To Explain or not to Explain: the Effects of Personal Characteristics when Explaining Music Recommendations

Martijn Millecamp
Department of Computer Science, KU Leuven
Leuven, Belgium
martijn.millecamp@cs.kuleuven.be

Nyi Nyi Htun
Department of Computer Science, KU Leuven
Leuven, Belgium
nyinyi.htun@cs.kuleuven.be

Cristina Conati
Department of Computer Science, UBC
Vancouver, Canada
conati@cs.ubc.ca

Katrien Verbert
Department of Computer Science, KU Leuven
Leuven, Belgium
katrien.verbert@cs.kuleuven.be

ABSTRACT

Recommender systems have been increasingly used in online services that we consume daily, such as Facebook, Netflix, YouTube, and Spotify. However, these systems are often presented to users as a "black box", i.e. the rationale for providing individual recommendations remains unexplained to users. In recent years, various attempts have been made to address this black box issue by providing textual explanations or interactive visualisations that enable users to explore the provenance of recommendations. Among other things, results demonstrated benefits in terms of precision and user satisfaction. Previous research had also indicated that personal characteristics such as domain knowledge, trust propensity and persistence may also play an important role on such perceived benefits. Yet, to date, little is known about the effects of personal characteristics on explaining recommendations. To address this gap, we developed a music recommender system with explanations and conducted an online study using a within-subject design. We captured various personal characteristics of participants and administered both qualitative and quantitative evaluation methods. Results indicate that personal characteristics have significant influence on the interaction and perception of recommender systems, and that this influence changes by adding explanations. For people with a low need for cognition are the explained recommendations the most beneficial. For people with a high need for cognition, we observed that explanations could create a lack of confidence. Based on these results, we present some design implications for explaining recommendations.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **Information visualization**; *User models*; *User interface design*; *Visualization design and evaluation methods*; • **Social and professional topics** → **User characteristics**; • **Information systems** → **Personalization**; **Recommender systems**.

KEYWORDS

recommender system; explanations; personal characteristics; music; Spotify; user characteristics; need for cognition

ACM Reference Format:
Martijn Millecamp, Nyi Nyi Htun, Cristina Conati, and Katrien Verbert. 2019. To Explain or not to Explain: the Effects of Personal Characteristics when Explaining Music Recommendations. In *24th International Conference on Intelligent User Interfaces (IUI '19), March 17–20, 2019, Marina del Rey, CA, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3301275.3302313

IUI '19, March 17–20, 2019, Marina del Rey, CA, USA

ACM ISBN 978-1-4503-6272-6/19/03…$15.00
https://doi.org/10.1145/3301275.3302313

1 INTRODUCTION

In the field of both Human-Computer Interaction and Recommender Systems, personal characteristics have been researched substantially. For instance, Knijnenburg et al. [29] have shown that the preference for interaction methods in recommender systems depends on several personal characteristics such as domain knowledge, trust propensity and persistence. In the Music Recommender Systems domain, the effects of both personal characteristics and controllability have been investigated on acceptance, perceived diversity and trust [26, 27, 34]. In recent years, the idea of explaining the inner process of recommender systems to users has attracted increased research interest. For example O'Donovan et al. [37] argued that existing recommender systems are not transparent to users to be able to make informed decisions and thus designed PeerChooser, a film recommender system providing users with a visual explanation of the recommendation process and the ability to steer recommendations [37]. Smallworlds made the recommendation process transparent by providing visual explanations of social links and similarities between the user and others [22]. In the Music Recommender Systems domain, TasteWeights [5] provides visual connections between users' favourite artists and recommended trending item. However, to the best of our knowledge, no previous studies have investigated the effects of personal characteristics on the need or effectiveness of visual explanations in recommender systems. A few researchers [14, 15, 43, 45] have investigated the effects of personal characteristics on the perception of visualisations, but did not focus on explanations in recommender systems.

To address this gap, we present in this paper the results of an online user study using a within-subject design where we researched the effect of personal characteristics on two versions of a music recommender system: one with and one without explanations.

To conduct this study, we built a music recommender system which uses the Spotify API [1] on top of existing music recommender systems [27, 34]. The design provides bar charts and a scatter plot to explain and compare the recommendations.

The specific user characteristics that we tested in the study include a personality trait (locus of control [41]), a variety of cognitive skills (need for cognition [10], visualisation literacy [6] and visual working memory [28]), and abilities related to the specific task of using a music recommender (musical sophistication [36], Spotify usage and tech savviness [34]).

The objective of the study was to answer the following research questions:

RQ1: How do personal characteristics impact user perceptions of the system when recommendations are explained?

RQ2: How do personal characteristics impact user interaction with the system when recommendations are explained?

The results of our study show that explanations are raising the confidence of users with a low need for cognition in contrast to users with a high need for cognition. For the latter group explanations seem to be lowering their confidence. Additionally, there are main effects for visualisation literacy and musical sophistication.

The main contribution of this paper is twofold: first, we address the gap between the research about personal characteristics and the research about explanations in recommender systems. Second, we provide evidence that personal characteristics have a significant influence on the interaction and perception of explanations.

This paper is organised as follows: we first discuss previous work related to personal characteristics and explanations in recommender systems. We then describe our music recommender system and the design of the experiment in detail. Next, we present the results of the study and discuss the implications of our findings. Concluding remarks as well as limitations of this study finalise this paper.

2 BACKGROUND

2.1 Personal characteristics

Previous research [14, 15, 43, 45] has investigated the influence of personal characteristics on the perception of visualisation, but we have yet to understand the influence on recommender explanations. In this study, we captured a number of different personal characteristics, including some suggested by [3]. In the following sub-sections, we present each of these characteristics in detail.

2.1.1 Locus of Control. *Locus of control* (LOC) is defined by Fourier [18] as "the extent to which people believe they have power over events in their lives." Users with an internal locus of control may believe that they have control over events and their outcomes, whereas users with an external locus of control tend to submit to outside forces. We used the LOC test proposed by Rotter et al. [41], which measures the degree of control individuals perceive towards outcomes. LOC was an important characteristics to measure for this study since our experimental system provided options to control and steer the recommend process (see the Implementation section for details of the system).

2.1.2 Need for Cognition. Cognitive skills have been investigated by numerous previous work [1, 8, 11, 14, 17, 46, 47]. Unlike measuring cognitive skills, *need for cognition* (NFC) is a measure of the tendency for an individual to engage in, and enjoy, effortful cognitive activities. Cacioppo et al. [10] presented a test for NFC, which comprises of 18 questions on a 5-point Likert scale. We used this test because we were interested in participants' natural willingness to engage with interface components when creating playlists for different occasions.

2.1.3 Visualisation Literacy. *Visualisation literacy* (VisLit) is the ability to interpret and to make meaning from information presented in the form of images and graphs. Boy et al. [6] proposed a test to check VisLit, which is designed in a way that it captures an individual's ability to confidently use a graph to translate data and interpret visual patterns. Boy et al. created a number of tests for different visualisation techniques, including bar charts and line graphs. We selected the specific test designed for bar charts, as we used bar chart the most for explanations in our system.

