# Measuring Social Influence Using Eye-Tracking in a Between-Groups User-Study

GEORGE POPESCU, École Polytechnique Fédérale de Lausanne

The decision success of a group is strongly related to its members' satisfaction while adapting to common outcomes that, in turn, relate to familiarity and trust. Thus, the evaluation of determinant factors enhancing group dynamics is essential for understanding social influence in a group setting. This paper presents the design and results of a group influence study in a music scenario, in which people share their tastes and participate in-group decisions. We designed and ran a between-groups experiment with the purpose of tracking group influence in our GroupFun music recommender system. Specifically, we were interested in measuring the extent to which people change their opinions when facing various individual preferences supported by other group members. Two groups of 9 and 10 members are evaluated one against the other by means of comparing decision acceptance, participant satisfaction and decision changes using an online questionnaire, songs' ratings and eye-tracking data. The eye tracking outputs we used present clear associations between various areas of the interest while fixation times are useful to understand group influence relative to closest members. Our results focus on group conformity, or the degree to which individuals change their own rating aligning to the group decision. We prove that the stronger social relationships (through familiarity and trust) are in a group, the more its members adapt to a common decision and are more inclined to change their preferences.

Categories and Subject Descriptors: **H.5.2 [Information Interfaces and Presentation (e.g. HCI)]**: User Interfaces – Evaluation/methodology, theory and methods, user-centered design.

General Terms: Design, Experimentation, Human Factors, Measurement.

Additional Key Words and Phrases: Eye-gaze, Eye-tracking, Familiarity, Group, Interface, Music, Rating Change, Recommender System, Social Influence, Trust, User Study.

## 1. INTRODUCTION

Imagine you and your friends have to decide the location for your upcoming holidays, new activities that you would like to plan together in the vicinity of your city, or the playlist songs for a party you are all attending. How would you participate? Would your preferences be strict or would you adapt to the group decision? Which members do you know best? Which members do or would you trust most? For which of them do you already know or could guess their music preferences? Would you enjoy discovering such tastes? A group recommender system helps your group make a qualitative decision together!

Whether selecting the ideal holiday trip from an overwhelming list of destinations, choosing a best movie to watch with your friends in the evening or deciding upon a collection of entertaining music for carpooling every day to work, individual satisfaction is essential for accepting group decisions. Preference aggregation systems are based on algorithms which aim at maximizing the overall group satisfaction [Agotnes et al. 2007; Bonhard et al. 2006; Bu et al. 2010; Grossi 2009; Popescu and Pu 2011; Popescu and Pu 2012] and/or provide intermediate suggestions that the group would agree or improve upon [Baskin and Krishnamurthi 2009; Chen and Chen 2001; Gartrell et al. 2010; Masthoff 2005]. The state-of-the-art scientific

Author's address: G. Popescu, LAMS, BC167, Computer Science Department, École Polytechnique Fédérale de Lausanne.



DOI10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

literature in the group decision field has pursued two research directions: one that deals with group dynamics [Kang et al. 2007; Postmes et al. 2001; Sassenberg 2002] and the other that analyses decision accuracy and user acceptance [Bonhard et al. 2006; Celma and Cano 2008; Masthoff 2005; Pu et al. 2011; Purpura et al. 2011].

In recent academic publications researchers have used tools to describe and understand decision behavior in both online and off-line environments for specific domains such as apartment search, music, games, TV series, etc. Results showed that group dynamics are correlated with individual and group performance [Prada et al. 2007]. In addition, with the emergence of social networks, various visualization techniques are able to capture temporal group interactions [Kang et al. 2007].

Unfortunately, little is known about how the group influences the individual (and vice-versa) in online systems and how it contributes to individual satisfaction and decision change. Large-scale social network studies have the advantage of presenting an overview of social dynamics but are limited by an in-depth comprehension of factors that determine decision change at the individual level [Tang et al. 2009]. Nevertheless, trust-based group models provide a reliable framework for understanding individual decision-making [Fleishman 1998; van Maanen et al. 2007].

A group recommender system (GRS) is a recommender system aimed at generating a set of recommendations that will satisfy a group of users, with potentially competing interests. Traditionally, GRSs have evolved by developing individual preference specification models that correspond to users interests [Bonhard et al. 2006; Bu et al. 2010; Chen and Chen 2001; Pu et al. 2011]. These preferences might change significantly over time and are context dependent: e.g. a group of friends would like to listen to one type of music for Saturday night's party and a different type while carpooling to work together on Monday morning. In addition, preferences may vary according to the group's characteristics: one person may share certain music preferences while with closest friends and/or family and different ones while with strangers. Thus, GRSs propose new algorithmic and design approaches by: incorporating social relationship interactions [Gartrell et al. 2010] and combining user and item information together [Bonhard et al. 2006; Bu et al. 2010]. Personalized recommendations have the key advantage of removing popular, uninteresting, recommendations, favoring the discovery of new items [Celma and Cano 2008; Popescu and Pu 2011].

Despite the fact that algorithms for GRSs are quite well adapted for increased accuracy, precise user modeling and serendipitous recommendations [Bonhard et al. 2006; Bu et al. 2010; Grossi 2009], one significant limitation in this field is that such algorithms demand an extensive user effort and produce only a small increase in user acceptance [Pu et al. 2011; Popescu and Pu 2011]. An alternative approach to this type of research envisions the "pursuit of satisfaction". This justifies the role of an affective state in decision-making and individual satisfaction [Masthoff 2005].

Nowadays, eye-tracking devices represent a precise and reliable technology for measuring users' eye movements when facing a digital interface. Eye-tracking technology has successfully proved its potential for understanding individual decision making, for example when people look at various items with the aim of identifying objectively and indirectly the features that are of interest for an individual and generalize for a target customer segment. However, such technology was not previously used for analyzing group influence, but only in online marketing related studies for which retailers assess their website features by capturing their potential visitors' interest for a specific item and/or recommendation technique through eye-gaze [Castagnos et al. 2009; Castagnos et al. 2010; Castagnos and Pu 2010; Chen and Pu 2011]. Finally, social influence analysis was addressed in large-scale online networks with the purpose of identifying the most influential person(s) and understanding the influent-follower role [Cialdini and Trost 1998; Fleishman 1998; Tang et al. 2009]. It was previously reported that people switch their preference when receiving recommendation from friends or people they know well, thus conforming to the group choice [Turner 1991; van Maanen 2007; Zhu et al. 2012].

One of the closest related research works to the current paper is a study that explores personal impact aimed to enhance group recommendation [Liu et al. 2012]. The authors present an overview of the group recommendation context showing how recommendation to a group of people can be made using social networking services. A key aspect is the analysis of the group decision-making process based on which the authors propose a personal impact topic (PIT) model, which serves to generate group recommendations. In the model, at first an algorithm creates a group preference profile for any group, based on two factors: personal preferences of members and their personal impacts. Personal impact is based on social network information and serves as fundament for the PIT model. Using this methodology, the authors show that their group recommendation technique performs better than similar approaches. The research idea of exploring social influence for item recommendation represents an interesting approach showing the potential of enhanced and more accurate recommendation techniques compared with previous approaches [Ye et al. 2012]. The probabilistic generative model naturally unifies social influence, collaborative filtering and content-based methods. This unified model is based on algorithms that learn parameters to describe hidden social influence. One of the most significant contributions of the study [Ye et al. 2012] is the model validation using data from last.fm and whrrl.com. The results show that the enhanced model with social influence performs significantly better that state-of-the-art group recommendation algorithms without social influence. Moreover, users from whrrl.com are more likely to be influenced by their friends than those form last.fm.

People's opinions can often be influenced or changed by others. In a social influence experiment aimed at measuring social influence in online recommender systems [Zhu et al. 2012], the authors measure the frequency of decision swapping grounded on different levels of acceptance and conformity pressure. The experiment design first asks the participants to provide personal preferences between pairs of items without any social influence or knowledge of others' tastes. At a later stage they make a decision about the same pairs with the knowledge of other's preferences. The general result proves that other people's opinions influence one's own and, consequently, it contributes to the swap of individuals' previous choice. The longer the time between the two decisions the more people are inclined to change: 22.4% changed their decision at a later moment compared with 14.1% immediately. The number of influential opinions is another important factor for personal decision-making: people are more likely to change when facing a moderate rather than large number of opposing preferences. One of the main differences between the experiment design of this study and the one presented hereafter is that we consider the influence of closest people rather than online opinions of others/strangers. Thus, the personal connection is a key aspect in our setting.

The human decision mechanism represents a complex system and is influenced by numerous factors that we aim to identify through a between-groups user study. More precisely, we investigate the topic of group influence in a music context for the members of two groups who select their desired songs for a party they plan and attend together. This issue corresponds to the preference aggregation problem [Agotnes et al. 2007; Baskin and Krishnamurti 2009; Grossi 2009] in which multiple tastes are combined by a system that recommends best options to the whole group. We designed a group influence experiment in which we aimed at quantifying people's decision making through: (1) the GroupFun group music recommender system, in which participants submit individual music preferences, and (2) an eye-tracking-based system including a set of interfaces, for which subjects are exposed to other group members' preferences and are asked to make a rating decision. As mentioned before, the group decision-making area of research lacks experiments and analysis of social influence sustained by both rating/preference submission and eye-tracking information. Therefore, correlating subjective group factors, such as familiarity and trust, with participants' objective eye-gaze, yield interesting results for both the social science and group decision research fields. In the present study we draw correlations between the two groups' connected networks' properties and use similarity and group influence metrics to interpret individual decision change.

Our results show that social influence is stronger in groups in which participants are most familiar with and trust their peers' tastes. We found trust to be a better measure than familiarity for predicting people eye-gaze behavior. Most importantly, group conformity is correlated with group familiarity and trust. Thus, the experiment sustains social conformity: people align their preferences to those of the group.

The specific contributions of this paper are:

(1) The design of a between-groups experiment combining subjective information consisting in familiarity and trust values with objective observations extracted from eye-gaze correspondences and focus times. In addition, we include the main features of our group music recommender system that participants used to contribute music to their group and submit individual ratings.
(2) A comparison between the two groups' characteristics. We show that for the more connected a group, i.e. with denser familiarity and trust networks, the individual rating change is higher. In other words people adapt more to the group decision by changing their own evaluation when social bounds are stronger. We present an extensive group satisfaction comparison as well as rating change statistics together with preference correlation values.
(3) The analysis of the overlay between the subjective values from each group, computed as normalized familiarity and trust, and the objective eye-gaze correspondences per participant, aimed at highlighting the most important members that influence the current participant's decision. We map the two values one to another and report that the eye-tracking data represents a reliable source of information for constructing a group's social network with high accuracy. By opposition, the music correlation values among participants do not represent a reliable source for deriving social bonds. As a consequence, the objective data gathered from the eye-tracker support the social alignment results, which are based on familiarity and trust subjective information.

The remainder of this report is structured as follows. First, we describe the experiment setup in detail presenting all structural elements. We continue by emphasizing the 3 building blocks of our study: an online questionnaire, the GroupFun system and the eye-tracking experiment introducing the two groups, main user tasks, the objectives of the study and the dataset for our analysis. We next present extensively our main results based on rating (preference correlation, change rate), eye-tracking data (eye-gaze patterns) and a computed group influence score (with respect to group size). We then discuss the major research implications focusing on certain advantages and limitations of our study design. Finally, we summarize our ideas through the conclusions and define future work.

## 2. EXPERIMENT DESIGN

The user study design is based on the following preference elicitation scenario: submission of individual ratings first without and then with others' ratings while listening to music alone in front of a computer. The main goal is to collect useful subjective and objective information relevant for social influence analysis. Both the GroupFun system and the eye-tracking interfaces were conceptualized to be easy to use and facilitate participants' understanding of the information displayed to them. Experiments are firstly performed as a written questionnaire to determine the social relationship through familiarity and trust, as well as eye-tracking to determine various areas of interest and fixation times on these areas, which can help understand group influence relative to closest members.

