4	5
UNI Leiden	2	2	2	/
UNI Klagenfurt	1	1	4	/
TU Delft	1	2	1	/
TU Delft2	1	8	14	8
TU Wien	2	1	5	/
SUM	7	14	26	8

Table 2 Study participants demographics

Age	Gender	Country
Min: 17		Netherlands: 114, Austria: 31
Median: 21.50	Male: 166	Spain: 7, China: 4
Mean: 22.46	Female: 34	Russia: 4, Singapore: 4
Max: 48		USA: 4, Other: 32

4 Measurements

In this section we describe the collected data in detail as well as the instruments used to collect it: the pre-questionnaire, the template for documenting the observations of the group behavior and the post-questionnaire. The instruments were designed based on existing literature (see Sect. 2) with the goal to cover different aspects, that might have an impact on the group decision-making process and its outcomes.

The first data collection instrument, i.e., the pre-questionnaire¹ captured a rich user profile of the participants. The questionnaire comprises 68 statements separated into four sections:

1. Demographic data (i.e., age, gender, country of origin, university affiliation and student identification number).
2. Tourist roles and Big Five factors:
 - 30 questionnaire statements were related to the 17 tourist roles (see Sect. 2).
 - 20 questionnaire statements were related to the Big Five factors (see Sect. 2).
3. Ratings or ranking of the ten predefined destinations and the experience related to those destinations:
 - *Destinations* Amsterdam (at TU Wien and UNI Klagenfurt), Berlin, Copenhagen, Helsinki, Lisbon, London, Madrid, Paris, Rome, Stockholm and Vienna (at TU Delft and UNI Leiden).

¹ <https://survey.aau.at/2012/index.php?sid=49577&lang=en>.

- *Experience* Participants were asked how many times they have visited each destination.
- *Ratings and ranking* Participants at the TU Wien rated, while other participants ranked the ten destinations (implications of this distinction are discussed in Sect. 7).

4. Ranking of decision criteria for choosing a travel destination (i.e., budget, weather, distance, social activities, sightseeing and other).

A five-point Likert scale was used for the questionnaire statements related to the 17 tourist roles and the Big Five factors. To obtain the scores, i.e., the level to which a person belongs to a certain tourist role or to a certain personality trait, ratings of the statements were normalized (i.e., summed and divided by the number of related questionnaire statements).

In the second phase the group decision task took place. As previously mentioned, at the TU Wien the observers recorded behavior of the decision-makers. As explained previously, the report template for the observers' recordings was designed based on the Bales's IPA (see Sect. 2). Thus, the task for observers was to audio record group discussion and to fill out the provided report template. The report template consisted of the following sections:

1. Decision-making process planning and execution: whether a specific plan for the group decision process was used or not and if yes the duration of the different decision process phases.
2. Group members' roles: e.g., leader, follower, initiator, information giver, opinion seeker.
3. Group members' behavior: Bales's IPA system and twelve categories of behavior (see Sect. 2).
4. Social decision scheme: when groups engage in a decision-making task, usually they adopt a type of a decision scheme to make a final choice, i.e., averaging—the group makes decisions by combining each individuals preference using some type of computational procedure; voting—the group selects the destination favored by the majority of the members; reaching consensus—the decision is made when everyone agrees on a course of action and expresses satisfaction with the decision; observers could also provide a description of the decision scheme in their own words.
5. Strength of group members' preferences: the observers rated group members' willingness to give up on their initially preferred options, on a scale from 1—*very unwilling* to 5—*very willing*.

To complete this task properly, observers attended a lecture with instructions on how to perform observations. At the lecture, each part of the report template was explained in detail. Furthermore, each behavior category of the IPA system was thoroughly elaborated with examples applicable to the decision-making task at hand.

Finally, a post-questionnaire² was used to collect data about the participants' experience with the group decision-making process and the overall study. It asked for:

1. Group choice: i.e., the first and the second preferred destination of the group.
2. The usage of the provided info about the ten destinations: i.e., the participants were asked whether or not they used the provided information about the destinations during the group decision-making process.
3. Textual description of the group decision-making process employed by the group: i.e., "Shortly describe how you reached the group decision?".
4. Overall attractiveness of the ten predefined destinations: e.g., "Many destinations were appealing.", "I did not like any of the destinations.".
5. Satisfaction with the group choice: e.g., "I like the destination that we have chosen.".
6. Difficulty of the decision process: e.g., "Eventually I was in doubt between some destinations.".
7. Participant's perceived group identification: e.g., "I identify with the other students in my group.", "I see myself as a member of this group.") and preferences similarity with the other group members (e.g., "I considered myself similar to the other members in my group in terms of our preferences.", etc.
8. Assessment of the task: participants were asked to select the statements to which they agree regarding the organization of the task, i.e., "The task was well described.", "More and better instructions on what we should have done would have been helpful.", "I did not understand what we should do.", "Most people in our group had no idea what we should do."), their feedback (e.g., "The exercise was chaotic.", "I learned something.", etc.), and willingness to participate in the same or similar study (i.e., "Would you like to participate more often in exercises like this one?").

A five-point Likert scale was used to assess 4., 5., 6. and 7.