2.1.4 Visual Working Memory. Working memory is the part of our cognitive system that is responsible for short-term holding of information for further processing [35] and it can be categorised into visual and verbal working memory. Previous work has found that visual working memory can influence cognitive load when interacting with visual information systems [14, 32, 45]. As the main focus of our study involved visually explaining recommendations, we measured *visual working memory* (VM) of our participants. Following the approach of Millecamp et al. [34] and Jin et al. [27], VM was measured using a Corsi block-tapping technique[2], which captures the longest sequence of block tapping a participant can correctly repeat.

2.1.5 Musical Experience. Musical experience of a participant may also influence the way they interact with music recommender systems [27]. The Goldsmiths Musical Sophistication Index (Gold-MSI) [36] has been considered an effective way to measure domain knowledge of users and has shown a strong correlation with personality [21] and music preferences [36]. Previous research in music recommender systems [26, 27, 34] have also used the Gold-MSI as a factor of personal characteristics. Similarly, in this study, we used a total of 18 Gold-MSI 7-point linear scale questionnaires to measure individual participants' *musical sophistication* (MS). The minimum and maximum possible MS scores are 18 and 126 respectively [36]. Since we used the Spotify platform to provide recommendations, we additionally measured the participants' *Spotify usage* in number of hours per week, similar to [34].

2.1.6 Tech-Savviness. We define *tech-savviness* of participants as their confidence in trying out new technology. Previous studies have investigated a similar personal characteristic, namely internet-savviness [19, 50]. However, these studies focused on participants' familiarity with the internet rather than their confidence in trying out new technology. It has been shown that tech-savviness can influence the behaviour of mobile device usage [16]. We believe

[1] https://developer.spotify.com/documentation/web-api/

[2] https://www.psytoolkit.org/experiment-library/corsi.html

that tech-savviness also plays an important role in influencing the way participants interact with new music recommender interfaces. Therefore, using the approach of Millecamp et al. [34], we captured tech-savviness by asking the participants to identify themselves between *confident*, *not confident* and *somewhere in-between* when it comes to trying new technology.

2.2 Explanations in Recommender Systems

Recommender systems rely on different algorithms to predict items a user might like. Although a number of algorithms have been proposed to provide the best possible suggestions for users, insights into the logic or justification of the recommendations are not transparent to users [42]. This raises an issue of not understanding why certain recommendations are being provided which in turn could lead to doubts against the recommendations [23].

Tintarev and Masthoff [44] listed seven possible aims of an explanation facility that have been found in the existing systems, namely: transparency, scrutability, trust, effectiveness, persuasiveness, efficiency and satisfaction. Although not all of the reviewed systems have focused on facilitating all seven aims, the types of explanations fall under three categories: content-based explanation (e.g. "We have recommended X because you liked Y"), collaborative-based explanation (e.g. "People who liked X also liked Y") and preference-based explanation (e.g. "Your interests suggest that you would like X").

In addition to textual explanations, a number of other researchers have looked into visual explanations. Interactive visualisations allow users to not just understand, but also to manipulate recommender components that may affect the way recommendations are presented. For instance, by visualising user profiles and the recommendation process, transparency and user control of the system can be significantly improved [25]. Donovan et al. [37] designed a movie recommender system, PeerChooser, to provide users with a visual explanation of the recommendation process and the opportunity to manipulate input weights to steer the recommendations. SmallWorld, a social recommender system designed by Gretarsson et al. [22], is similar to PeerChooser in that it visually explains social connections between recommended users using a node-link diagram. Another social recommender system, SFViz (Social Friends Visualisation) [20], used a radial space-filling technique [13] to show relationships between users and their interests in order to suggest potential new friends with similar interests. Verbert et al. [51] presented a system that increases the effectiveness of decision-making by explaining the provenance of recommendations and offering control to users. Tsai and Brusilovsky [48] designed a diversity-enhanced recommender system that visualised recommendations with multiple attributes in a two-dimensional scatter-plot, which inspired our visualisation approach.

In the Music Recommender domain, TasteWeights [5] provides visual explanations for user's favourite artists and recommended musics that are based on trending topics across social media and Wikipedia. MoodPlay [2] is an emotion based music recommender system. It allows users to explore music by inspecting the explanation of recommendations and by modifying affective data, which could further increase acceptance and understanding of recommendations process. Jin et al. [27] designed a Spotify-based music recommender system and investigated the effects of personal characteristics on different levels of controllability. Unlike the work of Jin et al. [27], we focus on scrutable explanations rather than only on user control.

3 IMPLEMENTATION

For this study, we implemented a music recommender system using the Spotify Web API to query artists and recommendations. We designed two different versions of the system: 1) an interface without explanations (baseline) and 2) an interface with explanations (explanation interface). Figure 1 shows the explanation interface which is comprised of seven different components: a) artists, b) preference, c) list of recommendations, d) bar chart , e) scatter plot, f) number of rated songs in a sidebar, and g) list of rated songs in a sidebar. The only difference between the baseline and explanation interfaces were parts d and e. In the next sub-sections, we describe the components in detail.

3.1 Artist

On the upper left side, Figure 1(a), users can search for an artist or select one of the artists that is already in the list. This list is based on the user's listening history in Spotify and shows their five most favourite artists. The user can select one artist who is used as a seed for recommendations.

3.2 Preference

Under the artist component, Figure 1(b), six different musical attributes are displayed as sliders, namely acousticness, danceability, energy, instrumentalness, tempo and valence. The attributes are part of 14 different musical attributes offered by the Spotify recommendation API and are selected based on the results of Millecamp et al. [34]. Definitions for these attributes are displayed in the demonstration stage and are also available in the interface with a simple mouse hover on the ❓ icon next to the attributes.

3.3 List of Recommendations

Using the selected artist and attribute values as the seeds, recommendations are queried from Spotify[3] and presented in a list of ten songs at a time, as presented in Figure 1(c). The list layout was selected because it is the most applied method for presenting recommendations [49]. In this list, we showed the recommended songs by the title and the artist. The user can click on 👍 to like a song, ▶ to listen to a thirty seconds preview of the song and 👎 to dislike a song. In the explanation interface, there is also a `Why` button for each song. Clicking on this button shows or hides a bar chart which is part of the explanation component. The `Why` buttons are not present in the baseline interface.