### 2.1 Participants

19 subjects belonging to 2 groups where recruited to participate in the group influence experiment. Their ages ranged from 23 to 30. Participants were mainly university students and employees. They had 12 different national backgrounds: Switzerland, France, China, India, Russia, Italy, Romania, etc. and 3 different educational backgrounds: Master, PhD and Post-Doc. 9 participants formed Group1 and 10 formed Group2. In the first group there were only male participants whereas in the second there were 6 male and 4 female members. One important aspect for the group structure is that members in each group knew all other members but not the members from the other group. More specifically, Group1 members were students attending the same course whereas Group2 members were Master, PhD and Postdoc students working on their research in two closely related labs. Subjects only interacted with the other peers in their own group, but not the other.

The study was carried out over a period of 3 weeks. In the first week the two groups used the GroupFun music recommender system (to be described later) to create a group, upload music and give ratings individually, without any group interface and influence. During weeks 2 and 3, eye-tracking experiments were carried out with the purpose of recording users' eye-gaze and analyze their newly updated ratings by comparing them with previous ones. In week 2, all members of Group1 performed the eye-tracking phase of the experiment whereas in week 3 all members of Group2 did the same task. In total we collected more than 4 hours of continuous eye-tracking gaze data and 399 ratings for 86 songs.

The study proposed a music incentive. The reward that all subjects received at the end of the experiment was a collection of all group songs uploaded to their group. We created an archive and sent the music items to all members of the same group. Group1 received 39 contributed songs and Group2 47. Besides the specific songs, we also shared similarity music profiles and rating correlation values among members.

### 2.2 User Tasks

Our experiment's main user task was to "choose songs for a party that all group members would organize together".

The steps that each participant chronologically pursued are detailed in Fig. 1.

Subjects were asked to perform the 3 following tasks:

(1) Complete an online questionnaire evaluating familiarity and trust with respect to other group members (Fig. 1 top - Phase I). For assessing familiarity and trust subjects used ratings from "1" to "5" for each other group members.
(2) Use GroupFun to create a group, upload music and give individual ratings to group songs (Fig. 1 middle - Phase I). Users logged in to GroupFun and joined

the virtual group corresponding to the real one. Then, each uploaded several songs that the system displayed in an individual preference elicitation interface. Using this interface users were able to listen to the songs, fast forward, pause and give a rating from “1” to “5” to any number of contributed songs in the group.

(3) Listen and give ratings to a selection of 24 group songs while watching group interfaces containing other group members’ names and their ratings, being recorded by an eye-tracking device (Fig. 1 bottom - Phase II in red). All 24 songs were extracted from the total group contribution.

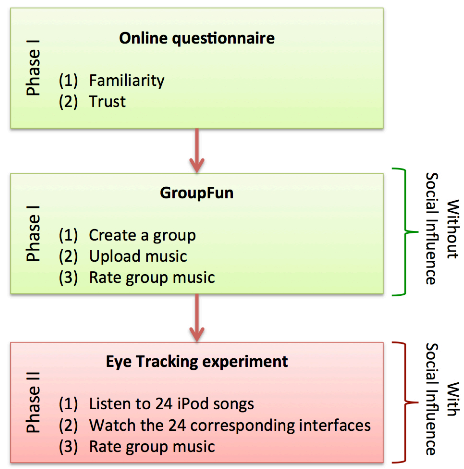


Fig. 1. Experiment setup and phases

### 2.3 Objectives

In a social context people adapt their decision to that of the group. We expect our volunteers to change decision given their group orientation rather that keep strict ratings corresponding to their previously stated preferences. Our main objective is to analyze group influence by comparing the two groups of users in order to understand the factors that produce social alignment. Furthermore, we draw correlations between the implicit familiarity and trust network information and explicit eye-tracking data.

Given the specific characteristics of each group we aim at analyzing how the familiarity and trust networks influence each group’s preference dynamics. More specifically we analyze group satisfaction before and after the eye tracking session and draw correlations among music tastes. Our hypothesis is that higher group connectivity yields greater social influence and rating changes among group members. Conversely, the more distant group members are, the more selfish they behave and don’t change their ratings reaching a similar satisfaction under both the individual setting and the group influence conditions.

We also compute rating correlations between members of the same group in order to determine closest members preference-wise and compare this objective metric with the subjective evaluation given by our users after answering the familiarity and trust online questionnaire. We next correlate social bounds between group members with data obtained from the eye-tracking device to understand specific group factors that model individual decision. At a more detailed level, our main goals are to: (1) understand how the group characteristics model individual decision making and (2) compare members of two different groups among them-selves in terms of familiarity, trust, preference, group satisfaction, eye-gaze and rating change.

### 2.4 Materials

During Phase I (Fig. 1, top) volunteers used an online form to fill in their responses. The link to the form was provided by e-mail so that respondents could fill in their answers privately at any time. Then, they proceeded to use the GroupFun music recommender system[1] for creating a group, uploading music and contributing to the group decision via an individual evaluation of group songs through ratings.

In addition to this, we used the Tobii 1750 eye-tracking device[2] consisting of a monitor with embedded infrared cameras to capture pupil movement. We attached a desktop computer running the ClearView 2.7.1 software capable of capturing users' eye gaze, display fixation points and generate heat-maps. After an initial calibration for each participant, we asked our users to look at the screen in a comfortable way with the help of a head-mount.

The second device for our experiment was an Apple iPod Shuffle on which we recorded 24 recommended songs were played synchronously according to the 24 interfaces displayed on the eye tracker. Members of each group had the same 24-songs playlist recorded in the iPod but songs in Group1 and in Group2 were different.

## 3. ONLINE QUESTIONNAIRE

Before the experiment we asked our participants to fill out an assessment questionnaire. Two questions were used to assess familiarity and trust information: (1) "How much do you know your friends' musical tastes?" and (2): "How much do you trust your friends (music-wise)?" Both questions were mandatory and each participant submitted a rating value for each of his peers. The first one assesses self-revealed familiarity whereas the second one self-revealed trust on a 5 point Likert scale. Through the second question we asked subjects to rate how much they trust each other based on music, while expecting them to know very little about each other's music tastes. This kind of "blind" trust is different from the "informed" trust in trust-based recommender systems.

The completion time of the questionnaire was around 2 minutes / participant. The role of the two above questions is double-folded. First, we explore the groups' internal subjective network structure (based on familiarity and trust) as seen through the eyes of its members. Secondly, we map this subjective information with eye-tracking data to analyze the extent to which these familiarity and trust values correspond to participants' actual behavior. As such, we compare subjective and objective parameters that model group behavior yielding results that are useful for the understanding of group decision-making and social influence.

The directed graphs (Fig. 2) show the familiarity (top) and trust (bottom) networks of the two groups: Group1 ("Cool party") to the left (P1 to P9) and Group2 ("Legendary party") to the right (P10 to P19). The directed arrows show the answer to each question given by Pi to Pj for familiarity and trust, respectively. A thick arrow shows a high score (of 5). We note from the upper-left graph that participants P1 and P4 are not familiar with other members' tastes. Conversely, P2 stated that he is most familiar with the majority of other members' tastes. For example, he is very familiar with P3, P6 and P7's music preferences (thickest arrows). The relationship is reciprocal for P3 (two-ways oriented arrow) but not for P6 and P7 (one-way). By opposition, thinnest one-way or two-ways oriented arrows indicate smallest ratings (scores of 1).

[1] http://grpupc1.epfl.ch/~george/GroupFun/
[2] http://grpupc1.epfl.ch/~george/Tobii1750.gif

Overall, the two familiarity graphs (Fig. 2, top) show that members who are at the exterior are most familiar with their peers' music tastes. Reciprocally, the participants positioned at the center of the trust graphs (Fig. 2, bottom) trust others most, given a large number of thick edges oriented away from them. Overall, the 4 graphs suggest that familiarity and trust are not symmetrical.

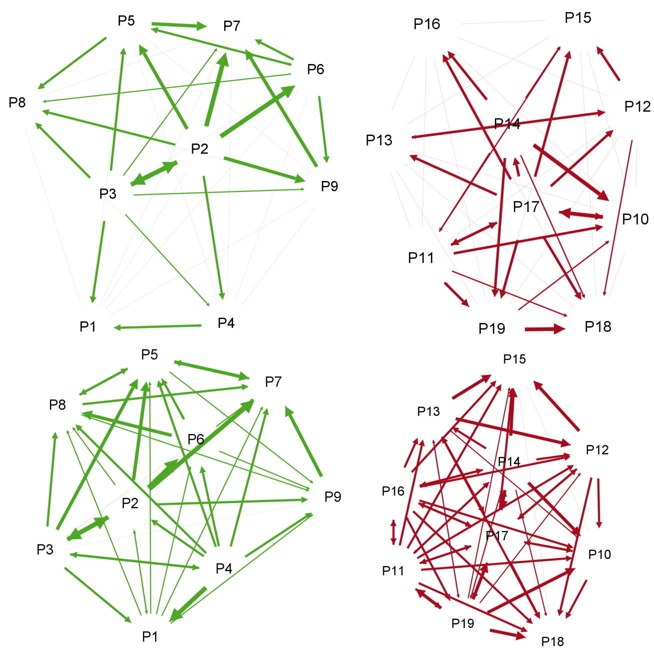


Fig. 2. Familiarity (left) and trust (right) networks for the 2 groups. Thin lines mark low scores of 1 while thick, most visible lines, mark scores of 5.

## 4. GROUPFUN EXPERIMENT

In addition to the online questionnaire, all 19 subjects participated in the GroupFun phase of the experiment that lasted approximately 3 hours per group.

### 4.1 The GroupFun System

GroupFun is a web application that helps a group of friends agree on a common music playlist for a given event they will attend, e.g. party. First, it is a music application that allows individuals to manage and combine their favorite music with others. In GroupFun users can listen to their own collection of songs as well as their friends' music. With the collective music database, the application integrates friends' music tastes and recommends a common playlist to them. Therefore, it aims at satisfying music tastes of the whole group by aggregating individual preferences.

Our design is based on previous user studies and pilot tests assessing group decision and interaction aspects. The current version of GroupFun was customized to fit the purpose of the experiment. Thus, it contains 3 sub-pages: "Home", "My Groups" and "My Music". In the first one, users see 4 playlists: one containing most popular songs, one used at a previous event, another one including recent uploads and the last one from a group party. In the "My Groups" page users create groups, upload and rate their music, invite friends and hear the group's songs (Fig. 3). Finally, in the "My Music" page members see their contribution to GroupFun: for each song the interface displays the associated group, the user rating and its name and artist.

One of the most important characteristics of GroupFun is that it combines music, friends and groups together. This distinguishes the system from related work. First, it meets users' expectations for conveniently organizing their individual music. Then, it supports effective communication among friends through ratings. Thus, users can participate in social activities while enjoying their common collection of music.