The overall structure of the data, and the different aspects that were collected with the three instruments is shown in Fig. 2. Moreover, different colors indicate different study phases, i.e., rose: pre-questionnaire, blue: groups meetings/discussions, and yellow: post-questionnaire. Central entity in the diagram is the group member, i.e., the decision maker who is connected to all the other data dimensions.

5 Findings

In this section we present the results obtained by the several data analyses conducted on a sample of 200 participants in 55 groups.

² <https://survey.aau.at/2012/index.php?sid=98597&lang=de>.

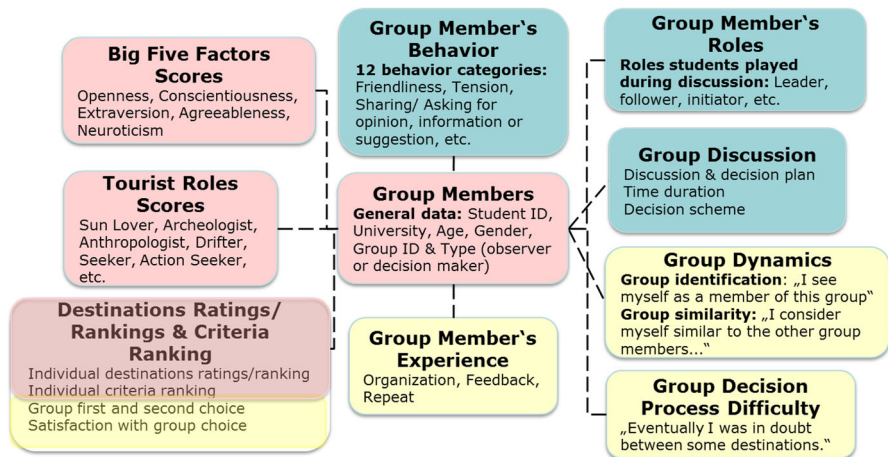


Fig. 2 Structure of the collected data

5.1 Exploratory analysis on choice satisfaction and aggregation strategies

In a first data analysis we studied whether or not the decision-makers were satisfied with the outcome of the group decision-making process, and we tried to understand the impact of their initial preferences on that satisfaction. The vast majority of participants showed a high satisfaction for the destination chosen by the group, i.e., they indicated that they were excited about this destination. Obviously, for those whose individual top choice matched the group selection (73 out of 200, 36.5%), this was particularly true (67 out of 73, 91.8%). However, most decision-makers whose top-choice was not the group choice (127 out of 200, 78.0%) were also satisfied (99 out of 127, 63.5%), see Table 3. To some extent this might be related to the fact that the decision-makers perceived the ten offered destinations overall as very attractive, or the best attainable compromise given the group members' preferences. However, why people are satisfied with a choice that is not their preferred item is a focus of our second analysis, summarized in Sect. 5.2.

A Chi-square test for the contingency Table 3 shows that the two dimensions are not independent ($p = 0.01$), hence significantly more people are excited about a destination when it matches their pre-discussion preferences.

However, as demonstrated in the further text and supported by the second analysis, individuals' satisfaction does not only depend on the match between individual and group preferences but on a great variety of factors, including the group decision-making process, and characteristics of the individuals as well as of the groups. Thus, below in this paper we will show that: (1) group choice is not just an aggregation of the group members' individual preferences, but that it is rather constructed during the process, and (2) individuals' satisfaction is related to certain characteristics of the individuals and groups.

The first statement is supported by the fact that common aggregation strategies used in GRSs are hardly able to predict the outcome of the group decision-making

Table 3 Contingency table: preferences match and excitement

	Excited	Not excited
Match	67	6
No match	99	28

process. To this end, we calculated the prediction precision of the first and second group choice computed by those aggregation strategies. As introduced in Sect. 2, an aggregation strategy is applied on the group members' individual preferences, e.g., ratings, to compute a group recommendation. In our case, with the aggregation strategies we try to "predict" the actual group choice based on the group members' individual ratings of the ten pre-defined destinations (acquired within the pre-questionnaire). This analysis is important as it can provide the first insights in clarifying the relevance of the group members' individual, pre-discussion preferences, as well as the relevance of an aggregation strategy in predicting the opinion or satisfaction of an individual group member with the actual group choice:

$$Precision = \frac{|TP|}{|TP| + |FP|}. \quad (1)$$

In this formula, *true positives* (*TP*) are group choices that a strategy correctly puts in the top-k items (i.e., top-1 or top-2), and *false positives* (*FP*) are the options in the top-k set, as predicted by an aggregation strategy, that were not selected by a group. The results i.e., the average precision computed on 55 groups, are shown in Table 4.

The multiplicative strategy, in general, outperformed other strategies, which is in-line with previous results (Masthoff 2004). The general satisfaction of participants with the final group choice indicates that the performance of an aggregation strategy in terms of predicting the actual group choice might be of minor relevance, since a group member might be satisfied even when her individual top-choice is not selected by a group. However, the performance of an aggregation strategy in terms of individuals' satisfaction with the group choice is of great importance and requires a user study, after all, as it was shown, the pre-discussion preferences are not always an indicator of what a group member will say about the actual group choice.

Therefore, it is clearly relevant to identify other factors that play a substantial role in determining outcomes of group decision-making processes. In our analysis, by the outcomes of the group the decision-making process we consider, (1) the actual group choices, and (2) the choice satisfaction of individual group members.