3.4 Explanations

To explain why a song is recommended, we used a grouped bar chart, as seen in Figure 1(d), which shows a comparison of the attributes between the selected song and the user's preference. For each of the six attributes, we showed the value of the user's

[3] https://developer.spotify.com/web-api/get-recommendations

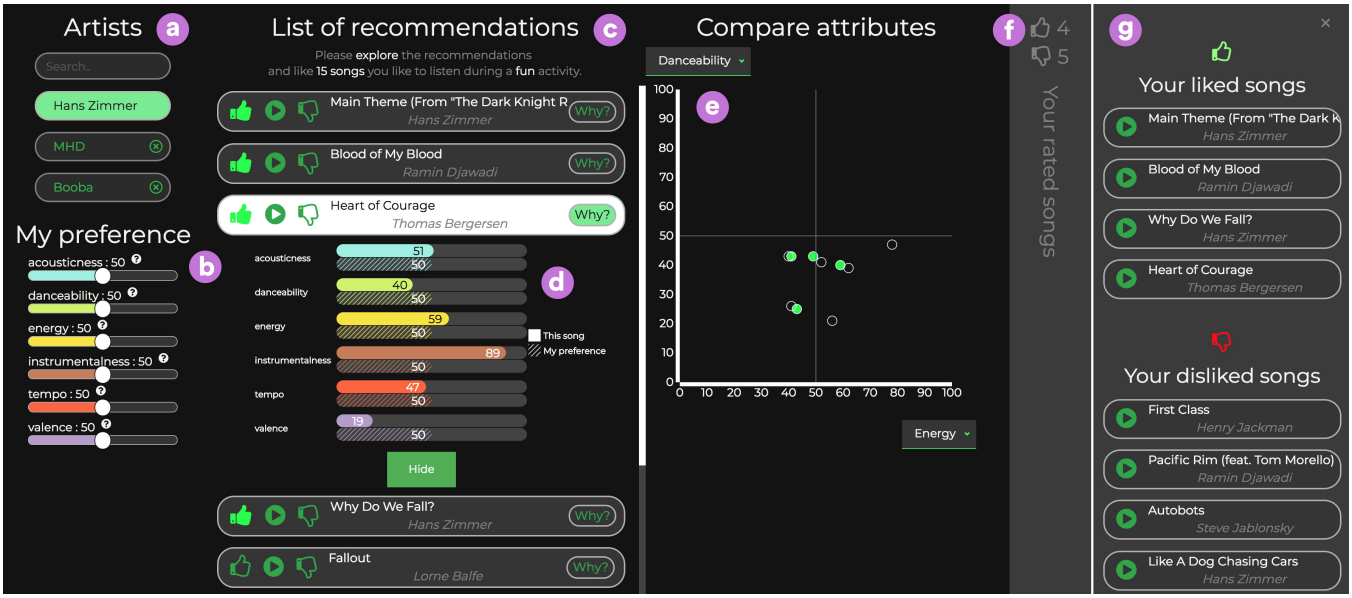


Figure 1: The explanation interface highlighting all its components: a) artists, b) preference, c) list of recommendations, d) bar chart, e) scatterplot, f) rated songs in a sidebar, and g) list of rated songs in a sidebar.

preference (striped bar) and the value of the the selected song (solid-coloured bar). The colours of the bars match those of the sliders from the *preference* component in order to show their relations. We chose for bar charts because they are one of the most popular and effective techniques to compare different values [46]. Under the bar chart, there is a button `More/Hide` that shows or hides a scatter plot.

With the bar charts, users can compare their attribute preference with attributes of only one recommended song, but they cannot visualise differences between multiple recommended songs. To enable comparison of different attributes of all the recommended songs, we used a scatter plot, as it provides an intuitive way to present multidimensional data in an interactive way [49]. As seen in Figure 1 (e), the recommended songs are displayed in a two dimensional plot where users can select the attributes they wish to see on the X and Y axes. The songs that have been liked are shown in green, and the user's attribute preference is shown by two lines forming a cross. Hovering over a songs highlights the song in the recommendation list and vice versa.

3.5 Side Bar

To assist the users with their task of liking 15 songs, users can see the number of songs they liked and disliked, see Figure 1(f). Clicking on this bar shows a list of all these songs as presented in Figure 1(g). Users are also able to play the songs in this list to enable them to get an impression of the *"sound"* of their playlist.

4 EXPERIMENTAL DESIGN

In the following sub-sections, we describe in more detail the participants, procedure and measurements of our study.

4.1 Participants

All participants of this study were recruited via Amazon Mechanical Turk (MTurk)[4]. We required them to have a Spotify account and to fill out the study within one hour. A total of 105 participants completed the study. Out of these 105 participants, we removed 34 participants due to the inconsistencies in answers to questions which were built in to verify the validity of the responses. As a result, we had 71 valid participants (22 females and 49 males) who were rewarded with $2 via the MTurk platform. A total of one hour was allowed to complete the study although most participants completed within 30 minutes. The distribution of the participants across the demographic and personal characteristics are highlighted in Table 1.

4.2 Study Procedure

As explained in the Implementation section, we designed two different versions of the system, one with and one without explanations. Participants evaluated both interfaces following a within-subject design with a Latin-Square counterbalancing method. To optimise study duration, tasks, questionnaires and interface components, a pilot study was conducted with 20 participants. The pilot study indicated that creating a playlist of 15 songs was optimal to evaluate each interface and that a time limit of one hour was realistic. We also found that recommending up to 10 songs at a time was the best approach as it reduces overloading of information.

In the next paragraph, we will explain in detail the different steps of the study. When recruiting participants, a URL to the study was provided with the recruitment posts on MTurk. Once the URL was loaded, participants were presented with detailed information about the study and a consent form. If they agreed to participate

[4]https://www.mturk.com

Demography/PC	Category	Frequency	Mean	SD
Age	15-24	13		
	25-34	28		
	35-44	21		
	45-54	7		
	55-64	0		
	65+	2		
Gender	Male	49		
	Female	22		
Tech-savviness	Low	0		
	Between	8		
	High	63		
Spotify Usage	0-5 hr/week	40		
	6-10 hr/week	15		
	11-15 hr/week	10		
	16-20 hr/week	5		
	21+	1		
MS	low	34	67.634	20.051
	high	37		
VWM	low	20	5.93	1.477
	high	51		
LOC	low	40	12.099	4.583
	high	31		
NFC	low	35	62.845	14.962
	high	36		
VisLit	low	21	-0.073	0.977
	high	50		

Table 1: The distribution of 71 participants across the demographic and personal characteristics (PC). MS: musical sophistication, VWM: visual working memory, LOC: locus of control, NFC: need for cognition, VisLit: visualisation literacy, and SD: standard deviation. High and low categories for PC are divided at median.

in the study, they were asked to authorise their Spotify account to the system in the next step, which allowed the system to access their favourite artists and to create a playlist for them. After authorisation, participants were presented with a series of questionnaires which included demographic questions (e.g. age and gender) and personal characteristics questions (e.g. tech-savviness, Spotify usage, MS, VWM, LOC, NFC and VisLit).

After completing the questionnaires, participants were given a video demonstration of the interface components (i.e. either baseline or explanation interface depending on the counterbalanced rotation), and the task they needed to perform using the interface. We devised two different realistic tasks to accommodate the two different interfaces. The tasks asked the participants to create a playlist of 15 songs for either 1) a fun activity or 2) a relaxing activity. The order in which the tasks were presented was rotated independently from the interfaces using a Latin-Square counterbalancing method.