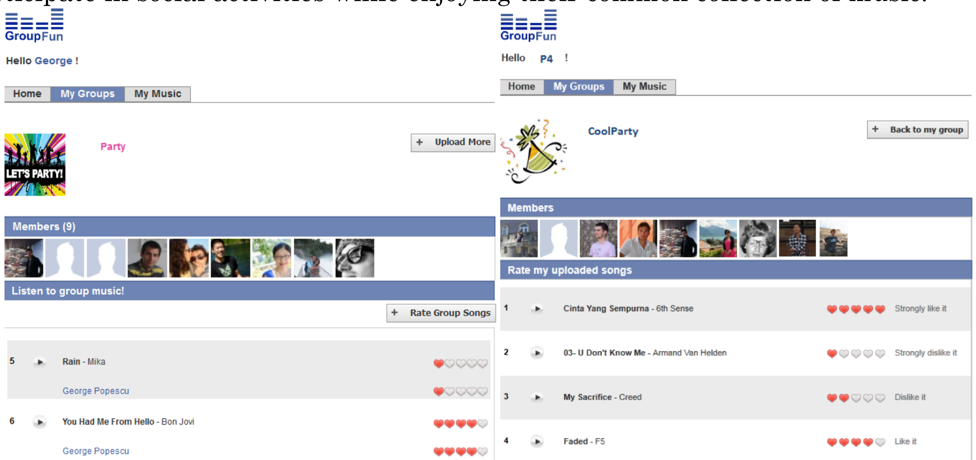

Fig. 3. Group (left) and individual (right) preference elicitation interfaces in the GroupFun music recommender system

Each subject was assigned the main task of uploading several individual songs and rating as many as desired group contributions. After uploading and rating self-songs each user was shown a list of songs already uploaded to the group by other members. The next step was to rate group songs without any authoring information. The interface (Fig. 3 right) displayed only the name of the song and the artist without the group member who contributed / uploaded the song (such as in Fig 3 left). Each song was indexed chronologically to the left hand side. A multi-functional play button allowed users to play, pause and fast-forward the selected song. In the center the song's name was included in bold and followed by the artist's name. Finally, members rated group songs by clicking on the appropriate hearts corresponding to a rating, ranging from "Strongly dislike it" (1 heart) to "Strongly like it" (5 hearts). The interface also includes GroupFun's logo (at the top), the 3 tabs of the application, the group's name and a top box displaying all group members with their profile photos.

### 4.2 Contributed Group Songs

Members of the "Cool party" group contributed with 39 songs (2h and 51m listening time, 212 MB) while those of the "Legendary party" with 47 songs (2h and 55m listening time, 202 MB) all in .mp3 format: an average of more than 4 songs per user.

### 4.3 Contributed Group Ratings

In the first group the 9 participants submitted a total of 203 ratings - most ratings submitted by P4 (39 ratings) and least by P2 and P6 (9 ratings). The average of those ratings, which we interpret as the initial group satisfaction score, was 3.27 (standard deviation SD=0.43). Table I shows the most satisfied member, P2 (average individual rating = satisfaction of 3.9), and least satisfied P7 (average satisfaction of 2.7)

In the second group the 10 participants submitted a total of 196 ratings for all 47 songs yielding a higher group satisfaction score (average of all group ratings) of 3.61

(SD=0.47). In this group most satisfied members are P10, P12 and P13 whereas least satisfied are P17 and P19 (Table I, second sub-table, 3rd row).

By performing the same computation (taking the average of ratings) only for the 24 songs that were used for the 3rd phase of the experiment we obtain the values from the last row of the two sub-tables. We notice that the average values decrease slightly. This is because for this 24 songs selection, few contributed others were eliminated. Those songs had highest ratings since their contributors also rated them. Group1's average eye-tracking ratings (average satisfaction) shows a decrease from 3.27 to 3.12 while Group2's average satisfaction from 3.61 to 3.52. Thus, the second group is still the most satisfied during both GroupFun and eye tracking. We also notice slight differences per participant for the two conditions.

Table I. GroupFun rating statistics: Group1 (top) and Group2 (bottom)

| **Group1** | **P1** | **P2** | **P3** | **P4** | **P5** | **P6** | **P7** | **P8** | **P9** | **#** |
|---|---|---|---|---|---|---|---|---|---|---|
| **#ratings** | 20 | 9 | 19 | 39 | 26 | 9 | 37 | 29 | 16 | **203** |
| **Average (39)** | 3.3 | 3.9 | 3.6 | 3.2 | 3.7 | 3.4 | 2.5 | 2.7 | 3.1 | **3.27** |
| **Average (24)** | 3.3 | 4.0 | 3.6 | 3.0 | 3.8 | 3.5 | 2.1 | 2.2 | 3.2 | **3.12** |

| **Group2** | **P10** | **P11** | **P12** | **P13** | **P14** | **P15** | **P16** | **P17** | **P18** | **P19** | **#** |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **#ratings** | 10 | 13 | 21 | 16 | 10 | 16 | 13 | 23 | 33 | 41 | **196** |
| **Average (47)** | 4.2 | 3.7 | 4.2 | 4.4 | 3.7 | 3.4 | 3.2 | 3 | 3.3 | 3 | **3.61** |
| **Average (24)** | 4.1 | 3.6 | 4.2 | 4.3 | 3.6 | 3.7 | 3.2 | 2.5 | 3.4 | 2.9 | **3.52** |

## 5. EYE-TRACKING EXPERIMENT

We designed our eye-tracking experiment aiming at understanding how people perceive other members' ratings. Furthermore, we were interested in analyzing the type of correspondences they make through visual contact. Our eye-tracking interfaces were designed to include the minimum relevant information: song name and artist, members' names and ratings information.

### 5.1 Experiment Steps

The eye-tracking study lasted for 25-30 minutes. During the gaze-recording experiment each participant pursued the following steps:

Step1. The experiment's admin first debriefs each participant on the nature of the experiment explaining the main process flow consisting in the 3 sub-steps presented in Fig. 1 – Experiment design (bottom, red box).

Step2. The admin assists each participant during the calibration process of the eye tracker. He then saves the calibration data for each user and loads it at the beginning of the recording session.

Step3. The user fixates the iPod headsets in his ears and positions him-self comfortably in front of the eye-tracking device. He uses his right hand to locate the "forward" and "pause" buttons on the iPod which are useful to shift to the next song (forward button one-time press), fast-forward the current one (forward button kept pressed), pause the song when ready (pause button one-time pressed) or continue to listen to the current song (pause button one-time pressed again).

Step4. At this moment the most important part of the experiment starts. Each user listens to one song and sees a corresponding interface at a time. In each interface are presented the title and artist information (at the top), a list of other members' names (to the left) and a list of their ratings (to the right). There are 24 recommended songs on the iPod corresponding to the 24 interfaces customized for each user. Each participant sees only the ratings of others and not his own. When ready with the song evaluation the participant gives a rating from "1" to "5" to the song he just listened to. He proceeds to the next song by clicking on the "forward" button with his right hand and any key on the computer keyboard controlling flow of interfaces on the eye-tracker with his left hand.

Step5. To conclude the study, the admin collects general comments through open discussions in order to assess the participants' overall perception of the experiment.

### 5.2 Interfaces

There are several essential observations we first need to make about our experiment. First, it is tailored for each user based on personalized interfaces: all interfaces contained peers' names and ratings for each current participant. We considered separately 3 types of interfaces: first 8 corresponding to all remaining group members, 8 with closest 5 members and the last group of 8 interfaces with 2 closest members. The terminology "closest members" is used to denote the highest trust values extracted from the questionnaire. Each of the last 16 interfaces was adjusted by the trust values. We determined that, in general, trust ratings were more reliable in assessing the closeness between group members compared with familiarity. Not only that trust ratings were higher than familiarity ratings but subjects rated others considering a more sparse distribution which allows for a better distinction between 5 and 2 most trusted members making our selection more accurate and meaningful.

The design of each interface is based on the fact that participants, as decision makers undertaking the rating task, use the social information displayed to them to adjust their ratings to the group preferences, result which is sustained by the social conformity theory.

In specifying a rating for a song that he listens to, a user explores the interface presented to him - the general design is shown in Fig. 4. The eye-tracking device records his eye movements and fixation times while browsing throughout the interface. Through our design we expect our participants to create numerous horizontal correspondences between the names' area (marked with a red border) and the ratings area displaying the rating (marked with a green border). These correspondences are essential for identifying the other members who contributed the most to one's personal decision.

In order to have clear regions on the interfaces for eye-tracking data collection we divided all interfaces into 3 regions, which we call areas of interest (or AOIs, Fig. 4):

- AOI Song = the grey rectangle at the top (AOIS blue label);
- AOI Users = the left part containing participants' names (AOIU red label);
- AOI Ratings = the right part containing user ratings (AOIR green label).

The interface from Fig. 4 contains a random list of names and surnames, which replaces real participants' names for anonymity. This figure was generated with the purpose of showing the reader a clear interface containing the information that our participants saw on their eye-tracking screens.

Altogether, the 24 interfaces were conceptualized as follows:

- First group of 8 interfaces: all other group members (Fig. 5, top row);
- Second group of 8 interfaces: 5 closest members (Fig. 5, middle row);
- Third group of 8 interfaces: 2 closest members (Fig. 5, bottom row).

Furthermore, we split each of the 8 into:

- Non-controversial: lowest standard deviation of ratings (Fig. 8, consensus);
- Controversial: highest standard deviation of ratings (Fig. 5, divergence).

Every next set of 2 consecutive interfaces corresponds to one of the two conditions:

- Trust: the ordering of names shows the most trusted members at the top and least trusted at the bottom;
- Popularity: complete names are sorted by their corresponding rating and the order is from highest ratings at the top to lowest at the bottom.

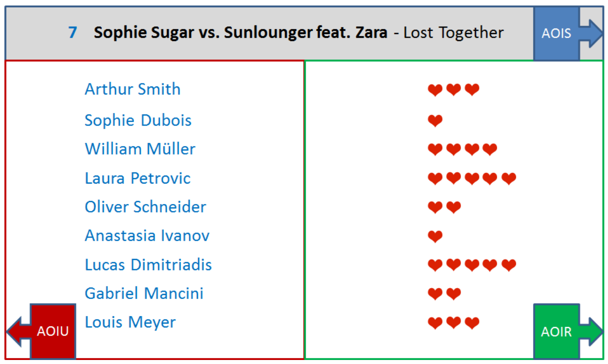

Fig. 4. Example of eye-tracking type interfaces

The fictive example from Fig. 4 corresponds to the following conditions: all other members (9 other members), divergence (highest standard deviation across ratings from “1” to “5”) and trust (the ordering follows the names column to the left).

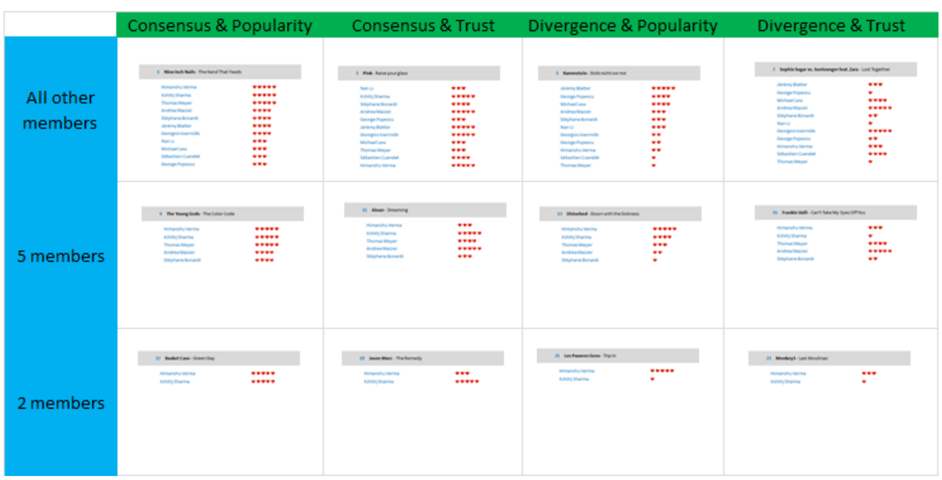

Fig. 5. Design examples for all 24 eye-tracking interfaces

A summary of all of the above conditions is visually presented in Fig. 5. For each box we displayed 2 songs consecutively. Thus, the same set of conditions would benefit from 2 ratings and 2 eye-gaze data sets. The eye-tracking experiment flow is from top-left to bottom-right. Each participant first sees one interface displaying Song1 in which all other members agree on similar ratings (consensus or lowest standard deviation of ratings) and songs are sorted by ratings (popularity): from highest to lowest values. In the second interface he sees Song2 in the same condition. For Song3 in the third interface, all other members’ ratings are sorted by trust: most trusted members at the top and least trusted at the bottom, from highest to lowest ratings. Song4 has the same condition. In interface 5 the current subject faces divergent ratings of all other members sorted by the song's rating. Song6 displays the same information. Song7 is included in an interface in which all other members submitted divergent ratings and their names are sorted by trust. Same condition applies for Song8. From Song9 until Song16 only the names and ratings of 5 most trusted members are displayed. Finally, from Song 17 to Song 24 only the names and ratings of 2 closest members are included.