Table 4 Performance of aggregation strategies

Strategy	Precision top-1	Precision top-2
Additive	0.3291	0.2471
Multiplicative	0.3333	0.2571
Median	0.2788	0.2146
Least misery	0.2281	0.1922
Most pleasure	0.1634	0.1492

In the next step, we studied in more details the relationship between the choice satisfaction and characteristics of the individuals and the groups. We found that the choice satisfaction was significantly and positively correlated with *Agreeableness* and *Conscientiousness*, and negatively correlated with *Neuroticism* (Delic and Neidhardt 2017). Obtained correlations are in-line with the personality theory—people with more agreeable and open personalities are easier to be satisfied, compared to those scoring high on *Neuroticism*. Moreover, behavioral categories that were observed and recorded during the decision-making process were found to be related to the choice satisfaction as well as to the perceived difficulty of the decision-making task. Choice satisfaction was significantly and negatively correlated with the *Give opinion* and *Ask for suggestion* behavioral categories, and the perceived difficulty of the decision-making task was significantly and positively correlated with the *Give opinion* and *Ask for opinion* behavioral categories.³

5.2 Analysis on determinants of choice satisfaction

To better understand when and why the decision-makers were highly satisfied or, on the other hand, not so satisfied with the final group choice, we conducted a second analysis (Delic et al. 2017). Firstly, we explored to which extent the choice satisfaction was related to the distance between the group members' individual preferences and the final group choice. Thus, we calculated the Kendall-tau distance between individuals' ranked destinations and groups' top two choices, and correlated it with the satisfaction measure. As expected, a significant correlation was found, but only with a moderate correlation score (-0.35 , $p < 0.001$). Therefore, to examine what other factors may influence the level of individuals' satisfaction, we identified high and low satisfied decision-makers, and we analyzed differences between the two. A *t*-test revealed that high satisfied decision-makers scored higher on the *Conscientiousness* and *Agreeableness* personality traits, and also on the *Social and Sport* and *Action and Fun* travel types. At the same time they scored lower on the *Neuroticism* personality trait. Additionally, they perceived the group decision process as easier, the group similarity as higher, and their group identification was stronger. Finally, the analysis showed that decision-makers with a more collaborative personality were generally more satisfied with the final group choice.

In the next step of the analysis, two additional categories of decision-makers were introduced: (1) winners, i.e., those whose individual preferences were close to the final group choice, and (2) losers, i.e., those whose individual preferences were further away from the final group choice. It was especially appealing to investigate what are the differences between high and low satisfied losers in this case. We found out that those who fall into the losers category and who were still satisfied with the final group choice, in general, were more open to new experiences, extroverted and agreeable, and, again, less neurotic. These findings were consistent with theory and research results on the five-factor model of personality (Donnellan et al. 2006;

³ In an analysis that we conducted on a smaller data sample of 78 participants in 24 groups, published in Delic et al. (2016b), we found similar, but slightly different results.

McCrae and Costa 1987; McCrae and John 1992). Finally, the results showed a significant difference in reported choice satisfaction for individuals with active (not avoiding) or passive (avoiding) style of resolving a conflicting situation. Active and passive behavior styles were identified based on the Thomas–Kilmann conflict resolution styles (Kilmann and Thomas 1977). To assign each decision-maker to one of the Thomas–Kilmann styles, a relationship between personality traits and the conflict resolution styles was established as suggested in Wood and Bell (2008). Therefore, a passive person, i.e., with an avoiding conflict resolution style, scores low on the *Agreeableness* as well as on the *Extraversion* personality trait. It was found that decision-makers with a passive (avoiding) conflict resolution style were highly satisfied with the final group choice when it matched their own initial preference, but they were extremely dissatisfied with the final group choice in case of a mismatch with their own initial preference.

5.3 Choice satisfaction at the group level

Of course, the satisfaction of the individual is of crucial importance, but the satisfaction of a group as a whole plays an important role as well. To capture the satisfaction of a group, we studied the average choice satisfaction of the group members. Statistical tests identified significant differences between highly and less satisfied groups with respect to a number of factors. These factors captured, on the one hand, whether or not the group perceived the task as difficult. On the other hand, they were related to aggregated travel behavioral patterns (i.e., more satisfied groups scored higher on the *Social and Sport* and lower on the *Sun and Chill-out* travel factor), as well as personality traits of the group members (i.e., more satisfied groups scored higher on the *Openness to new experiences* and lower on the *Neuroticism* personality trait). Furthermore, in less satisfied groups, the observers recorded a significantly higher level of disagreement during the group decision-making process.

5.4 Qualitative insights into the adopted group decision-making processes

The aim of the qualitative analysis is to provide more details on the actual decision-making processes adopted by groups for our study task (i.e., the selection of a destination to visit together as a group). Moreover, the goal is to identify aspects in which the adopted decision-making processes differed among groups. The overall objective would then be to explore the relationship between the different types of the decision-making process, characteristics of the group and the decision-making process outcomes.

Several types of group decision-making processes were adopted by group members to reach their final decisions. The processes mainly differed in three identified aspects: (a) preferences disclosure technique (i.e., how the decision-makers expressed their individual preferences); (b) discussion type (i.e., whether they exhaustively discussed different options or not); and (c) decision reaching technique (i.e., whether the decision-makers voted for their final choice or they tried to convince each-other until they reached a consensus).