Next, participants performed their given task with the first interface. This included selecting their favourite artist(s), manipulating the musical attributes and liking (or disliking) the recommended songs until they had collected 15 favourite songs for the given task. After they had liked at least 15 songs, participants were given an option to keep the playlist in their Spotify account. After completing the task, they were presented with a number of evaluation questions which were based on similar previous work [9, 12, 30, 39]. Details of the post-task questions are explained in the Measurements section.

Once the first interface had been evaluated, participants were demonstrated with a different interface and task. The remaining steps were exactly the same as the first iteration. After participants had evaluated the second interface, they were provided with a series of *comparative questions* between the baseline and explanation interfaces (see the Measurements section for details).

4.3 Measurements

We performed measurements of personal characteristics and evaluation outputs in this study. An exhaustive list of personal characteristics captured include: Spotify usage, MS, VWM, LOC, NFC and VisLit. Detailed descriptions of the personal characteristics can be found in the Related Work section. During data analysis, the personal characteristics were used as independent variables. Dependent variables included evaluation outputs measured for each interface using both qualitative and quantitative metrics. These metrics included a total of 15 Likert scale questionnaire items, 4 comparative questionnaire items, 3 open-ended questions and 5 interaction log outputs. In the following paragraphs, we describe these evaluation metrics in detail:

4.3.1 Post-task Questions. Post-task questions were administered after performing a task with each interface. These included 5-point Likert scale questions which were based on a number of previous work [9, 12, 30, 39] and focused on capturing participants' perception towards 7 different aspects of the system, namely: recommender effectiveness, good understanding of recommender rationale, trust, novelty, use intentions, satisfaction and confidence. An overview of the questions is shown in Table 2. The answer for each question ranged from 1 (strongly disagree) to 5 (strongly agree), with 3 being neither agree nor disagree.

Recommender effectiveness was measured by how well the participants agreed that the recommended songs matched their interests and that the system helped them find good songs for creating the playlist. *Good understanding* of recommender rationale was measured by how well the participants agreed that they understood why the songs were recommended, thought the information was sufficient to make a good decision and understood the recommended songs' attributes in relation to their preferences. *Trust* was measured simply based on how well the participants agreed that they trusted the system to suggest good songs. *Novelty*, *use intentions*, *satisfaction* and *confidence* were respectively measured by how well the participants agreed that the system helped them find new songs, that they will use the system again, that they were overall satisfied with the system, and that they were confident about the playlist they had created.

Besides these questions, we also added five additional questions for the explanation interface to capture the perceptions towards the explanation components. These questions are as follows:

- Being able to ask why a song is recommended improved my experience with the recommender system (Expl-Q1).

Metric	Question(s)
Recommender Effectiveness	The songs recommended to me match my interests. The recommender helped me find good songs for my playlist.
Good Understanding	I understood why the songs were recommended to me. The information provided for the recommended songs is sufficient for me to make a decision for my playlist. The songs recommended to me had similar attributes to my preference.
Trust	I trust the system to suggest good songs.
Novelty	The recommender system helped me discover new songs.
Use Intentions	I will use this recommender system again.
Satisfaction	Overall, I am satisfied with the recommender system.
Confidence	I am confident about the playlist I have created.

Table 2: An overview of 5-point Likert scale questionnaires designed to capture recommender effectiveness, good understanding of recommender rationales, trust, novelty, use intentions, satisfaction and confidence

- The bar chart explaining why a song was recommended was easy to understand (Expl-Q2).
- The scatterplot comparing songs based on specific attributes was easy to understand (Expl-Q3).
- The bar chart explaining why a song was recommended was useful (Expl-Q4).
- The scatterplot comparing songs based on specific attributes was useful (Expl-Q5).

Following the Likert scale questions, a pair of optional open-ended questions were presented to capture a broad range of feedback regarding each interface. Specifically, these questions asked the participants to describe parts of the interface that were the most and least useful to them. Answers to these questions were analysed using thematic analysis [7].

4.3.2 Post-evaluation Comparative Questions. The post-evaluation questions were designed to understand whether the participants preferred the baseline or explanation interface in relation to a set of statements. For each of the following statements, the participants had to choose an answer between "with explanation" and "without explanation".

- If I had to choose, I would use again the recommender system ___.
- I trust the recommender system ___ to suggest good songs.
- I am the most confident about the playlist I have created in the recommender system ___.
- The songs recommended to me match my preferences the best in the recommender system ___.

4.3.3 Interaction Logs. Participants' interactions with interface components were captured using mouse-clicks together with their timestamps. Based on this log data, we then calculated the following metrics for both interfaces:

- *nb_slider:* the number of times a participant interacted with the sliders.
- *precision*: the ratio between the number of songs liked and the number of distinct songs seen by a participant. Unlike traditional precision measure, true positives in this metric are determined by individual's perception towards a song in relation to a given task.
- *nb_play*: the total number of songs played divided by the total amount of time spent by a participant in a given interface.

In addition to these, we calculated a few metrics that were unique to the Explanation interface, namely:

- *nb_why*: the number of click on the "Why" button by a participant.
- *du_why*: the duration, in minutes, the "Why" feature was used by a participant.

5 RESULTS

5.1 Statistical Analysis

Since we used a within-subject design, participants were asked to make a playlist both in the baseline and in the interface with explanations. In our analysis, we wanted to verify the relationship between the personal characteristics (LOC, MS, VisLit, NFC, VWM) and the kind of interface. To include both the by-interface and the by-subject variance, a linear mixed effect analysis is a suitable option [14, 52]. For each dependent variable (DV), we ran a linear mixed effect model with the personal characteristics and the kind of interface as fixed-effects. To deal with the individual differences between subjects, we included subjects as a random-effect[5], resulting in the following model:

$$\text{DV} \sim \text{LOC} + \text{MS} + \text{VisLit} + \text{NFC} + \text{VWM} + \text{interface} + (1|\,\text{subject}) \quad (1)$$

To avoid overfitting our models, we performed a backward elimination of random-effect terms followed by a backward elimination of fixed-effects [31][6]. P-values to verify the significant influence of the interface and the personal characteristics on the DV are obtained by a t-test with Satterwaite's method between the full model with the effect in question against the model without the effect in question [31, 52].

To test whether the personal characteristics are significantly inter-dependent on the interface, we performed a likelihood ratio test of the model without interactions against the model with interactions between the interface and the personal characteristics [52]. As we performed this test for each DV and for each of the personal characteristics, we adjusted the obtained p-values with the Benjamini-Hochberg procedure [24, 38]. In the remaining sections, we first discuss the interaction effects we found, followed by the main effects of interface type and personal characteristics.