### 5.3 Eye-Gaze Output

Eye-tracking technology allows to output different types of visual representation in the form of: eye-gaze, fixation points, circles, bubbles, gaze maps, etc. We preferred the fixation-point representation (an example is included in Fig. 6) as the most useful for extracting horizontal correspondences and fixation times. Participants' names from AOI Users in Fig. 6 (left hand-side) were erased in order to preserve their anonymity. The blue circles represent the main fixation points that the eye-tracker detected, the numbers inside the circles correspond to the milliseconds for each fixation and the blue lines show the eye's movements from one element to the next.

This visualization is a powerful representation being useful to: (1) extract fixation times for each of the areas of interest based on which we detect users that are most influential for each participant by summing up the fixation times per interface and (2) analyze fixation directions: (nearly) parallel lines (between AOI Users and AOI Ratings) account to the fact that participants associated other members with their ratings whereas vertical lines (within AOIs) denote the fact that users were in search of other information skipping either user or rating information.

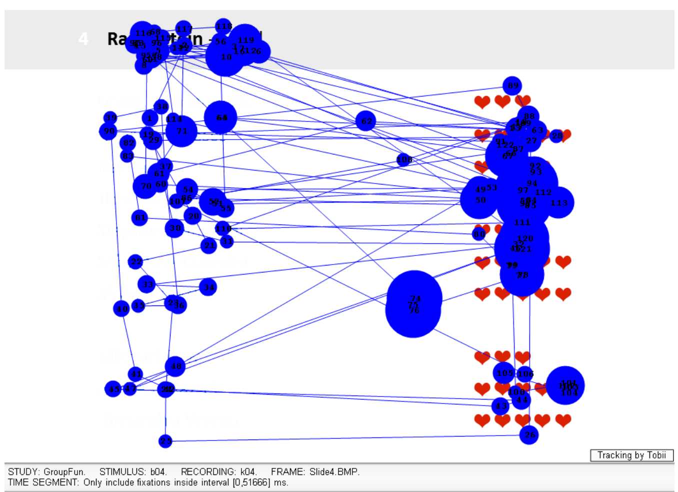


Fig. 6. Eye-tracking output using fixation-point map

### 5.4 Time in Areas of Interest

An extensive dataset containing fixation points (measured in seconds) was collected based on 19 participants x 24 interfaces = 456 eye-tracking output images. We recorded the total time gaze time per participant, further divided into AOI.

The total time indicated is the time indicated by the eye-tracker at the end of the experiment: 4,867.5 seconds for Group1 and 6,436.8 seconds for Group2. The total gaze time for the 19 participants was 11,304.3 seconds, i.e. an average of 595 seconds per participant or 10 minutes of continuous looking at the interface. The continuous gaze, without interruptions, is 9,417.1 seconds for both groups, i.e. 495.6 seconds per participant or 8.26 minutes of gaze recording per subject.

The proportion between the "Total Time Computed" (TTC) and the "Total Time Indicated" (TTI) gives the eye-tracker's efficiency for each group: 84% (4,080.7/4,867.5) for Group1 and 83% (5,336.4/6,436.8) for Group2. For the total time spend in front of all interfaces 84% of eye-gaze information was collected for the first 9 participants and 83% for the last 10 subjects. The TTC is further divided into the 3 conditions: "8/9" other group members, 5 closest members and 2 closest members.

Several preliminary conclusions yield from Table II:

(1) For all interfaces, participants spent a different amount of time in all 3 AOIs.
(2) For all members, users of both groups gazed mostly at AOIU, then at AOIR and finally at AOIS.
(3) For 5 closest members the situation is different per group. The first group gazed mostly at AOIS and least at AOIR. Since the songs' title and artist does not contain any social influence elements, but only the horizontal correspondences between AOIU and AOIR, users were least influenced. For Group2 users looked mostly at AOIU, then at AOIR and finally at AOIS, being most influenced.
(4) For 2 closest members, participants from both groups looked most at the song's name and artist, then at other members and finally at their ratings. This suggests that they were least influenced. However, since the interface area was reduced in size it was expected to obtain less gaze time than before.

Table II. Time view-statistics (in seconds) per group

| Group | | Total Time Indicated | Total Time Computed | Time in AOIS | Time in AOIU | Time in AOIR |
|---|---|---|---|---|---|---|
| **1** | | 4,867.5 | 4,080.7 | 1,536.7 | 1,382.2 | 1,161.8 |
| | 8 other members | | 1,642.1 | 386.6 | 640.0 | 615.5 |
| | 5 closest members | | 1,284.7 | 543.8 | 406.9 | 334.0 |
| | 2 closest members | | 1,153.9 | 606.3 | 335.3 | 212.3 |
| **2** | | 6436.8 | 5,336.4 | 1,517.4 | 2,188.9 | 1,630.1 |
| | 9 other members | | 1,793.5 | 274.6 | 891.7 | 627.2 |
| | 5 closest members | | 1,872.4 | 464.7 | 786.5 | 621.2 |
| | 2 closest members | | 1,670.5 | 778.1 | 510.7 | 381.7 |

The total time computed is higher for the second group compared with the first. The relative difference of about 80 seconds on average per participant (e.g. 5,336.4 / 10 = 533.64 seconds / member in Group2 vs. 4080.7 / 9 = 453.41 seconds / member in Group1) is not only due to the one extra group member, but also to the overall interest of this group's members to see others' tastes.

## 6. RESULTS

In this section, by comparing the two groups, we prove that the network structure is strongly correlated with the eye-tracking data. To do so, we extract eye-tracking fixation point times and produce a qualitative analysis of all participants' gaze. Furthermore, we compare average rating data submitted using GroupFun (individual satisfaction without social influence) and similar information gathered during the eye-tracking group study (individual satisfaction with social influence). Following a statistical analysis and computing p-values for each pair of 2 conditions we report that it is unclear if user interfaces produce big differences.

In the following we present the eye-gaze patterns and summarize the main findings related to the relative total gaze time in each area of interest. Taking into account all the submitted ratings for each of the 24 songs per group we compute preference correlation between each 2 participants and identify subjects with closest music preference. In the "ratings change" subsection, we include significant results that prove group alignment. First, we extract the ratings from the individual and group interfaces. Based on these contrasting values we compute the relative change rate per participant and per group determining a group influence score. We further enhance our results by removing participants' own songs and ratings. Finally, we analyze social influence with respect to group size for: "all other members", "5 closest members" and "2 closest members".

### 6.1 Eye-Tracking Gaze

In the current sub-section we present our results based on the eye-gaze output provided by the eye tracker, which includes time in AOIs and eye-gaze patterns.

#### *6.1.1 Time in Areas of Interest*

One principal question we address is: "What is the relationship between the dense / sparse familiarity and trust networks and the eye gaze patterns of our subjects?"

Firstly, we recall that members in Group2 are more connected than members in Group1: familiarity and trust networks are denser as it was inferred in Fig. 2. Our data shows that for all interfaces and any number of members, users in Group2 looked less at AOI Song and spent more time by looking at AOI Users and AOI Ratings: 38% vs. 28% for all interfaces, 24% vs. 15% for "all other members", 42% vs. 25% for "5 closest members" and 53% vs. 47% for "2 closest members". This implies that they were mainly interested in seeing and analyzing other members' ratings. Members in the first group behaved more egocentric, by looking mostly at AOI Song and spending less time in the other 2 AOIs. This observation can be summarized as follows: the more connected a group is through familiarity and trust, the more interested its members are to discover other members' music preferences.

The results also show that, within the same group, participants looked less at AOI Songs for the first 8 interfaces, i.e. when all other members and their ratings where displayed. This is an expected behavior since the heavy-information interfaces were included at the beginning and users needed more time to scan through the names and ratings: the screen surface of the AOIU and AOIR together is the biggest.

Another important finding is that for the interfaces containing "2 closest members", participants in both groups looked approximately half of the time at the top (AOIS: grey rectangle at the top in Fig. 4 or Fig. 6) and half at the bottom of the interface (AOIU and AOIR): 53% for Group1 and 47% for Group2 for AOI Song.

Next, we compare the time in AOIU with the time in AOIR within the same group. Over all interfaces, the time in AOI Users is longer than the time in AOI Ratings as follows: 34% vs. 28% overall, 39% vs. 37% for all members, 32% vs. 26% for "5 closest members and 29% vs. 18% for "2 closest members". The relative percentage difference between AOI Users and AOI Ratings is less for "all 8 members" in Group1 and for "2 closest members" in Group2. Turning back to the network structure we find this result very encouraging. We explain it by the fact that the more distant the relationship between users is (Group1) the more they are inclined to look at the whole list of members rather than closest ones.

### 6.1.2 *Eye-Gaze Pattern*

The eye-gaze pattern of each participant offers many important details showing the attention a subject paid to a certain part of the interface before making a decision. As it can be inferred from the below figures there are two types of correspondences which help understand the importance of information displayed to each participant through an objective measurement: horizontal lines (between users (left) and their ratings (right)) and vertical lines (within AOI Users or AOI Ratings). The (almost) horizontal correspondences are the ones that are important for our analysis as they denote the associations a subject made for his decision. For these horizontal lines we record both the other members' names (to the left) and their ratings (to the right).

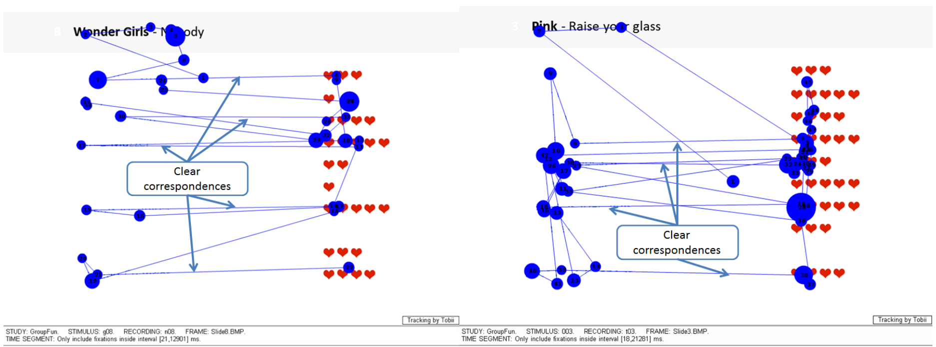


Fig. 7. Examples of clear, horizontal visual correspondences.

In Fig. 7 are marked clear horizontal correspondences between users and their ratings. Out of the entire number of subject names to the left (again hidden for privacy concerns) the current participant in Fig. 7 (left) looked at only 5 other members' ratings and drew associations with their ratings in evaluating his own rating. This member belongs to Group2 because there are displayed 9 users and their corresponding ratings. We consider only the correspondences that connect names and ratings on the same row, which we call "clear correspondences". They might not be completely horizontal and may vary with a small deviating angle. The screen-shot in Fig. 7 (right) shows the eye-gaze of another participant. This subject looked more to the bottom of the interface and created 4 horizontal correspondences that were recorded. Overall, the outputs we obtained were similar to those from Fig. 7 (left and right) including very good eye-gaze recording and fixation times.