5.4.1 *Preferences disclosure technique*

To disclose individual preferences, decision-makers, employed one of the following techniques: (1) top-choice disclosure (“Every group member stated his favorite locations.”); (2) the elimination process or the least misery approach (“We discussed which cities everyone did not want to visit because he/she has already been there/hates it/doesn’t find it appealing.”); (3) disclosure of the general expectations, criteria, pros and cons of the ten destinations (“..we talked about what are the criteria to rule out cities. We came up with architecture and the distance to the sea.”, “Firstly we described our expectation from the vacation.”).

5.4.2 *Discussion type*

Whether the group discussed their options in length or not was, of course, related to the preferences disclosure technique. Groups that started with their top-k choices or the elimination process, in general, seemed to spend less time on discussion, since they could identify similarities in group members’ individual preferences early in the decision-making process. On the other hand, groups that started with their expectations and criteria spent more time on discussion, since usually they discussed each destination in the choice set before making a decision “Discuss each of the destinations and each member explains why he/she wants or doesn’t want to go there.”.

5.4.3 *Decision reaching technique*

To reach the final decision, decision-makers either voted or managed to convince each-other on a certain choice. It was consistently observed that groups with a higher preferences diversity employed the majority voting strategy as they did not have other way to agree on a final choice “Our initial plan was to discuss the destinations until everyone was consent and happy about the decision. This was probably very naive, since this is very unlikely to happen. Our interests for vacations were very different, so it was not possible to find a location where every could do the things he or she wanted to do. Therefore we later on decided to do a majority vote between the two most popular destinations.”. Finally, a very specific approach was adopted by a certain number of groups, i.e., they assigned points to each destination, and then made their decision based on the number of points that each destination received. In some cases they obtained points from individually ranked lists and in some cases they explicitly assigned points to a number of destinations “We all named our top-3, then gave the No-1 10 points, the No-2 5 points, and the No-3 3 points. Everybody also got the opportunity to give one city -5 points if they did not want to go there. If cities ended with the same amount of points, we did a separate vote including only those cities.”.

Clearly, the groups adopted the decision-making approach that fit them the best—different groups managed to reach their final decisions in different ways still being satisfied with the outcome. Therefore, to deliver group recommendations, the question is not only what to recommend, but also how given the group at hand. The

goal, however, should be clear and driven by the maximization of decision-makers' satisfaction.

To summarize, in this section we have showed the following results:

- Majority of the decision-makers were satisfied with the final group choice even when their top choice was not selected by their group.
- Aggregation strategies applied on the group members' individual, pre-discussion preferences can hardly predict the actual group choice.
- The choice satisfaction of group members is related to their personality.
- Collaborative behavior style is related to the greater choice satisfaction regardless of the group decision-making outcome, while the satisfaction of those with a passive behavior style is profoundly related to the match/mismatch between their individual preferences and the final group choice.
- The choice satisfaction of the group as whole is related to the difficulty of the decision-making process, personality, travel behavioral patterns and the degree of agreement/disagreement among the group members during the decision-making process.
- Finally, groups adopted various decision-making approaches, which mainly differed in (a) preferences disclosure technique; (b) discussion type; and (c) the decision reaching technique.

6 Implications for recommender systems

As previously mentioned, the proposed observational study is ultimately motivated by the goal of being able to design more effective GRSs. This means that the system should better predict, and therefore recommend, which items will make the group members more satisfied. We will now discuss some important benefits that the analysis of the data acquired by observing users' interactions in group decision-making tasks can bring to recommender systems, and we will also illustrate some already achieved results.

First of all, GRSs require the design of ranking functions that can highlight which items a group must primarily look at. Ranking functions for GRSs are based on preference aggregation strategies. While we already mentioned that there is not a single best aggregation strategy that fits all possible recommendation tasks and decision contexts, observational study data can be used to choose and customize the aggregation function to the specific contextual conditions of the group. We conjecture that, having a family of candidate aggregation functions, one can optimally choose the right one by fitting the observation data. For instance, experimental results of the study showed that the social role and personality of the group members influence group choices which was also confirmed in other studies (Gartrell et al. 2010; Quijano-Sanchez et al. 2013; Recio-Garcia et al. 2009). Hence, for instance, among a family of multiplicative aggregation models one can fit the importance weights of the group members depending on their roles and personality.

This conjecture is furthermore addressed by a recent simulation study (Nguyen and Ricci 2017b) analyzing how long-term and session-specific preferences can be optimally combined in different group scenarios. It is observed that a combination strategy that weighs more the long-term preferences is fitted to the scenarios when the group setting has no impact on group members' preferences, but when the group context pushes users to be either cooperative or uncooperative, users seem to benefit more from a recommender that takes into account the preferences observed from the group discussion, which reflect their newly emerging interests.