5.2 Interaction Effects

A significant interaction effect ($\chi^2(1)$=8.73, p=0.003) was found between the need for cognition and the interface in terms of *confidence*. To show the interaction effect, we divided the participants into four quartiles based on their NFC score, as shown in Figure 2.

Figure 2 shows that participants with low NFC are reporting a higher confidence in their playlist created with the explanations interface than created with the baseline. An explanation might be that low NFC participants benefited from the explanations because

[5]We implemented this using `R` [40] and the packages `lme4` [4] and `lmerTest` [31]
[6]More information about this backwards elimination with $\alpha = 0.05$, can be found in the documentation of the `step` function of the `lmerTest` package or in [31]

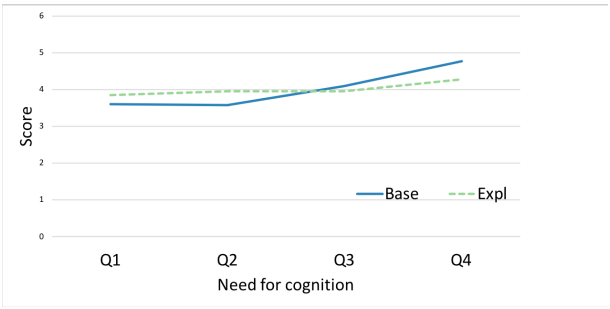
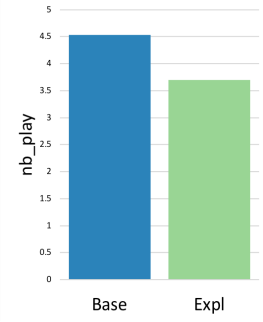
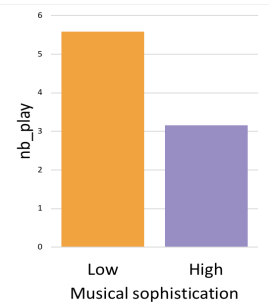


Figure 2: The interaction effect between NFC (divided into 4 quartiles Q1-Q4) and interfaces in terms of the *confidence* metric. The blue solid line indicates the score for the baseline and the green dotted line indicates the score for the explanation interface.



(a) The effect of each interface on *nb_play* showing that participants played more songs in the baseline than in the explanation interface.



(b) The effect of MS on *nb_play* showing that participants with a low MS played more songs than those with a high MS.

Figure 3: The main effect results of a) interfaces and b) MS on *nb_play*.

they did not spontaneously engage in much extra reasoning to justify the recommendations they received. Additionally, when they received the rationale from the explanation, this increased their confidence in their songs selection.

Figure 2 also indicates that as the NFC increased, the confidence of participants in the playlist created in the baseline also increased. This seems to imply that participants with a high NFC were more willing to understand their own musical preference in relation to the attributes of the recommended songs even without explanation. This may have resulted in a higher confidence in their playlist.

We did not see the same increase in trust as NFC increases in the explanation interface. As Figure 2 shows, the NFC scores in the third quartile are almost the same for both interfaces. At the highest NFC level, participants had a higher confidence in the baseline than in the explanation interface. The reduced confidence within the explanation interface could be an indication that users with a high NFC have less need for explanations.

5.3 Main Effect of Explanations

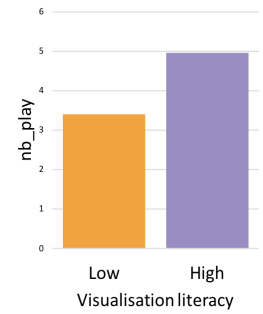
We found a significant effect of interface type on the number of songs a participant played each minute. As Figure 3a shows, participants played fewer songs in the interface with explanations than in the baseline (t=-2.421, p=0.01807). A possible reason for this difference might be that with explanations, participants have insights in the attributes of the recommended songs. This insight may help reduce the need to listen to every song when searching for an ideal song. Another reason could be that they need more time to process the information resulting in playing fewer songs.

5.4 Main Effect of Personal Characteristics

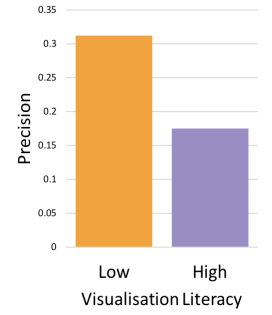
Several main effects of personal characteristics are found on some of our dependent measures, which are discussed in the next section.

5.4.1 Visualisation Literacy. The analysis shows main effects of visualisation literacy on both the number of songs played, *nb_play* (t=2.052, p=0.4408), and the *precision* (t=-2.795, p=0.00592). The direction of these effects is presented in Figures 4a and 4b respectively, where participants are grouped into high and low VisLit, based on a median split. Figure 4a shows that as VisLit increases, the number of songs a participant plays also rises. In contrast, Figure 4b shows that an increase in the VisLit score lowers the precision.

A lower visual literacy score means that one has limited ability to interpret visual patterns. A possible consequence could be that these users need more time to translate visual representation of the explanations and musical attributes in relation to their actual preferences. This increase in processing time could have led to a decrease in *nb_play*. Such delays in processing the visual elements may have also lead to a decrease of engagement, resulting in less exploration and instead clicking the like buttons quickly to compensate for the time lost, or out of frustration (hence, increasing the precision score). Further explanations for these effects are provided in the Discussion section. To gain a better understanding of these effects, we plan to run a user study with eye tracking and



(a) The effect of VisLit on *nb_play* showing that participants with a low VisLit played less songs than those with a high VisLit.



(b) The effect of VisLit on *perceivedprecision* showing that participants with a low VisLit have a higher precision than those with a high VisLit.

Figure 4: The main effect results of VisLit

investigate whether users with a low visualisation literacy need more time to process the explanations and musical attributes.

5.4.2 Musical Sophistication. The last significant effect was the effect of *musical sophistication* on the number of songs played each minute (t=-2.755, p=0.00757). In Figure 3b, we divided the participants into low and high MS based on a median split. Results indicated that as *musical sophistication* increased, the number of songs played was lowered. We assume that participants with a higher MS enjoyed the top majority of recommended songs as they tend to have a broad spectrum of musical taste, resulting in not trying out other songs that may be at bottom of the recommended list.

5.5 Thematic analysis

In addition to the quantitative data analysis discussed above, we also analysed the open-ended questions collected after the post-task questions. To do so, we followed the step-by-step guides of Braun and Clarke [7] and Maguire et al. [33]: first we made ourselves familiar with the the results of the open ended-questions by reading and discussing them. Once we were familiar with the data, we organised the comments in a meaningful and systematic way, and coded each segment of data that was relevant to our research question. In a third stage, we grouped the codes in themes and discussed if these themes made sense. We then iterated a few times, changing the themes until we agreed on four different themes: explanations, musical attributes, quality of recommendations and discoverability. We discuss these themes in the following sub-sections.