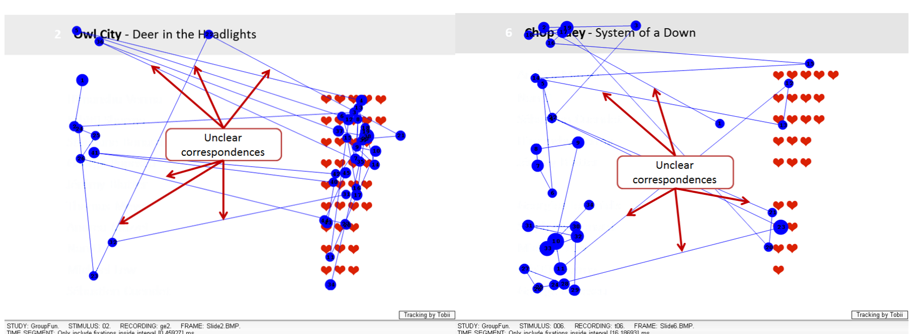


Fig. 8. Examples of unclear, skewed visual correspondences

However, not all eye gaze patterns were very clear (Fig. 8). We noticed some outputs in which users looked:

(1) mostly (90% of the time or more) at the song's name (AOI Song);
(2) at parts of the interface displaying no information (not song, names or ratings);
(3) from top to bottom on both users and ratings columns drawing few associations;
(4) in a non-uniform manner, skimming from one corner of the interface to the next.

In this case fixation times do not correspond well to the names on the left and corresponding lines are not horizontal but skewed. Consequently, in this case we report that the participant combined the 3 AOIs drawing associations from AOI Song to AOI Users (top names) and from AOI Song to AOI Ratings (middle and bottom hearts). Here we cannot make any judgment about group influence for any of the members on the left. However, we do use the fixation times for each of the 3 AOIs.

### 6.2 Music Preference Correlation

We use the Pearson correlation score, which reflects the degree to which 2 variables are related, to determine subjects who are closest preference-wise (Fig. 9).

As an example, in the first group, P1 estimated to be closest to P4 (highest familiarity score) but the ratings show him closest to P8 (preference correlation values, Fig. 9, left). Reciprocally, P4 believed to be closest to P1 (again, highest familiarity score) but rating correlations prove him to have closest tastes to P9 (Fig. 9, left). Another case shows P7 to know most P5 and P9's music tastes. Once again the correlation coefficient shows that he likes music similarly to P8 (Fig. 9, left).

In the second group P19 is the rating reference for all other members as P11, P13, P14, P16, P17 and P18 all have highest rating correlation with P19 (Fig. 9, right). Furthermore, P18 and P19 stated to known each-others' preferences the most. P12 stated to know P15's tastes and these tastes are also highly correlated (Fig. 9, left).

We plot the correlation values into two un-oriented graphs in Fig. 9 and compare them with the familiarity and trust networks from Fig. 2. The thick lines between two participants indicate a strong correlation (close to 1) between their music tastes. The Pearson score is based on similar rating values they submitted for the same songs. By contrast, thin lines indicate a negligible correlation (close to 0). In the center of the two graphs below there are the participants with highest music correlation with others. This cross-correlation (Fig. 9) shows the objective measurements of the individual preference elicitation whereas the first graphs (Fig. 2) denote the subjective measurement of familiarity and trust values. Thus, P1 and P3 positioned in the center of the green graph and P11, P18 and P19 positioned in the center of the red graph are closest to the two groups' "average" music preference even though they have not necessarily received highest familiarity or trust scores.

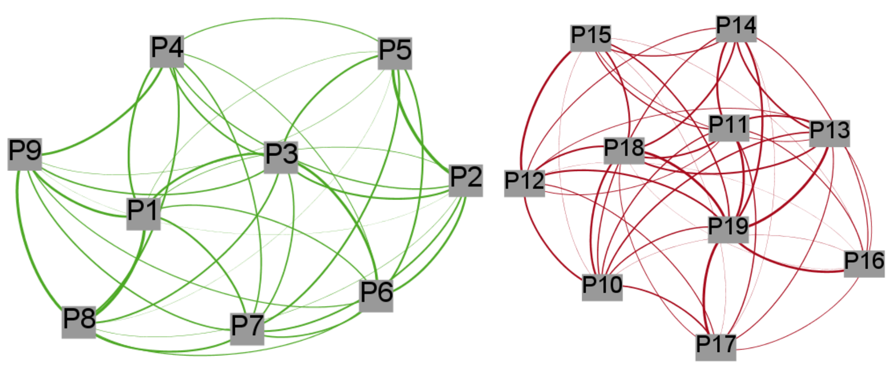


Fig. 9. Pearson correlation networks for the two groups

Comparing the familiarity and trust networks in Group1 (Fig. 2 in green) with the Pearson correlation (re-scaled from [-1, 1] to [0, 5]) music-preference network (Fig. 9 in green) we observe that the overlay between values is small. P4, for instance, mentioned to know P1's music preference but he has most similar preferences with P7 whereas all other correlation values are close to 0. He also trusts P1 the most and P2, P3, P5, P6, P8, P7 P8 and P9 equally. P12 from Group2 is most similar music-wise with P11, P15 and P18 (Fig. 9 in red). However, he trusts most P10, P15, P17 and P18 and knows most P13, P15 and P18's preferences (Fig. 2 in red). We conclude that only to a small extent the music-preference correlation network match the familiarity and trust networks.

### 6.3 Ratings Change

In this sub-section we measure the change in ratings for the two conditions: individual rating in GroupFun and group rating during eye tracking. Based on these changes we compute a social influence score per group. We further elaborate our analysis after excluding own contributed songs. In addition to our experiment setup we compute the change rate for the consensus and divergence conditions.

#### *6.3.1 Ratings in Individual and Group Interfaces*

In Table III is presented the average group rating (satisfaction) per type of interface for the 24 songs that were rated and displayed both in GroupFun and in the interfaces during the eye-tracking experiment. These 24 songs were the same for the 2 conditions (out of 39 for Group1 and 47 for Group2) but different per group.

The "Average Before" condition in the third column of the table below includes the average values of all ratings that a group submitted in GroupFun, i.e. before seeing the eye-tracking interface. The "Average Interface" column includes the average of ratings displayed in the interfaces. Finally, the last column, "Average After", includes the average rating values per group after the participants saw the group interfaces. Comparing columns 3 and 5, it yields that generally both groups slightly reduced their ratings (less satisfied). For all interfaces the average rating in Group1 dropped from 3.21 (SD = 0.64) to 3.15 (SD = 0.41) while in Group2 it dropped from 3.53 (SD = 0.57) to 3.44 (SD = 0.31). By opposition, a significant increase is observed for the last condition: "2 closest members". Here both groups increased their average ratings from 3.10 to 3.27 for Group1 and from 2.93 to 3.33 for Group2. This means that subjects gave higher ratings when seeing the ratings from most trusted peers.

Table III. Average rating comparison per group, sets of interfaces and experiment phase

| Interface | Group | Average Before | Average Interface | Average After |
|---|---|---|---|---|
| **All** | **1** | 3.21 (SD = 0.64) | 3.52 (1.24) | 3.15 (SD = 0.41) |
| **All** | **2** | 3.53 (SD = 0.57) | 3.55 (1.24) | 3.44 (SD = 0.31) |
| **8** | **1** | 3.11 (SD = 0.79) | 3.43 (1.14) | 3.05 (SD = 0.38) |
| **9** | **2** | 3.69 (SD = 0.57) | 3.50 (1.13) | 3.46 (SD = 0.50) |
| **5** | **1** | 3.05 (SD = 1.02) | 3.65 (1.17) | 3.11 (SD = 0.63) |
| **5** | **2** | 3.75 (SD = 0.84) | 3.65 (1.17) | 3.52 (SD = 0.50) |
| **2** | **1** | 3.10 (SD = 1.01) | 3.50 (1.41) | 3.27 (SD = 0.49) |
| **2** | **2** | 2.93 (SD = 0.72) | 3.50 (1.41) | 3.33 (SD = 0.55) |

Despite the high standard deviation of ratings displayed in the interfaces (all bigger than 1), we report higher preference alignment after seeing the interfaces for both groups: "Average After" column in Table III. The numbers in the brackets show very low standard deviation values (max. of 0.63) for all conditions and both groups. Comparing this with the standard deviation from the 3rd column in which values are much higher we find that, despite the fact that individual preference diverged more when taken individually, subjects' ratings converged when shown the same reference, expressed in the eye-tracking interfaces, thus proving social conformity.

One final observation targets the average ratings in the interface. As it can be observed from the "Average Interface" column the average rating in all the 24 interfaces was balanced: 3.43 smallest average for "8 other members" in Group1 and 3.65 largest average for both groups for "5 closest members". With this design we aimed at influencing our subject's decisions through social influence coming from most trusted members and not through high ratings.

### 6.3.2 Change Rate

We computed a normalized group influence score per participant measuring the difference in ratings that each subject submitted while being exposed to the individual and group interfaces, respectively.

For each group member we compute an average rating for the GroupFun experiment in Phase I as a measure of individual satisfaction without social influence (Eq. 1). These results were summarized in Table I. When referring to individual / group satisfaction we always compute it as an average across individual / group ratings. Next, we normalize this by the number of GroupFun rated songs since this number is different among users. U is the set of all users (participants or subjects), $R$ is the set of all ratings and S the set of all songs. $r$ represents the rating submitted for each of the subset of 24 songs individually, i.e. without any external influence from the other participants and $r_i$ is the rating submitted by participant $i$. We mapped these values with the familiarity and trust network to observe if the subjective evaluation of familiarity and trust corresponds with objective ratings.

$$satisfaction = \frac{\sum_{1\leq j\leq |S|} r_{ij}}{\#r_i}, \forall i \in U, r_{ij} \in R, |S| = 39 \ or \ |S| = 47 \quad (1)$$

Similarly, we compute a satisfaction score for the eye-tracking experiment ($r^*$) considering only the ratings given by the 24 songs in the interface (Eq. 2):

$$satisfaction_i^* = \frac{\sum_{1\leq j\leq 24} r_{ij}^*}{\#r_i^*}, \forall i \in U, r_{ij}^* \in R^* \quad (2)$$

$\#r_i$ is the number of rated songs for each participant out of the 39 and 37 songs, respectively (Eq. 1). For (Eq. 2) $\#r_{ij}^* = 24$ as all participants rated all 24 songs.

Next, we compute a normalized group influence index per participant measuring the difference in ratings when exposed to the individual interface compared with the group interface, only for the same songs (rated twice).

$$\Delta_i = \sum_{1..j} |r_{ij} - r_{ij}^*|, \forall i \in U, r_{ij} \in R, r_{ij}^* \in R^* \quad (3)$$

$$\varepsilon_i = \sum_{1..j} r_{ij} - r_{ij}^*, \forall i \in U, r_{ij} \in R, r_{ij}^* \in R^* \quad (4)$$

In Table IV we denote the following:

- Avg_i: average rating for the individual, GroupFun interface: 24 songs;
- Avg_g: average rating for the group, eye-tracking interface: 24 songs;
- t: number of songs rated twice (in both interfaces) in 3 categories: "+", "-" and "0";
- +: number of positive changes: increase of ratings;
- -: number of negative changes: decrease of ratings;

- 0: number of neutral changes: ratings kept the same;
- Δ: absolute rating difference;
- ε: signed rating difference;
- c: change rate percentage.