A second important usage of observational data is the construction of a more dynamic model of recommendations that integrates preference information derived by the observation of the discussion process into the baseline user preference model. In fact, it is clear from our study that the final group choice is not completely determined by the initial preferences of the users, i.e., the preferences expressed while evaluating domain items without any reference to or influence of the group. We conjecture that the observed dynamics of within group interactions must be carefully considered in order to better predict which items may suit the group at the precise point in time when the discussion in the group takes place. We have, for instance, mentioned the observed correlation between the decision-maker's activity in providing information or criticizing options and the choice satisfaction. As we suggested in the paragraph above, this data can also be used to identify a better aggregation strategy. However, we also hypothesize that this type of information can be exploited to revise the initial user models learned by the system using the historical preference data of the users. For instance, if a content based model was fitted to the known ratings of a user, this model can then be revised by considering the items that the user liked or criticized during the group discussion. Clearly, performing observations within the system is a much simpler task than conducting an observational study with human observers. The system could easily track decision-makers' reactions to each-others' proposals and system-generated recommendations. However, even though the classification of decision-makers' behavior might be a harder task for a system, it is certainly possible to introduce and detect a set of basic behavior categories. Moreover, in this study we aim at learning which behavioral aspects play an important role for the group decision-making outcomes, i.e., the group choice and the choice satisfaction, and as the exploratory analysis has shown, not all the categories seem to be of critical importance.

This idea has been implemented in a mobile system called STSGroup (Nguyen and Ricci 2017c). The system allows group members to be engaged in a group discussion where they can exchange messages together with proposing items that are thought to be suitable for their group and react to other group members' proposals by giving feedback such as likes, dislikes or best-choice (see Fig. 3a). The interactions between the members and the system during the group discussion are monitored and taken into account in order to provide appropriate recommendations and choice suggestions for group members (see Fig. 3b, c). The group recommendations are accompanied by explanations that are computed on the base of the group members' actions and contexts. Hence, this system builds up on the observational study, and it convincingly demonstrates (1) the importance of the study scope and

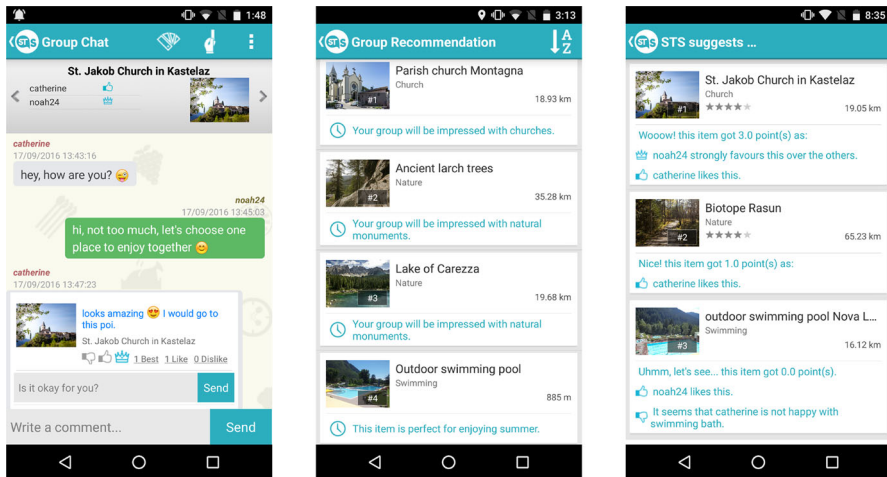


Fig. 3 Screen-shots of STSGROUP, from left to right: **a** group discussion, **b** group recommendations, and **c** choice suggestions

focus in the area of GRSs; and (2) why the research in the area of GRSs needs more similar studies that better tackle into the behavior of users and not only preferences.

A live user study was conducted to assess the usability of STSGROUP, the perceived quality of the group recommendations and the choice satisfaction (i.e., the satisfaction of the users with the item that was finally selected by the group for a visit). The results of the user study has shown that the usability of the system is superior to a standard benchmark. Particularly, most of the participants indicated that the system is not complex and it is easy to use. It also leads to high perceived recommendation quality and choice satisfaction. This conclusion was supported by the fact that more than 70% of the participants confirmed that they found the new item recommendations for a group relevant, and even though only 60% of the participants thought that the chosen place fits their preference, more than 85% of the participants indicated that they were excited about the group choice.

Moreover, the information observed and collected during the group interaction, such as, the duration of the discussion and how much users interact with each other, can be further exploited to assess the “situation” that each individual member is likely to experience in the group setting. In fact, there are several different kinds of social response to group pressures Forsyth (2014). For example, group members may be consistent with their personal standards, or show conformity to the group opinion, or alternatively react negatively to the group setting. The “situation assessment” is essential since for different group settings the trade-off between long-term and session-based preferences has to be fine tuned in order to quickly meet the users’ needs and requirement. This hypothesis has been confirmed through a recent follow-up simulation experiment that was conducted by reusing STSGROUP data and its group recommendation model. Moreover, evaluating the group situation will pave the way for making a GRS proactive. More concretely, based on the estimated circumstance, the recommender can automatically adapt and better

choose its actions, e.g., giving group recommendations, acquiring more information or suggesting a final choice to support the group decision-making process. In fact, adaptive action selection was successfully introduced and employed in a conversational travel recommender system for individuals (Mahmood et al. 2009), and we believe that such an approach can bring even a greater benefit in a group decision support system.