5.5.1 Explanations. The first theme reported by nineteen participants in the open-ended questions concerned with "explanations". We analysed these results to verify our hypothesis proposed in the "Interaction Effects" section: namely that low NFC users benefit from the explanations because these users do not spontaneously engage in much extra reasoning to justify the recommendations they receive, and when they get the rationale from the explanation this increases their confidence in their songs selection. To verify this, we divided the users in four quartiles based on their NFC score and compared their reactions.

Three participants with the lowest NFC reported that the why component is the most useful part. Two of them mentioned the explanations in general *"Where it showed you why it was chosen for you"* (P50) and *"I liked the "WHY""* (P57), whereas one user explicitly mentioned the scatterplot as *"The dot chart."* (P56). In addition to these three participants in the first NFC quartile, five other participants in the second NFC quartile reported that explanations were the most useful. Two participants reported that they liked the explanations because they explain why the songs are recommended: *"it explained why a song was being recommended"* (P6) and *"The bars showing why it matched search"* (P11). The three other users reported that they liked the Why component in general: *"why?"* (P35), *"the chart"* (P33) and *"It was really cool seeing the Why? bar graph and scatter plot, cause it kinda lets you compare songs you enjoy."* (P15). For these two groups, only one user reported that the explanation were not needed: *" I didn't really need the Why thing. I didn't find it useful, I know why the thing recommended all the songs."* (P25).

In Figure 2, we also see that for the highest NFC users, the baseline scored higher than the explanation interface. We expected that the spontaneously reasoning about the explanations would give users with a high NFC even more confidence in their playlist in the explanation interface than in the baseline. In the results of the open-ended questions, we saw two possible reasons why explanations were not increasing confidence. The increased cognitive engagement could make the explanations redundant and showing the working of the system could lead to distrust of the system.

Three participants with a high NFC (P9, P52 and P53) reported that they did not need the explanations. They answered respectively: *"The Why function wasn't really necessary. I already know what I like about songs and why and it was obvious enough why the songs were being recommended when I changed the settings."*, *"The why button is not needed."* and *"The Why? question. It's pretty obvious, looking at the genres."*.

In the results, we saw that one participant with a high NFC reported that he distrusted the system: *"[...] now that I see why the system is selecting many songs that I do not want to listen I know that it is not my mistake, but an arbitrary algorithmic decision."* (P70)

5.5.2 Musical attributes. Another theme that came back in 22 answers was the theme "Musical attributes". Based on the results of the baseline in Figure 2, we proposed in section 5.2 that participants with a high NFC are more willing to understand their own musical preference in terms of the provided attributes and that this increased understanding results in a higher confidence in their playlist. To verify this hypothesis, we divided the users in four quartiles based on their NFC score and compared their reactions. The results indicated that low NFC users are not willing to put a lot of cognitive effort to understand their own preferences and that they do not see the added value of the sliders. For the lowest two quartiles, nine participants answered that they did not like the sliders. For example P50 and P49 reported that they just did not like or need the control: *"I didn't like the idea of the sliders on the side. Danceability, tone, etc. It doesn't really help and didn't suggest any great songs because of it."* and *"The fine tuning adjustments. Helpful yes indeed, but at the same time, I don't know if I have a real need to fine tune this stuf."*. One participant reported that he was unsure how to adapt the musical attributes to get better recommendations *"I wasn't sure which of the attributes were causing the lack of songs that I wanted."* (P3). For the high NFC users in the highest two quartiles, only two participants reported that the sliders were not useful: *"The selectors ad too many factors to something that should be simple and automatic."* (P70) and *"the sliders"* (P61).

5.5.3 Novelty. In total, 11 participants reported that the system helped them to find new or forgotten songs and artists. For instance, one participant said *"Broadening my horizon with groups I had not heard of that were thematically similar."* (P45). Another reported something similar: *"The different songs that showed up that I haven't heard of before and ended up liking."* (P49).

Three of the participants reported "novelty" in both the explanations and the baseline. Five participants reported this only for the explanations interface. While explanations are not directly related to discovering new music, we believe that seeing the values of the song in bar chart as well as in a scatter plot could help to discover new songs.

To research the reason that participants with a high musical sophistication are playing less songs (see Figure 3b), we analysed the open-ended questions of participants with a low and high musical sophistication (median split). We see that for the baseline, all participants reporting novelty as the most useful have a low musical sophistication. For the explanation interface, the eight participants reporting novelty are equally divided between high and low musical sophistication. These results seem to indicate that explanations enable especially also high MS users to find novel songs. For the baseline, results seem to indicate that these participants played less songs because they discovered less new music.

5.5.4 Quality of Recommendations. Seventeen participants reported positively about the recommendation component for both the explanation and the baseline interfaces (n=8 and n=7 respectively). For example, P28 mentioned, *"Just the recommendations overall. It was good at guessing similar songs."* and P67 mentioned *"I just really loved the music it chose for me"*. While the number of participants satisfied with the recommendations was almost equal for the two interfaces, this was not the case for unsatisfied participants. Out of the twelve participants that reported that they felt that the system was not recommending good songs, three participants reported this for the explanation interface, while nine participants were not satisfied with the recommendations in the baseline without explanations. These numbers seem to indicate that this difference in perceived accuracy is partly caused by the lack of explanations. The answers of six of the nine participants confirmed this: they reported that the recommendations were not good because they didn't know what attributes to change or because they had the feeling that the recommendations were not following their preference. P3 mentioned *"I wasn't sure which of the attributes were causing the lack of songs that I wanted."* P45 mentioned *"The least useful would be the valence adjustment. I was still a little confused how that altered my preferences."*. Another participant, P23, also reported something similar: *"Some of the sliders didn't seem to make a lot of difference"*.

6 DISCUSSION

In the following sections, we discuss the findings from this study in relation to our research questions.

6.1 Impact on User Perceptions

To answer the first research question (*RQ1: How do personal characteristics impact user perceptions of the system when recommendations are explained?*), we performed a statistical analysis of the post-task questions. We found that there was an interaction effect for need for cognition with confidence. Participants with a low need for cognition were more confident about their playlist when recommendations were explained, as opposed to those with a high need for cognition. The results of the open-ended questions suggests several reasons for this effect. Participants with a low need for cognition benefited from explanations because they did not spontaneously engage in much extra reasoning to justify the recommendations they received, and when they obtained the rationale from the explanation this increases their confidence in their songs selection. In contrast, as users with a high need for cognition spontaneously engaged in reasoning to justify the recommended songs, they had less need for explanation. Otherwise, they became frustrated as they realised that their effort was not rewarded because of the limitations of the system.

Another difference between the participants with a high and low need for cognition was that those with a high need for cognition were more willing to understand their own musical preference in relation to the provided attributes. Previous research has also shown that users with a high need for cognition are more likely to accept recommendations [47].

6.2 Impact on User Interaction

For the second research question (*RQ2: How do personal characteristics impact user interaction with the system when recommendations are explained?*), we looked at the interaction logs to see how personal characteristics impact user interactions with the system when recommendations are explained. We found four main effects: one for the interface, one for musical sophistication and two for visualisation literacy. Three main effects showed a difference for the number of songs a user played per minute. These effects are discussed below in detail.