In Table IV we compare the ratings and change rate only for the 24 songs displayed for the second part of the experiment. One first observation is that participants did indeed change their ratings at a very high rate (45.17% on average for both groups). Most importantly, the more connected group (Group2) changed more (50.19%) than the other (40.15%). To strengthen this result and prove that the 10% difference is significant a regression analysis that uses familiarity/trust to predict the rating change is essential. Here we focus on more general results and less on the analytic details, which we plan to address in future work. In the table we marked in bold the participants who had a high change rate. Participants in Group1 rerated 152 songs out of which 42 where an increase in values, 25 a decrease and 85 where kept the same. In Group2 the 10 participants submitted 271 new ratings out of which 63 where increases, 64 decreases and 144 where kept the same.

Table IV. Rating change statistics

| | Avg_i | Avg_g | t | + | - | 0 | Δ | ε | c |
|---|---|---|---|---|---|---|---|---|---|
| **P1** | **3.29** | 3.58 | 17 | 8 | 6 | 3 | 21 | 7 | **82.35%** |
| **P2** | **4.00** | 3.33 | 7 | 0 | 0 | 7 | 0 | 0 | 0.00% |
| **P3** | **3.63** | 3.42 | 16 | 5 | 2 | 9 | 9 | 3 | 43.75% |
| **P4** | **3.08** | 3.29 | 24 | 10 | 7 | 7 | 29 | 5 | **70.83%** |
| **P5** | **3.78** | 3.63 | 23 | 0 | 2 | 21 | 2 | -2 | 8.69% |
| **P6** | **3.50** | 2.88 | 8 | 1 | 1 | 6 | 2 | 0 | 25.00% |
| **P7** | **2.14** | 2.54 | 22 | 9 | 3 | 10 | 16 | 10 | **54.54%** |
| **P8** | **2.24** | 2.54 | 21 | 5 | 2 | 14 | 7 | 3 | 33.33% |
| **P9** | **3.21** | 3.17 | 14 | 4 | 2 | 8 | 7 | 3 | 42.86% |
| **Group1** | **3.21** | 3.15 | 152 | 42 | 25 | 85 | 93 | 29 | 40.15% |
| **P10** | **3.63** | 4.14 | 7 | 2 | 1 | 4 | 6 | 0 | 42.85% |
| **P11** | **3.54** | 3.56 | 9 | 1 | 3 | 5 | 4 | -2 | 44.44% |
| **P12** | **3.96** | 4.21 | 14 | 1 | 4 | 9 | 5 | -3 | 35.71% |
| **P13** | **3.83** | 4.27 | 11 | 5 | 0 | 6 | 8 | 8 | 45.45% |
| **P14** | **3.5** | 3.57 | 7 | 1 | 2 | 4 | 4 | -2 | 42.86% |
| **P15** | **2.96** | 3.56 | 9 | 3 | 2 | 4 | 6 | 2 | **55.56%** |
| **P16** | **3.13** | 3.18 | 11 | 2 | 6 | 3 | 11 | -5 | **72.73%** |
| **P17** | **3.29** | 2.54 | 13 | 2 | 7 | 4 | 20 | -14 | **69.23%** |
| **P18** | **3.25** | 3.41 | 17 | 2 | 3 | 12 | 5 | -1 | 29.41% |
| **P19** | **3.33** | 2.91 | 22 | 3 | 11 | 8 | 16 | -8 | **63.64%** |
| **Group2** | **3.44** | 3.54 | 120 | 22 | 39 | 59 | 85 | -25 | 50.19% |
| **Total** | **3.32** | 3.34 | 271 | 63 | 64 | 144 | 178 | 4 | 45.17% |

The signed difference and absolute difference between the two arrays of ratings, $r_{ij}$ and $r^*_{ij}$, are computed in order to determine the number of changes: positive: "+" (increase), negative: "-" (decrease) and neutral: "0" (same).

In the second column of Table IV are marked in bold the higher average satisfaction of the two conditions per participant: individual (second column) and group (third column). We report that only 7 (out of 19) participants were more satisfied with the music they rated in the first condition. All others increased their ratings afterwards. Overall, the group satisfaction varies only slightly: 3.15 for the group interface compared with 3.21 for the individual interface for Group1 and 3.54 compared with 3.44 for Group2. Additionally, both Δ and ε help understand the variations of ratings. The decrease and increase may correspond to a large or small variation in rating change according to each rating update. Despite the fact that the number of re-ratings does vary per participant (minimum of 7 songs in both groups and maximum of 24 in Group1 (P4) and 22 in Group2 (P19)) we report large change rate across the majority of participants (largest values in bold – last "c" column).

Another important result is that members in the first group have achieved about the same satisfaction in both conditions (4 members were more satisfied and 5 less satisfied) whereas in the second group 8 out of 10 members have improved their ratings overall - only P17 and P19 were less satisfied. Group2 is also more balanced (minimum of 29.41% rating change for P18 and maximum of 72.73% for P16) whereas Group1 shows larger variance (minimum of 0.00% for P2 and 82.35% for P1).

Overall, the more songs people rate the higher the group influence and social alignment - the bold percentages from the right column mark group influence higher than 50%: P1 (82.35%), P4 (70.83%), P7 (54.54%) from Group1 and P16 (72.73%), P17 (69.23%), P19 (63.63%) and P15 (55.56%) are the participants that changed the most.

### 6.3.3 *Group Influence Score*

Next, we consider the following 2 functions:

$$\bar{r}_i = \frac{\Delta_i}{\#r_i}, \forall i \in U, r_i \in R \tag{5}$$

$$\eta_i = \frac{\sum_{1..j}(\#r_{ij}^{+} + \#r_{ij}^{-})}{\sum_{1..j}(r_{ij}^{+} + r_{ij}^{-} + r_{ij}^{0})}, \forall i \in U, j(i) \in R \tag{6}$$

The first one (Eq. 5) computes the average rating difference (on a 1 to 5 scale) for the absolute rating difference whereas the second one is a value between 0 and 1 measuring the number of rating changes vs. the total number of ratings. The second variable (Eq. 6) represents the normalized group influence score.

For both groups the high $\eta$ suggests that individuals were strongly influenced by their peers' ratings. An average $\eta$ for Group1 is 0.4 and for Group2 0.5 which corresponds to the fact that, on average, each participant in Group1 changed his previous rating with 0.54 (on a 1 to 5 scale) and each participant in Group2 changed each of his ratings with 0.72 points. Indeed social influence in the second group is more homogenous given the group's higher connectivity compared with the first one (Fig. 2). In Group1 some participants tended to have a very strong preference for which they care about and would not adapt to the group decision.

### 6.3.4 *Quantifying Change Rate*

Our goal in this sub-section is to provide more rating changes details. We categorize rating differences into "1", "2", "3" and "4" (maximum) points corresponding to the absolute difference between the initial submitted ratings (GroupFun) and the updated ones (eye-tracking). "1" represents small rating differences such as "3-2" or "4-5" whereas 4 corresponds to the maximum difference: either "1-5" or "5-1".

For each participant in each group we sum up the number of times each rating difference falls into the "1", "2", "3" and "4" categories corresponding to "small", "medium", "large" and "very large" changes. In total, for both groups, we obtained 128 changes out of which the majority, 90, were 1-point (70% "small changes"), 28 were 2-point (22% "medium changes"), 8 were 3-point (6% "large changes") and 2 were 4-point rating changes (2% "very large changes"). Indeed, in the 2 last cases one individual changed his rating from "1" ("strongly dislike the song") to "5" ("strongly like the song") and vice-versa.

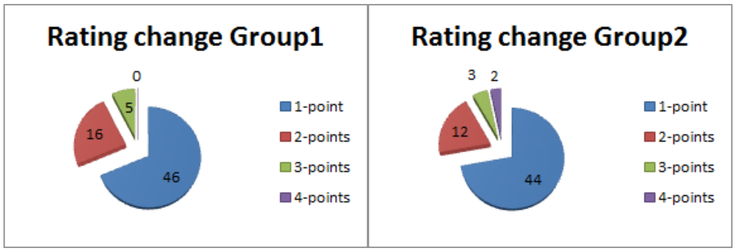


Fig.10. Rating change per group for 1, 2, 3 and 4-point(s) differences

The two groups exerted similar change behavior. In total we noted 67 rating changes (out of 152 songs) in Group1 and 61 (out of 120 songs) in Group2. The rating differences between groups also show similar results: 46 vs. 44 "1"-point rating changes, 16 vs. 12 "2"-point changes, 5 vs. 3 "3"-point changes and 0 vs. 2 "4"-point changes in Group1 and Group2, respectively. We expected to obtain more changes for the second group given the overall social alignment and the results obtained from the eye-tracking data. Members' preferences prove that this is not the case and that variations are small. A more detailed analysis considering each individual user and performed separately yield that individuals exert very diverse preference behavior. The vast majority of subjects changed their rating only slightly which corresponds to a more general psychological result that people have intrinsic preferences which they value. Their peers / the context influence them but these produce only a small change.

To conclude this section, our data proves that individuals do not have fixed preference values that they conform to but they adapt to group suggestions.

#### 6.3.5 *Social Influence and Group Size*

For each type of eye-tracking interfaces displaying: (1) all other group members, (2) 5 closest members, and (3) 2 closest members we compute the average change rate for all participants and per group. The aim is to understand if individuals align more when they see more or fewer opinions coming from the people they trust most. Our hypothesis is that they would change more when they face the preferences of 2 closest members and re-state own preferences when facing all others' preferences.

First of all, when comparing the two groups among themselves we notice that the members of the second group adapted more to the group decision as the number of members in the interface was reduced. This validates our hypothesis: Closest 5 and Closest 2 have similar change rate scores: 58.71% and 56.96%, respectively. However, the change rate decreased in Group1 from the first to the last interfaces. This proves that Group1 members changed their preferences less when seeing fewer opinions, even if those opinions came from the members that they trusted the most!

The main reason for this phenomenon we associate with the familiarity and trust networks. Group1 is the one in which people are more isolated and adapt less. Consequently, this result is in agreement with this group's characteristics. The situation in Group2 is reversed: in this connected group individuals adapt more, gaze more at others' preferences and only slightly decrease their overall satisfaction. The change rate further proves that they also adapt more when they see their 5 and 2 closest members' preferences. It is also interesting to note that the "starting point" of both groups is about the same: 46.20% vs. 44.77% change rate for all other group members. Indeed, social relationships are beneficial for the individual decision making process in the second group (Fig. 11).

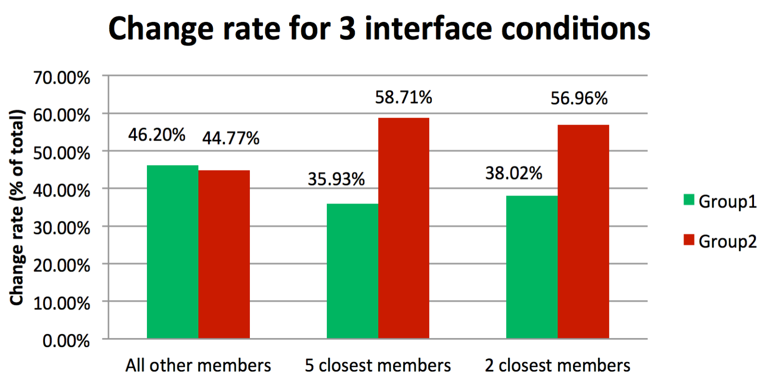


Fig. 11. Change rate per interface condition group comparison

#### 6.3.6 *Change Rate for Consensus and Divergence*

In the following we analyze the change rate for consensus and the divergence conditions for both groups. The data from the Table V below show that the second group, in general, changed more for all the (sub-) conditions. Furthermore, individuals changed most when seeing 2 members' preferences, 5 others and finally all others in the strongly connected group under both the consensus and divergence conditions. This detailed result was obtained after investigating the change rate. We find encouraging the increase in the change rate as the group size decreases.