A fourth, probably most fundamental issue, is related to the ultimate goals of observational data and the scope of a GRS. Should the recommender fit the data, i.e., suggests what the users in a given context are supposed to choose, or should the system act as a mediator instead, aimed at driving the group towards a more fair choice? In the first case, as illustrated in the two paragraphs above, the system pleases the group and let it more smoothly and efficiently converge towards the decision that the group may have taken even without the system intervention. In the second case, the system is instead assuming that the fairness of a sound aggregation strategy should prevail on the natural group dynamics and will stick to it. This contra-position is not new in recommender systems: it relates to the question whether a recommender should only suggest items predicted to be top choices for the user or inject in the recommendations items that would make the list of recommendations more diverse, novel, sustainable, or simply more trendy. In order to address these fundamental questions, and understand which role the recommender should play, live user studies are unavoidable.

A fifth implication of the study is related to the picture-based approach introduced in Neidhardt et al. (2014, 2015). The pre-survey questionnaire and the picture-based approach aim at capturing a user model described in terms of the same 17 tourist roles and Big Five factors that we used in the observational study described by this article. The picture-based approach uses the 17 tourist roles and the Big Five factors to extract, in a lower dimensional space, seven factors that describe tourist behavioral patterns: *Sun and Chill-out*, *Knowledge and Travel*, *Independence and History*, *Culture and Indulgence*, *Social and Sport*, *Action and Fun*, *Nature and Recreation*. But, to avoid long and tedious questionnaires to capture user's preferences, the authors use pictures. For each of the seven factors, pictures were identified and user preferences were captured by prompting the user to select pictures from this predefined set. By mapping the selected pictures onto the seven factors, a score for each of the factors can be determined for the user. Also, points of interest (POIs) can be represented using the same seven factors, so the recommendations for a user can be calculated by using the Euclidean distance between his/her user profile and the POIs. Figure 4 illustrates the picture selection environment and the travel profile feedback. Actually, the findings of this observational study can be related to the picture-based approach model and then generalized to design a GRS. The proposed research and related challenges are described in Delic (2016).

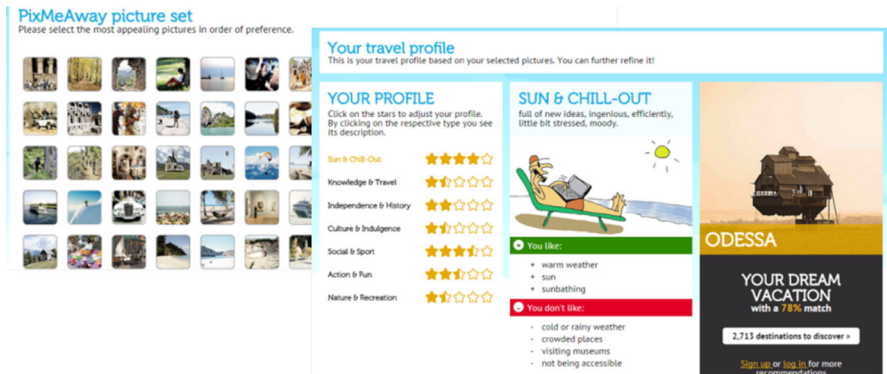


Fig. 4 Screen-shots of the picture-based recommendation engine PixMeAway

7 Conclusion and outlook

In this section we summarize the contributions of the paper and list some further challenges that we believe could be addressed in future analysis of the data we have already collected. Furthermore, we discuss potential variations and generalizations of the proposed observational study.

In summary, the main contributions of the paper are:

- A detailed description of a replicable study procedure and instruments used for the data collection that can provide insights into group decision-making processes (see Sects. 3 and 4).
- The implementation of the proposed study procedure in a concrete context of tourism and traveling group decision making.
- Experimental results showing that certain characteristics of individuals as well as groups, which go beyond group members' individual preferences and their straightforward aggregation, play an important role in the final group choice (see Sect. 5).
- Implications of the performed observational study to the design of GRSs and the identification of aspects that should be considered when building such systems (Sect. 6).

During the initial data analysis, we encountered several challenges related to data measurements. We believe that these challenges can be better addressed in future work:

- How to aggregate different individual scores, e.g., personality traits, at the group level?
- How to measure diversity among group members with respect to the different data dimensions?
- How to distinguish satisfied from not so satisfied groups?
- How to match and compare individual preferences to the preferences of the group as a whole?

- How to address ratings/ranking difference in different study implementations?
- How to relate participants' personalities to their preferences?

So far, we were mainly using the average of the individual scores when aggregating them at the group level (Delic et al. 2016b). However, more diverse approaches should be applied and compared in future work.

Different dimensions of the study procedure can be varied in order to grasp insights into the group dynamics in this particular context. In the following we present some of the variations and their potential implications:

- *Duration and timing of the study* In our implementations, we noticed different behaviors of the subjects when comparing the results of study conducted over the three weeks period with that conducted in one lecture session. In the first case students were not explicitly referring to their initial, individual preferences during the group discussion, but they were rather discussing their preferences in general. In the second case, students were comparing their initial preferences during the group discussion, therefore their final choice was usually based on these comparisons.
- *Diversity of the ten predefined destinations* (e.g., country side tourism vs. big city tourism; mountain destination vs. sea side destination; hot climate destination vs. cold climate destination): higher diversity could generate more conflicting preferences in groups and more intense discussions and decision processes.
- *Nature of the ten predefined destinations* In our study, the ten pre-selected destinations were all European capitals (except Amsterdam). Clearly, the participants were well informed about the destination set at hand. Therefore, the question is raised whether or not the usage of less known destination set and the participants' lack of knowledge would influence the decision-making process and if yes in which way.
- *Group size* The conducted data analysis showed differences in groups' satisfaction with respect to the group size—smaller groups tend to be more satisfied with the group choice than the larger groups, which is quite intuitive. Nevertheless, varying the group size in the study can provide insights on different aspects that should be considered.
- *Group diversity* We conjecture that controlling the diversity of the group with respect to the preferences as well as the personality could reveal novel and interesting characteristics of the group decision-making processes, and therefore, can lead to the design of better methods for supporting groups in action.
- *Budget* Including budget into the group discussion increases the complexity of the task for the participants and it also enables more realistic setting of the decision process in the context of traveling.
- *Group decision task* If the group were to choose a point of interest that they *actually* had to visit together right after the group discussion, then the group members might pursue their preferences and interests in a more natural manner and more persistently.