The results (see Figure 3a) showed that participants played more songs per minute in the baseline than in the explanation interface. This suggests that in the explanation interface, participants were able to find their ideal songs without having to listen quite often. It also shows that participants tend to rely significantly on explanations when judging the songs regardless of their contents.

The second main effect is shown in Figure 3b and indicates that users with a higher musical sophistication tend to listen less when judging the songs. We believe users with a higher musical sophistication have a broad spectrum of taste for music and liked the top majority of songs recommended by the system. Consequently, they failed to explore the remaining songs that may be at bottom of the recommended list.

The third effect, as presented in Figure 4a, indicates that users with a lower visualisation literacy also listened to less songs. We believe that users with lower visualisation literacy may have spent more time interpreting the different visualisations and were not able to listen to as many songs, but this hypothesis will need to be tested with eye tracking data.

The fourth main effect, also concerned with visualisation literacy, shows that a lower visualisation literacy results in a higher precision (see Figure 4b). It is important to note that true positives of the precision metric in this study are determined by the number of songs a user liked (i.e. user's perceived true positive). Therefore, a possible explanation for this could be that since users with a low visualisation literacy were believed to have spent more time interpreting the visualisations, they ended up clicking the like buttons frequently to compensate or out of frustration.

6.3 Design guidelines

As our results show, personal characteristics have an effect on the perception of, and interaction with, explanations. As guidelines for future interfaces, we recommend to take personal characteristics into account when designing explanations. In our study, we find that explaining recommendations could be beneficial because it helps users gain confidence in their choices and assist them to

judge the songs faster. However, explaining recommendations could also lower the confidence of users if they do not need the recommendations or if they see that putting cognitive effort does not always results in better recommendations. Based on these results we suggest three design implications: firstly, like the recommendations themselves, explanations should be personalised for different groups of end-users. Secondly, to reduce information overloading, users should be able to choose whether or not they wish to see explanations. Thirdly, explanation components should be flexible enough to present varying level of details depending on a user's preference.

7 LIMITATIONS

Although immense cares have been put into the planning, our study did not go without limitations. Firstly, we recruited the participants via Amazon Mechanical Turk and found that a large proportion of them reported being confident with trying a new technology. This may have had a bias on results particularly related to *tech-savviness*. Secondly, a different music taste could have affected the accuracy of the recommendations and could have created a bias. Thirdly, we tested the implications of bar chart and scatter plot to facilitate explanations, but there are a number of other possible visualisations that may facilitate explanations. Future research will explore other possible visualisations as means to explain recommendations.

8 CONCLUSION AND FUTURE WORK

To close the gap between personal characteristics and explanations in recommender systems, we performed an online study using a within-subject design. Two versions of a music recommender system were designed for the study: one with and one without explanations. Specifically, we investigated the effects of personal characteristics on user perceptions of and interaction with the system when recommendations are explained. A combination of qualitative and quantitative methods were used to evaluate the different versions of the system.

Our results have shown that there is an interaction effect between need for cognition and confidence. Users with a low need for cognition tend to benefit more from explanations as they help raise the confidence of users on their decisions. For users with a high need for cognition, explanations lower their confidence. In addition to this interaction effect, we also found a main effect of explanation on the number of songs users tend to play each minute. This main effect indicates that explanations assist users to judge the songs faster without having to listen quite often. A second effect shows that users with a higher musical sophistication also tend to judge the songs faster without having to listen quite long, possibly because they have a larger knowledge of songs. Despite this quality, we learned from the qualitative analysis that explaining recommendations could still help users find new songs. The last two main effects concern with visualisation literacy. Results indicate that users with a lower visualisation literacy tend to judge the songs faster without listening as often and, as a result, have a higher precision. A possible explanation could be that these users need more time to process the visual elements, resulting in a decrease of engagement, less exploration, faster liking of songs and higher precision. We are planning to verify this hypothesis in an eye-tracking study.

This paper contributes to the fields of Human-Computer Interaction and Recommender Systems by providing a better understanding of user perceptions towards explanations in recommender systems, and design guidelines that could benefit the design of transparent recommender systems. Future research should investigate the possibilities of other visualisations for explaining recommender systems. In addition, our results indicated that explanations could also lower the confidence of users when creating a playlist. Future research should look into the aspects of explanations that may have produced this effect. To gain further insights into user interaction, a eye tracking technique will also be deployed. Finally, we recommend that explanations, much like recommendations themselves, should be personalised for different end-users. Users may also be allowed to choose the type and level of explanations they prefer to see at any time.

9 ACKNOWLEDGMENTS

Part of this research has been supported by the KU Leuven Research Council (grant agreement C24/16/017).

REFERENCES

[1] Azzah Al-Maskari and Mark Sanderson. 2011. The effect of user characteristics on search effectiveness in information retrieval. *Information Processing & Management* 47, 5 (2011), 719–729.
[2] Ivana Andjelkovic, Denis Parra, and John O'Donovan. 2016. Moodplay: Interactive Mood-based Music Discovery and Recommendation. In *Proc. of UMAP '16*. ACM, 275–279.
[3] Nuray M Aykin and Turgut Aykin. 1991. Individual differences in human-computer interaction. *Computers & industrial engineering* 20, 3 (1991), 373–379.
[4] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01
[5] Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. 2012. TasteWeights: a visual interactive hybrid recommender system. In *Proc. of RecSys'12*. ACM, 35–42.
[6] Jeremy Boy, Ronald A Rensink, Enrico Bertini, and Jean-Daniel Fekete. 2014. A principled way of assessing visualization literacy. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 1963–1972.
[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
[8] Peter Brusilovsky and Eva Millán. 2007. User models for adaptive hypermedia and adaptive educational systems. In *The adaptive web*. Springer, 3–53.
[9] Andrea Bunt, Joanna McGrenere, and Cristina Conati. 2007. Understanding the utility of rationale in a mixed-initiative system for GUI customization. In *International Conference on User Modeling*. Springer, 147–156.
[10] John T Cacioppo, Richard E Petty, and Chuan Feng Kao. 1984. The efficient assessment of need for cognition. *Journal of personality assessment* 48, 3 (1984), 306–307.
[11] Giuseppe Carenini, Cristina Conati, Enamul Hoque, Ben Steichen, Dereck Toker, and James Enns. 2014. Highlighting interventions and user differences: informing adaptive information visualization support. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 1835–1844.
[12] Li Chen and Pearl Pu. 2005. Trust building in recommender agents. In *Proceedings of the Workshop on Web Personalization, Recommender Systems and Intelligent User Interfaces at the 2nd International Conference on E-Business and Telecommunication Networks*. Citeseer, 135–145.
[13] Mei C Chuah. 1998. Dynamic aggregation with circular visual designs. In *Information Visualization, 1998. Proceedings. IEEE Symposium on*. IEEE, 35–43.
[14] Cristina Conati, Giuseppe Carenini, Enamul Hoque, Ben Steichen, and Dereck Toker. 2014. Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. In *Computer Graphics Forum*, Vol. 33. Wiley Online Library, 371–380.
[15] Cristina Conati, Giuseppe Carenini, Dereck Toker, and Sébastien Lallé. 2015. Towards user-adaptive information visualization. In *Proc. of AAAI '15*. AAAI Press, 4100–4106.
[16] Ipshita Dewan and Pierre Benckendorff. 2013. Impact of Tech Savviness and impulsiveness on the mobile information search behaviour of young travellers.