Table V. Change rate comparison for consensus and divergence in the two groups

| Change Rate – Group1: 44.04% | | |
|---|---|---|
| **Group Size / Variable** | **Consensus** | **Divergence** |
| **All other members** | 2 – 44.79% | 2 – 60.18% |
| **5 closest members** | 1 – 56.48% | 3 – 52.77% |
| **2 closest members** | 3 – 33.33% | 1 – 66.66% |
| **Overall Consensus / Divergence** | 41.59% | 46.61% |
| **Change Rate – Group2: 52.92%** | | |
| **Group Size / Variable** | **Consensus** | **Divergence** |
| **All other members** | 3 – 63.63% | 3 – 45.83% |
| **5 closest members** | 2 – 65.00% | 2 – 55.83% |
| **2 closest members** | 1 – 70.00% | 1 – 68.33% |
| **Overall Consensus / Divergence** | 49.71% | 54.07% |

In Group2, for consensus we obtained: 63.63% change rate for all other members, 65.00% change rate for closest 5 members and 70.00% for 2 closest members. We report even greater relative differences for the divergence condition: 45.83% change rate for all other members, 55.85% for 5 closest members and 68.33% change rate for 2 closest members. In Group1 the situation is more complex. First of all, under both the consensus and the divergence condition, overall participants changed the same. For the consensus condition subjects changed most when seeing 5 closest members (56.48%), then all other members (44.79%) and finally 2 closest members (only 33.33%!). However, when facing divergent ratings they changed most when seeing 2 closest members (66.66%), then all other members (60.18%) and least for 5 closest members (52.77%).

We expected individuals to change less when consensus was presented to them and more when the variance of others' preferences was large. However, the data shows that this is the case only for the first group. Despite close scores, the change rate in both groups shows higher change for the divergence condition.

### 6.4 Absolute Rating Difference

We noticed that the rating differences among participants account for the social pressure of other group members. Thus, we investigate in more detail the rating differences. In this section we measure the rating difference between initial (individual) and eye-tracking (group) conditions for the same set of 24 songs. First of all, we select both sets of ratings and compute the unsigned (absolute) difference between them (Eq. 7). Then, we sum the difference for each song per participant and divide by the total number of songs rated twice in the two conditions (Eq. 8). We obtain the d-values for all participants and compare them for Group1 (Eq. 9) and Group2 (Eq. 10) across experiment conditions.

Table VI. Absolute difference values for opposing experiment conditions

| Overall | Trust vs. Popularity | Consensus vs. Divergence | Fixed Divergence | Trust vs. Popularity |
|---|---|---|---|---|
| All members | 0.586 | 0.586 | All members | 0.623 |
| 5 members | 0.652 | 0.659 | 5 members | 0.565 |
| 2 members | 0.738 | 0.729 | 2 members | 0.756 |
| All interfaces | 0.654 | 0.654 | All interfaces | 0.646 |

| Fixed Familiarity | Cons. vs. Div. | Fixed Consensus | Trust vs. Pop. | Fixed Popularity | Cons. vs. Div. |
|---|---|---|---|---|---|
| All members | 0.537 | All members | 0.543 | All members | 0.644 |
| 5 members | 0.512 | 5 members | 0.762 | 5 members | 0.800 |
| 2 members | 0.784 | 2 members | 0.700 | 2 members | 0.688 |
| All interfaces | 0.597 | All interfaces | 0.664 | All interfaces | 0.710 |

$$d_{ij} = |r_{ij} - r_{ij}^{*}|, \forall i \in U, \forall j \in S, r_i \in R, r_i^{*} \in R^{*} \tag{7}$$

$$d_i = \frac{\sum_j dij}{k_i}, \forall i \in U, \forall j \in S, k_i \in R \cap R^{*} \tag{8}$$

$$\bar{d}_i^1 = \frac{\sum_{i=1}^{9} d_i}{9}, \forall i \in U, r_i \in R, r_i^{*} \cap R^{*} \tag{9}$$

$$\bar{d}_i^2 = \frac{\sum_{i=10}^{19} d_i}{10}, \forall i \in U, r_i \in R, r_i^{*} \cap R^{*} \tag{10}$$

We populate the table below with all the d-values computed for all sets of conditions. The generic d-value corresponds to the value from Eq. 11 below, which calculates the average of all d-values across all 19 participants.

$$\bar{d} = \frac{\sum_{i=1}^{19} d_i}{19}, \forall i \in U \tag{11}$$

In Table VI similar d-values across conditions are reported. There are 272 recordings (out of 456) for which d-values were computed, i.e. 60% of all cases amounting for 128 changes and 144 same values yielding a change rate of 47.05%: participants changed their rating almost once at every 2 songs! Large d-values are computed in the following conditions: “2 other members” for trust vs. popularity (0.738) and consensus vs. divergence (0.729). For fixed divergence we report higher d-value again for “2 closest members” when comparing trust with popularity (0.756). The same pattern is observed for “2 closest members” in “consensus vs. divergence”. However, the greatest absolute rating differences are reported for the “5 closest members” when we have a fixed consensus and compare trust with popularity (0.700) and when popularity is fixed and compare consensus with divergence (0.800).

In summary, this data shows that people change their preferences to a greater extent when they see 2 and 5 closest members’ ratings. The extent in this case is the d-value, or the absolute difference between initial and final rating.

## 7. DISCUSSION

In this paper we described a user study for measuring social influence. In particular, we wanted to understand how group influence affects one’s own preference through rating. Towards this, we have conducted a between group user study where participants were asked to provide ratings of songs before and after they saw group ratings. In a separate survey, we also collected participants’ familiarity and trust evaluations towards others. In order to understand how people perceive other people’s ratings, we designed and run an eye tracking experiment for two groups of 9 and 10 participants. Through our experiments, we have demonstrated that the members of the more connected (in terms of familiarity and trust) group showed more change in their rating after seeing other members' recommendation. We have also extensively discussed the factors of change in rating behavior.

Much effort and time was put into organizing the user study and the eye tracking experiment. When preparing for its development we took into account the extensive feedback we received from both pilot studies and user interviews we previously carried out as well as other researchers’ advice for handling eye tracking data. The data collection phase depended on the 19 participants’ availability. As a result we carefully planned all individual rating and eye-tracking sessions for which each set of interfaces was individually customized considering users’ familiarity and trust data.

The specific evaluation method proposed in this paper is using ratings and eye gaze to measure people’s decision change without (first) and with (after) social influence. An alternative to this would have been to simply ask participants direct questions through an online questionnaire in order to collect their re-ratings. Or re-use our music recommender system, GroupFun, for the same purpose some time later, without the eye-tracker. The main limitation of these two alternatives is that they allow for a rating-based analysis only, without any eye-gaze information. Thus, the preference correlation and rating change computation would have been performed in the same way but we would have not been able to understand the details that account for a decision change. On the other hand, an advantage of these 2 methods is that they can be easily applied by reducing user effort and simplify the experiment setup. Our method is also different from others that emphasize the effects of memory on one’s decision. Previous research work has proved the role memory plays in decision-making: individuals tend to change their preferences / ratings more after longer time periods. A key advantage of our experiment setting is that despite the relatively short time lapse between the first and second part of the experiment (7-10 days), our participants still changed their decision to a great extent (40 to 50%) under the influence of close peers. When facing the opinion of other people only a few minutes later people change to a smaller extend (Zhu et al., 2012). An extension of our framework would be to carry out semi-structured interviews with our participants in order to reveal the reasons related to the produced change.

The key advantages of our study are:

(1) We tackle a novel problem by adopting the use of eye tracking in a music context to analyze participants' decision process in detail (not only ratings). Using a staged experiment we evaluate trust/familiarity, attention and social influence. Previous research employed social network information or social pressure to analyze people's decision-making, but our experiment design proposes the use of eye tracking technology to confirm social alignment.

(2) The experiment setup can be relatively easily extended for a greater number of participants deciding on other similar problems: holiday destinations, movies, news, books, games, etc. Results can be then easily generalized to these domains.

(3) A detailed discussion of the experiment design is provided. The proposed explanation has clear implication for the design of group recommender systems and research on group influence.

By opposition there are also a few limitations we report:

(1) The current study paper emphasizes the experiment setup and does not focus on very detailed analytical results. Our evaluation is not compared with other approaches given that this method was not previously used.

(2) The scale of the experiment is small (19 participants) but it is difficult to conduct even such a small-scaled user study due to: participants' availability for each experiment phase, the setup of the customized interfaces per participant given group characteristics, names and ratings manipulation, extracting and interpreting eye tracking data (horizontal correspondences), etc.

(3) The two subject groups are different, leaving many potential confounds. The group with higher inter-person trust (Group2) changed their ratings more after reviewing ratings from peers. We suggested that it is this factor that drives the rating change. However, there might be other factors as well. For example, it is possible that the observed difference in rating change is not (entirely) based on familiarity or trust, but on some unintended confound coming from the particularity of these people. For instance, one group might change less simply because they have stronger conviction in their music preference, or because their personality is more stubborn. Because Group 1 is recruited from a class while Group 2 is recruited from two related research labs, Group 2 has a more diverse age distribution, more diverse national background, longer time spent together as a group and stronger shared identity. One further needs to investigate if one of those factors drives rating change more than familiarity/trust. This can be achieved through a similar experiment setup. Results need to compare the causality between each factor and rating change. Overall, the current results are not so strong due to the small and confounded subject population, so future eye-tracking studies of more homogenous and larger populations are needed.

One of the eye-tracker's key roles in the experiment is that it proves the fact that individuals do pay more attention to who they trust. We acknowledge the fact that this alone does not directly demonstrate that they changed their mind because of attention or trust. The inferences we made are based on the data and, additionally, on participants' comments collected at the end of the experiment.

## 8. CONCLUSIONS AND FUTURE WORK

This paper introduced a novel experiment in which eye tracking is applied in a group recommender system to understand people's decision process in detail. The result is overall interesting and encouraging because it relies on both subjective (familiarity and trust) and objective (eye-tracking) data. Specifically, we study the extent to which people that belong to a particular group can change and align their opinion (in terms of music preferences) with other individual group members.

In this report we start by highlighting our experiment design. With the use of GroupFun, our music recommender system, we logged participants' individual ratings with no external influence and compared them with changed preferences. Given the scores of their peers in the same group, subjects were asked to re-submit new rating values. This between-group design facilitated in-depth discussions and analysis about group alignment and individual satisfaction driven by group connectivity.

We analyzed both rating and eye-tracking data. By comparing the subjective values from the familiarity and trust networks with the objective ones recorded with the eye-tracking device we concluded that the more connected a group is, the more correspondences subjects produce between users' names and their ratings and the more they change their preferences. Users create these visual associations in order to develop individual decisions.

The eye-tracking data confirm the social relationships among participants: individuals followed their closest members' preferences during the eye-tracking experiment as subjectively evaluated through familiarity and trust. By opposition, music preference is independent of participants' familiarity and trust.

The main results of our experiment prove the fact that individuals do not have innate preferences but adapt their choices to those of the group. We computed social influence scores per participant identifying individuals who are more group-oriented and change more from those who are more self-oriented and vary their preferences less. These results have implications for psychology and personality research as well as decision making strategies.

We acknowledge the fact that the evaluation of our experiment on such small networks, as well as only two networks, needs further development. The conclusions of the paper may change if more groups are studied. To address this issue we plan to carry out more experiments in our future studies by analyzing: more similarly sized-groups, other factors accounting for ratings change, e.g. personality, group identity, and other decision domains: movies, articles, social outings, etc.

## ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation, under grant SNSF-200021-119955. We would like to thank our participants, Master students of Group1 and the members of the HCI and LIA laboratories at EPFL, for their active participation in all phases of the experiment and, additionally, the open discussions and short interviews that followed. The experiment was made possible with the help of the Pedagogical Research and Support Center at EPFL (CRAFT) by providing the eye tracking infrastructure. Most importantly, we are grateful to Prof. Pierre Dillenbourg and Patrick Jermann for their attentive coordination and valuable feedback. Special thanks are addressed to Prof. Alain Wegmann for his continuous support in pursuing this publication.