- *Domain* The same study could be carried out in a different domain, such as music, movies, restaurant, etc. In some cases it could be easier to introduce a more realistic setting to participants, but the discussion process, could evolve in a different way.

Finally, other types of analyses can be conducted making use of the rich information that has been collected so far (see Sect. 4), such as, (1) identifying sources of influence in the group decision-making process, (2) analyzing different approaches that groups employed in order to reach their final decisions and relating those approaches to the satisfaction of group members, (3) identifying characteristics of groups that could determine the best preferences aggregation strategy to be applied, etc. Clearly such analyses would provide great insights for the future designers of GRSs.

Acknowledgements Open access funding provided by TU Wien (TUW).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Ali I, Kim SW (2015) Group recommendations: approaches and evaluation. In: Proceedings of the 9th international conference on ubiquitous information management and communication. ACM, p 105
- Ardissone L, Goy A, Petrone G, Segnan M, Torasso P (2003) Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices. *Appl Artif Intell* 17(8–9):687–714
- Bales RF (1950) A set of categories for the analysis of small group interaction. *Am Sociol Rev* 15:257–263
- Bekkerman P, Kraus S, Ricci F (2006) Applying cooperative negotiation methodology to group recommendation problem. In: Proceedings of workshop on recommender systems in 17th European conference on artificial intelligence (ECAI'06), pp 72–75
- Bell A (2008) Predicting interpersonal conflict resolution styles from personality characteristics. *Personal Individ Differ* 45(2):126–131
- Berkovsky S, Freyne J (2010) Group-based recipe recommendations: analysis of data aggregation strategies. In: Proceedings of the 4th ACM conference on recommender systems, pp 111–118
- Blanco H, Ricci F (2013) Inferring user utility for query revision recommendation. In: Proceedings of the 28th annual ACM symposium on applied computing, SAC '13, Coimbra, Portugal, March 18–22, 2013, pages 245–252
- Chen L, de Gemmis M, Felfernig A, Lops P, Ricci F, Semeraro G (2013) Human decision making and recommender systems. *ACM Trans Interact Intell Syst* 3(3):17
- Delic A (2016) Picture-based approach to group recommender systems in e-tourism domain. In: Conference proceedings of the 24th conference on user modeling, adaptation, and personalization (UMAP 2016), Halifax, Canada
- Delic A, Neidhardt J (2017) A comprehensive approach to group recommendations in the travel and tourism domain. In: Adjunct publication of the 25th conference on user modeling, adaptation and personalization. ACM, pp 11–16
- Delic A, Neidhardt J, Nguyen TN, Ricci F (2016a) Research methods for group recommender systems. CEUR-WS