Information and communications technologies in tourism (2013).
[17] Gitta O Domik and Bernd Gutkauf. 1994. User modeling for adaptive visualization systems. In *Visualization, 1994., Visualization'94, Proceedings., IEEE Conference on*. IEEE, 217–223.
[18] G. Fournier. 2018 (Retrieved on September 16, 2018). *Locus of Control. Psych Central*. https://psychcentral.com/encyclopedia/locus-of-control/
[19] Roger W Geyer. 2009. Developing the internet-savviness (IS) scale: Investigating the relationships between internet use and academically talented middle school youth. *RMLE Online* 32, 5 (2009), 1–20.
[20] Liang Gou, Fang You, Jun Guo, Luqi Wu, and Xiaolong Luke Zhang. 2011. Sfviz: interest-based friends exploration and recommendation in social networks. In *Proc. VINCI '11*. ACM, 15.
[21] David M Greenberg, Daniel Müllensiefen, Michael E Lamb, and Peter J Rentfrow. 2015. Personality predicts musical sophistication. *Journal of Research in Personality* 58 (2015), 154–158.
[22] Brynjar Gretarsson, John O'Donovan, Svetlin Bostandjiev, Christopher Hall, and Tobias Höllerer. 2010. Smallworlds: visualizing social recommendations. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 833–842.
[23] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 241–250.
[24] Yosef Hochberg and Yoav Benjamini. 1990. More powerful procedures for multiple significance testing. *Statistics in medicine* 9, 7 (1990), 811–818.
[25] Yucheng Jin, Karsten Seipp, Erik Duval, and Katrien Verbert. 2016. Go with the flow: effects of transparency and user control on targeted advertising using flow charts. In *Proc. of AVI '16*. ACM, 68–75.
[26] Yucheng Jin, Nava Tintarev, and Katrien Verbert. 2018. Effects of individual traits on diversity-aware music recommender user interfaces. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. ACM, 291–299.
[27] Yucheng Jin, Nava Tintarev, and Katrien Verbert. 2018. Effects of personal characteristics on music recommender systems with different levels of controllability. In *Proceedings of the 12th ACM Conference on Recommender Systems*. ACM, 13–21.
[28] Roy PC Kessels, Martine JE Van Zandvoort, Albert Postma, L Jaap Kappelle, and Edward HF De Haan. 2000. The Corsi block-tapping task: standardization and normative data. *Applied neuropsychology* 7, 4 (2000), 252–258.
[29] Bart P Knijnenburg, Niels JM Reijmer, and Martijn C Willemsen. 2011. Each to his own: how different users call for different interaction methods in recommender systems. In *Proc. of RecSys'11*. ACM, 141–148.
[30] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 441–504.
[31] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82, 13 (2017), 1–26. https://doi.org/10.18637/jss.v082.i13
[32] Sébastien Lallé, Cristina Conati, and Giuseppe Carenini. 2017. Impact of Individual Differences on User Experience with a Visualization Interface for Public Engagement. In *Proc. of UMAP '17*. ACM, 247–252.
[33] Moira Maguire and Brid Delahunt. 2017. Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *AISHE-J: The All Ireland Journal of Teaching and Learning in Higher Education* 9, 3 (2017).
[34] Martijn Millecamp, Nyi Nyi Htun, Yucheng Jin, and Katrien Verbert. 2018. Controlling Spotify recommendations: effects of personal characteristics on music recommender user Interfaces. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. ACM, 101–109.
[35] Akira Miyake and Priti Shah. 1999. *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge University Press.
[36] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart. 2014. The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS one* 9, 2 (2014), e89642.
[37] John O'Donovan, Barry Smyth, Brynjar Gretarsson, Svetlin Bostandjiev, and Tobias Höllerer. 2008. PeerChooser: visual interactive recommendation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1085–1088.
[38] Thomas V Perneger. 1998. What's wrong with Bonferroni adjustments. *Bmj* 316, 7139 (1998), 1236–1238.
[39] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 157–164.
[40] R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
[41] Julian B Rotter. 1966. Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied* 80, 1 (1966), 1.
[42] Rashmi Sinha and Kirsten Swearingen. 2002. The role of transparency in recommender systems. In *CHI'02 extended abstracts on Human factors in computing systems*. ACM, 830–831.
[43] Nava Tintarev. 2017. Presenting Diversity Aware Recommendations: Making Challenging News Acceptable. In *Proc. of FATREC 17'*.
[44] Nava Tintarev and Judith Masthoff. 2007. A survey of explanations in recommender systems. In *Data Engineering Workshop, 2007 IEEE 23rd International Conference on*. IEEE, 801–810.
[45] Nava Tintarev and Judith Masthoff. 2016. Effects of Individual Differences in Working Memory on Plan Presentational Choices. *Frontiers in psychology* 7 (2016).
[46] Dereck Toker, Cristina Conati, Giuseppe Carenini, and Mona Haraty. 2012. Towards adaptive information visualization: on the influence of user characteristics. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 274–285.
[47] Stephanie Tom Tong, Elena F Corriero, Robert G Matheny, and Jeffrey T Hancock. 2018. Online Daters' Willingness to Use Recommender Technology for Mate Selection Decisions. In *Proceedings of the 5th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems co-located with ACM Conference on Recommender Systems (RecSys 2018)*. ACM, 45–52.
[48] Chun-Hua Tsai and Peter Brusilovsky. 2017. Enhancing Recommendation Diversity Through a Dual Recommendation Interface. In *Proc. of RecSys IntRS'17*. 10.
[49] Chun-Hua Tsai and Peter Brusilovsky. 2018. Beyond the Ranked List: User-Driven Exploration and Diversification of Social Recommendation. In *23rd International Conference on Intelligent User Interfaces*. ACM, 239–250.
[50] AJAM Van Deursen, Ellen J Helsper, and R Eynon. 2014. Measuring digital skills. *From digital skills to tangible outcomes. Project Report. Recuperado de: www. oii. ox. ac. uk/research/projects* (2014).
[51] Katrien Verbert, Denis Parra, Peter Brusilovsky, and Erik Duval. 2013. Visualizing recommendations to support exploration, transparency and controllability. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 351–362.
[52] Bodo Winter. 2013. A very basic tutorial for performing linear mixed effects analyses. *arXiv preprint arXiv:1308.5499* (2013).