## REFERENCES


Thomas Agotnes, T., Wiebe van der Hoek and Michael Wooldridge. 2007. Reasoning about Judgment and Preference Aggregation. In *Proceedings of the International Joint Conference on Autonomous Agents and Multi-agent systems*. ACM Press. DOI:http://dx.doi.org/10.1145/1329125.1329228

Jacob Baskin and Shriram Krishnamurthi. 2009. Preference Aggregation in Group Recommender Systems for Committee Decision-Making. In *Proceedings of the ACM Conference on Recommender Systems*. ACM Press, 337-340. DOI:http://dx.doi.org/10.1145/1639714.1639782

Stacey Birkett, Adam Galpin, Simon Cassidy, Lynne Marrow and Sarah Norgate. 2011. How revealing are eye-movements for understanding web engagement in young children. In *Proceedings of the Human Factors in Computing Systems Conference*. ACM Press, New York, 2251-2256. DOI:http://dx.doi.org/10.1145/1979742.1979900

Philip Bonhard, Clare Harries, John McCarthy and Angela M. Sasse. 2006. Accounting for Taste: Using Profile Similarity to Improve Recommender Systems. In *Proceedings of the Human Factors in Computing Systems Conference*. ACM Press, 1057-1066. DOI:http://dx.doi.org/10.1145/1124772.1124930

Jiajun Bu, Shulong Tan, Chun Chen, Can Wang, Hao Wu, Lijun Zhang and Xiaofei He. 2010. Music Recommendation by Unified hyper graph: Combining Social Media Information and Music Content. In *Proceedings Multimedia*. ACM Press. DOI:http://dx.doi.org/10.1145/1873951.1874005

Peng Cui, Fei Wang, Shiqiang Yang and Lifeng Sun. 2011. Who Should Share What? Item-level Social Influence Prediction for Users and Posts Ranking. In *Proceedings of the 34th ACM SIGIR International Conference on Research and Development in Informational Retrieval*. ACM Press, 185-194. DOI:http://dx.doi.org/10.1145/2009916.2009945

Sylvain Castagnos, Nicolas Jones and Pearl Pu. 2009. Recommenders' Influence on Buyers' Decision Process. In *Proceedings of the ACM Conference on Recommender Systems*. ACM Press, 361-364. DOI:http://dx.doi.org/10.1145/1639714.1639788

Sylvain Castagnos, Nicolas Jones and Pearl Pu. 2010. Eye-tracking Product Recommenders' Usage. In *Proceedings of the ACM Conference on Recommender Systems*. ACM Press, Barcelona, Spain, 29-36. DOI:http://dx.doi.org/10.1145/1864708.1864717

Sylvain Castagnos and Pearl Pu. 2010. Consumer Decision Patterns Through Eye Gaze Analysis. In *Proceedings of the International Conference on Intelligent User Interfaces*. ACM Press, Hong Kong, 78-85. DOI:http://dx.doi.org/10.1145/2002333.2002346

Òscar Celma and Pedro Cano. 2008. From Hits to Niches? or how Popular Artists can Bias Music Recommendation and Discovery. In *Proceedings KDD Workshop on Large-Scale Recommender Systems*. ACM Press. DOI:http://dx.doi.org/10.1145/1722149.1722154

Hung-Chen Chen and Arbee L.P. Chen. 2001. A Music Recommendation System Based on Music Data Grouping and User Interests. In *Proceedings of the Information and Knowledge Management Conference*. ACM Press, 231-238. DOI:http://dx.doi.org/10.1145/502585.502625

Li Chen and Pearl Pu. 2011. Users' Eye Gaze Pattern in Organization-based Recommender Interfaces. In *Proceedings of the International IUI Conference*. ACM Press, Palo Alto, California, USA, 311-314. DOI:http://dx.doi.org/10.1145/1943403.1943453

Li Chen and Pearl Pu. 2010. Eye-Tracking Study of User Behavior in Recommender Interfaces. In *P. De Bra, A. Kobsa, and D. Chin* (Eds.): UMAP, LNCS 6075, Hawaii, USA, 375-380. DOI:http://dx.doi.org/10.1007/978-3-642-13470-8_35

Yan Chen, F. Maxwell Harper, Joseph Konstan and Sherry Xin Li. 2010. Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens. In *American Economic Review* 100, 4. 1358-1398. DOI:http://dx.doi.org/10.1257/aer.100.4.1358

Robert B. Cialdini and Noah J. Goldstein. 2004. Social Influence: Compliance and conformity. In *Annual Review of Psychology* 55. 591-621. DOI:http://dx.doi.org/10.1146/annurev.psych.55.090902.142015

Robert B. Cialdini and Melanie R. Trost. 1998. Social Influence: Social Norms, Conformity, and Compliance. In *The Handbook of Social Psychology*, D.T. Gilbert, Fiske, S.T., Lindzey, G. (Eds.). Oxford Univ. Press, 151-192.

Gautier Drusch, J. M. Christian Bastien, Jérôme Dinet. 2011. From Gaze Plots to Eye Fixation Patterns using a Clustering Method based on Hausdorff Distances. In *Proceedings of the 23rd French Speaking Conference on Human-Computer Interaction*. ACM Press. DOI:http://dx.doi.org/10.1145/2044354.2044356

Susan Dumais, Georg Buscher and Edward Cutrell. 2010. Individual Differences in Gaze Patterns for Web Search. In *Proceedings of the Symposium on Information Interaction in Context*. ACM Press, 185-194. DOI:http://dx.doi.org/10.1145/1840784.1840812

John A. Fleishman. 1998. The Effects of Decision Framing and Others' Behavior on Cooperation in a Social Dilemma. In *Journal of Conflict Resolution* 32, 1. 162-180.

Mike Gartrell, Xinyu Xing, Qin Lv, Aaron Beach, Richard Han Shivakant Mishra and Karim Saeda. 2010. Enhancing Group Recommendation by Incorporating Social Relationship Interactions. In *Proceedings* *of the International Supporting Group Work Conference*. ACM Press, 97-106. DOI:http://dx.doi.org/10.1145/1880071.1880087

Davide Grossi. 2009. Unifying Preference and Judgment Aggregation. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems*, 1. ACM Press, 217-224.

Hyunmo Kang, Lise Getoor and Lisa Singh. 2007. Visual Analysis of Dynamic Group Membership in Temporal Social Networks. In *Proceedings of Knowledge Discovery and Data Mining Conference*. ACM Press, 13-21. DOI:http://dx.doi.org/10.1145/1345448.1345452

Xingjie Liu, Yuan Tian, Mao Ye and Wang-Chien Lee. 2012. Exploring Personal Impact for Group Recommendation. In *ACM International Conference on Information and Knowledge Management*, Maui, Hawaii. ACM Press, 674-683. DOI:http://dx.doi.org/10.1145/2396761.2396848

Judith Masthoff. 2005. The Pursuit of Satisfaction: Affective State in Group Recommender Systems. In *Computer Science*, Vol. 3538. 297-306. DOI:http://dx.doi.org/10.1007/11527886_39

Jessica M. Nolan, P. Wesley Schultz, Robert B. Cialdini, Noah J. Goldstein and Vladas Griskevicius. 2008. Normative Social Influence is Underdetected. In *Personality and Social Psychology Bulletin* 34, 7, 913-923. DOI:http://dx.doi.org/10.1177/0146167208316691

George Popescu and Pearl Pu. 2011. Probabilistic Game Theoretic Algorithms for Group Recommender Systems. In *Proceedings of the Workshop on Music Recommendation and Discovery*, RecSys. ACM Press.

George Popescu and Pearl Pu. 2012. What's the Best Music You Have? Designing Music Recommendation for Group Enjoyment in GroupFun. In *Proceedings of the Human Factors in Computing Systems Conference*. ACM Press, 1673-1678. DOI:http://dx.doi.org/10.1145/2212776.2223691

George Popescu and Pearl Pu. 2012. A Different View on Group Analysis – (Eye-) Tracking Social Influence. In *EPFL Technical Report*. Lausanne, Switzerland.

Tom Postmes, Russell Spears, Khaled Sakhel and Daphne de Groot. 2001. Social Influence in Computer-Mediated Communication: the Effects of Anonymity on Group Behavior. In *Personality and Social Psychology Bulletin* 27, 10, 1243-1254. DOI:http://dx.doi.org/10.1177/01461672012710001

Rui Prada, Nuno Otero and Ana Paiva. 2007. The User in the Group: Evaluating the Effects of Autonomous Group Dynamics. In *Proceedings of Advances in Computer Entertainment Technology Conference*. ACM Press, 25-32. DOI:http://dx.doi.org/10.1145/1255047.1255054

Marco Pretorius, Helene Gelderblom and Bester Chimbo. 2010. Using Eye Tracking to Compare how Adults and Children Learn to use an Unfamiliar Computer Game. In *Proceedings of the South African Institute of Computer Scientists and Information Technologists Conference*. ACM Press, 275-283. DOI:http://dx.doi.org/10.1145/1899503.1899534

Pearl Pu, Li Chen and Rong Hu. 2011. A User-Centric Evaluation Framework for Recommender Systems. In *Proceedings of the ACM Conference on Recommender Systems*. ACM Press,Chicago, IL, USA, 157-164. DOI:http://dx.doi.org/10.1145/2043932.2043962

Stephen Purpura, Victoria Schwanda, Kaiton Williams, William Stubler and Phoebe Sengers. 2011. Fit4Life: The Design of a Persuasive Technology Promoting Healthy Behavior and Ideal Weight. In *Proceedings of the Human Factors in Computing Systems Conference*. Vancouver, BC, Canada. ACM Press, 423-432. DOI:http://dx.doi.org/10.1145/1978942.1979003

Kai Sassenberg. 2002. Common Bond and Common Identity Groups on the Internet: Attachment and Normative Behavior in On-Topic and Off-Topic Chats. In *Group Dynamics* 6, 1. 27-37.

Abhay Sukumaran, Stephanie Vezich, Melanie McHugh and Clifford Nass. 2011. Normative Influences on Thoughtful Online Participation. In *Proceedings of the Human Factors in Computing Systems Conference*, ACM Press, Vancouver, BC, Canada, 3401-3410. DOI:http://dx.doi.org/10.1145/1978942.1979450

Jie Tang, Jimeng Sun, Chi Wang and Zi Yang. 2009. Social Influence Analysis in Large-Scale Networks. In *Proceedings of the Knowledge Discovery and Data Mining Conference*. ACM Press, 807-816. DOI:http://dx.doi.org/10.1145/1557019.1557108

John C. Turner. 1991. Social Influence. Wadsworth.

Peter-Paul van Maanen, Tomas Klos and Kees van Dongen. 2007. Aiding Human Reliance Decision Making Using Computational Models of Trust. In *Proceedings Web Intelligence and Intelligent Agent Technology Conference*. ACM Press, 372-376. DOI:http://dx.doi.org/10.1109/WI-IATW.2007.108

Mao Ye, Xingjie Liu and Wang-Chien Lee. 2012. Exploring Social Influence for Recommendation - A Generative Model Approach. In *International Conference on Research & Development on Information Retrieval*. Portland, USA. ACM Press, 671-680. DOI:http://dx.doi.org/10.1145/2348283.2348373

Haiyi Zhu, Bernardo Huberman and Yaron Luon. 2012. To Switch or not to Switch: Understanding Social Influence in Online Choices. In *Proceedings of the Human Factors in Computing Systems Conference*. ACM Press, 2257-2266. DOI:http://dx.doi.org/10.1145/2207676.2208383