- Delic A, Neidhardt J, Nguyen TN, Ricci F, Rook L, Werthner H, Zanker M (2016b) Observing group decision making processes. In: Proceedings of the tenth ACM conference on recommender systems, RecSys'16
- Delic A, Neidhardt J, Werthner H (2016c) Are sun lovers nervous?—Research note at enter 2016 eTourism conference. Bilbao, Spain
- Delic A, Neidhardt J, Rook L, Werthner H, Zanker M (2017) Researching individual satisfaction with group decisions in tourism: experimental evidence. In: Information and communication technologies in tourism 2017. Springer, pp 73–85
- Donnellan MB, Oswald FL, Baird BM, Lucas RE (2006) The mini-ipp scales: tiny-yet-effective measures of the big five factors of personality. *Psychol Assess* 18(2):192
- Fernández-Tobías I, Brauhofner M, Elahi M, Ricci F, Cantador I (2016) Alleviating the new user problem in collaborative filtering by exploiting personality information. *User Model User Adapt Interact* 26(2–3):221–255
- Fesenmaier DR, Kuflik T, Neidhardt J (2016) Rectour 2016: workshop on recommenders in tourism. In: Proceedings of the 10th ACM conference on recommender systems. ACM, pp 417–418
- Fiske ST, Gilbert DT, Lindzey G (2010) Handbook of social psychology, vol 2. Wiley
- Forsyth DR (2014) Group dynamics, 6th edn. Cengage Learning
- Gartrell M, Xing X, Lv Q, Beach A, Han R, Mishra S, Seada K (2010) Enhancing group recommendation by incorporating social relationship interactions. In: Proceedings of the 16th ACM international conference on supporting group work. ACM, pp 97–106
- Gibson H, Yiannakis A (2002) Tourist roles: needs and the lifecycle. *Ann Tour Res* 29(2):358–383
- Guzzi F, Ricci F, Burke R (2011) Interactive multi-party critiquing for group recommendation. In: Proceedings of the 5th ACM conference on recommender systems, pp 265–268
- Jameson A (2004) More than the sum of its members: challenges for group recommender systems. In: Proceedings of the working conference on advanced visual interfaces, pp 48–54
- Jameson A, Smyth B (2007) Recommendation to groups. In: The adaptive web. Springer, Berlin, Heidelberg, pp 596–627
- Kilman RH, Thomas KW (1977) Developing a forced-choice measure of conflict-handling behavior: the “mode” instrument. *Educ Psychol Meas* 37(2):309–325
- Mahmood T, Ricci F, Venturini A (2009) Learning adaptive recommendation strategies for online travel planning. *Inf Commun Technol Tour* 2009:149–160
- Masthoff J (2004) Group modeling: selecting a sequence of television items to suit a group of viewers. In: Personalized digital television. Springer, Dordrecht, pp 93–141
- Masthoff J (2015) Group recommender systems: aggregation, satisfaction and group attributes. In: Ricci F, Rokach L, Shapira B (eds) Recommender systems handbook. Springer, Boston, MA., pp 743–776
- Masthoff J, Gatt A (2006) In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems. *User Model User-Adap Inter* 16(3–4):281–319
- McCarthy K, McGinty L, Smyth B, Salamo M (2006) The needs of the many: a case-based group recommender system. In European conference on case-based reasoning. Springer, Berlin, Heidelberg, pp 196–210
- McCrae RR, Costa PT (1987) Validation of the five-factor model of personality across instruments and observers. *J Personal Soc Psychol* 52(1):81
- McCrae RR, John OP (1992) An introduction to the five-factor model and its applications. *J Personal* 60(2):175–215
- Neidhardt J, Schuster R, Seyfang L, Werthner H (2014) Eliciting the users’ unknown preferences. In: Proceedings of the 8th ACM conference on recommender systems. ACM, pp 309–312 (2645767)
- Neidhardt J, Seyfang L, Schuster R, Werthner H (2015) A picture-based approach to recommender systems. *Inf Technol Tour* 15(1):49–69
- Nguyen TN (2017) Conversational group recommender systems. In: Proceedings of the 25th conference on user modeling, adaptation and personalization. ACM, pp 331–334
- Nguyen TN, Ricci F (2016) Supporting group decision making with recommendations and explanations. In: Posters, demos, late-breaking results and workshop Proceedings of the 24th conference on user modeling, adaptation, and personalization (UMAP 2016), Halifax, Canada
- Nguyen TN, Ricci F (2017a) A chat-based group recommender system for tourism. In: Information and communication technologies in tourism 2017. Springer, Cham, pp 17–30
- Nguyen TN, Ricci F (2017b) Combining long-term and discussion-generated preferences in group recommendations. In: Proceedings of the 25th conference on user modeling, adaptation and personalization. ACM, pp 377–378

- Nguyen TN, Ricci F (2017c) Dynamic elicitation of user preferences in a chat-based group recommender system. In: Proceedings of the 32nd ACM symposium on applied computing, pp 1685–1692
- Quijano-Sanchez L, Recio-Garcia JA, Diaz-Agudo B, Jimenez-Diaz G (2013) Social factors in group recommender systems. *ACM Trans Intell Syst Technol* 4(1):8
- Quintarelli E, Rabosio E, Tanca L (2016) Recommending new items to ephemeral groups using contextual user influence. In: Proceedings of the 10th ACM conference on recommender systems. ACM, pp 285–292
- Recio-Garcia JA, Jimenez-Diaz G, Sanchez-Ruiz AA, Diaz-Agudo B (2009) Personality aware recommendations to groups. In: Proceedings of the 3rd ACM conference on recommender systems, NY, USA, pp 325–328
- Ricci F, Rokach L, Shapira B (2015) Recommender systems: introduction and challenges. In *Recommender systems handbook*, 2nd edn. Springer, Boston, pp 1–34
- Stettinger M, Felfernig A, Leitner G, Reiterer S, Jeran M (2015) Counteracting serial position effects in the choicla group decision support environment. In: Proceedings of the 20th international conference on intelligent user interfaces, GA, USA, pp 148–157
- Tajfel H (2010) Social identity and intergroup relations. Cambridge University Press
- Tindale RS, Kameda T (2000) Social sharedness as a unifying theme for information processing in groups. *Group Process Intergroup Relat* 3(2):123–140
- Venturini A, Ricci F (2006) Applying trip@dvce recommendation technology to <http://www.visiteurope.com>. In: ECAI 2006, 17th European conference on artificial intelligence, August 29–September 1, 2006, Riva del Garda, Italy, including prestigious applications of intelligent systems (PAIS 2006), Proceedings, pp 607–611
- Werthner H, Alzua-Sorzabal A, Cantoni L, Dickinger A, Gretzel U, Jannach D, Neidhardt J, Pröll B, Ricci F, Scaglione M, Stangl B, Stock O, Zanker M (2015) Future research issues in IT and tourism. *J Inf Technol Tour* 15(1):1–15
- Werthner H, Ricci F (2004) E-commerce and tourism. *Commun ACM* 47(12):101–105
- Yiannakis A, Gibson H (1992) Roles tourists play. *Ann Tour Res* 19(2):287–303