

“Knowing me, knowing you”: personalized explanations for a music recommender system

Millecamp Martijn · Cristina Conati ·
Katrien Verbert

Received: date / Accepted: date

Abstract Due to the prominent role of recommender systems in our daily lives, it is increasingly important to inform users why certain items are recommended and personalize these explanations to the user. In this study, we explored how explanations in a music recommender system should be designed to fit the preference of different personal characteristics. More specifically, we investigated three personal characteristics that influence the perception of explanations in music recommender system interfaces: need for cognition, musical sophistication, and openness. For each of these personal characteristics, we designed explanations for users with lower and higher levels of the personal characteristic. Afterward, we conducted for each personal characteristic a within-subject user study in which we compared the two explanations. Based on the results of these user studies, we provide design suggestions to adapt explanations to different levels of these three personal characteristics. In general, we suggest providing explanations up-front for all recommendations at once. For users low in need for cognition, displaying these explanations must be optional. To support users with low musical sophistication, we suggest providing brief explanations that do not require domain knowledge. For users

M. Millecamp
Department of Computer Science, KU Leuven
Celestijnenlaan 200A bus 2402, Leuven, Belgium
E-mail: martijn.millecamp@kuleuven.be
Orcid: 0000-0002-5542-0067

C. Conati
Department of Computer Science, UBC
2366 Main Mall, Vancouver, BC, Canada
E-mail: conati@cs.ubc.ca
Orcid: 0000-0002-8434-9335

K. Verbert
Department of Computer Science, KU Leuven
Celestijnenlaan 200A bus 2402, Leuven, Belgium
E-mail: katrien.verbert@kuleuven.be
Orcid: 0000-0001-6699-7710

with low openness, we suggest providing explanations with a lower number of explanation elements.

This paper or a similar version is not currently under review by a journal or conference. This paper is void of plagiarism or self-plagiarism as defined by the Committee on Publication Ethics and Springer Guidelines.

1 Self-assessment

1.1 What is the main research question that your planned submission addresses?

In this study, we aim to answer the question of how explanations should be designed for three different personal characteristics: need for cognition, musical sophistication, and openness.

1.2 What makes your research results important and worth being reported in a top-ranked journal?

To our knowledge, this is the first study that designs and evaluates explanations for different personal characteristics and does not only explore the effect of personal characteristics on the perception of explanations. Additionally, it is important to get an overview of the three different studies which extends the scope of a conference paper.

1.3 Why does your planned submission fit into the scope of UMUAI?

This study fits into the scope of UMUAI because it investigates how **explanations** in a **music recommender system interface** can be **adapted to personal characteristics**. Each of these elements in bold is mentioned as a core aspect of the scope of UMUAI.

1.4 What are the main limitations of your approach?

One of the main limitations of this study is that the results are obtained in a very specific context (creating a playlist in a music recommender interface) and thus that it is not certain to which extent the results can be generalized to other situations and other domains. Additionally, this study only reports the results of short-term experiments which investigate the perception of explanations in a first visit to the system. A long-term study could shed light on the evolution of the need for (personalized) explanations in music recommender systems.

1.5 What is the relationship of your work to the closest 2-3 publications by others?

This work mainly builds onto the results of our previous studies that explored the effect of personal characteristics on explanations [54,57,53,55]. Additionally, Kouki et al. [44] also explored the effect of musical sophistication and openness on the perception of explanations, and Ribera et al. [77] proposed guidelines to tailor explanations to domain experience. We used the results of these papers to come up with our hypotheses.

Keywords music recommender system interface, explanations, personalization, openness, need for cognition, musical sophistication

2 Introduction

Due to the development of Internet applications, music streaming services are becoming the mainstream way of listening to music [?]. To help users find songs in the enormous database, these music streaming services often provide a music recommender system which suggests songs the user may like [?]. Moreover, recommender systems are not only present in music streaming services but are also available on several other platforms to suggest friends, jobs, movies, books, etc. It is clear that recommender systems gain importance and that users rely more and more on these systems to make decisions [76]. As a consequence, it is natural that there is also an increasing call to make these systems more transparent, so users do not have to follow recommendations in blind faith [81].

However, despite the increasing interest to make recommender systems more transparent through explanations, it remains unclear how to design such explanations in practice [81]. One of the reasons that this is still unclear, is the fact that users react differently to explanations and that only recently the effect of personal characteristics on the perception of explanations has become a pressing research topic [61,81]. Furthermore, in the field of explainable artificial intelligence, Gunning et al. [29] identified that tailoring explanations to the users is one of the open challenges in this field. This illustrates that there is a need to research how explanations can be adapted to the personal characteristics of the user.

In recent work that investigated the effect of personal characteristics on explanations in recommender systems, there are four main categories which have shown to impact perception: interest [13,15,48,52,59], cognitive style [54,57,53,55,63], domain experience [44,54] and personality [44,55]. Especially for the last three categories, most of the existing studies did not aim to design and evaluate explanations adapted to these characteristics but only explored which personal characteristics influenced the perception of explanations [44,54,55]. For example, in our previous work [54,55], we explored the effect of multiple personal characteristics on the perception of explanations in a music recommender system interface. In the first study [54], we found that users with

lower levels of need for cognition (NFC) had more confidence in their playlist when they created it with explanations than without. In a second study [55], we found that also musical sophistication (MS) and openness affected the perception of explanations.

As a follow-up, we evaluate explanations for a music recommender system that are designed for different levels of NFC, MS, and openness.

The final goal of this research is to devise system-driven support to customize explanations in a music recommender system based on the context and the personal characteristics of the users. As a first step, we want to investigate in this study how we can design explanations adapted to the needs of different levels of NFC, MS, and openness leading to the following research question:

Do users prefer explanations designed for their level of NFC, MS, or openness in a music recommender system interface, and why?

The main contribution of this paper is twofold. First, we report the results of three different user studies in which we investigate the preference of explanations designed for different levels of NFC, MS, and openness. The results show that low MS users and low openness users preferred explanations designed for their level of MS and openness over explanations designed for higher levels of MS and openness. For NFC, our results indicated that all users preferred the explanations designed for high levels of NFC over the ones designed for low NFC. For high MS, slightly more users preferred the explanations designed for low MS over the explanations designed for high MS, but this difference in preference was not significant. For high openness, the results showed that more users preferred the explanations designed for their level of openness, but this difference in preference was not significant.

Second, based on these results we provide design suggestions on how explanations can be adapted to NFC, MS, and openness in a music recommender system.

The rest of the paper is organized as follows: In Section 3, we provide an overview of related work in the field of adaptive systems, personalized recommender systems, and personalized explanations. In Section 4, we explain the study protocol, the characteristics of participants, the measurements, and the analysis that we have done for the three studies. Afterward, we explain the implementation of the recommender system interface and the design of the explanations in Section 5. The details, the results, and the discussion of the three experiments are explained in Section 6, Section 7, and Section 8. We conclude this paper with a summary of the results and a discussion of the implications.

3 Related work

In this section, we start by briefly introducing the different steps needed to implement an adaptive system in Subsection 3.1. Next, we provide an overview of explanations in recommender systems in Subsection 3.2. Afterward, we explain the different personal characteristics that have been shown to influence user perception of explanations and elaborate on previous studies that already investigated this influence in Subsection 3.3.

3.1 Adaptive systems

Despite the vast amount of research about explanations in recommender systems, Springer et al. [80], Naiseh et al. [62], and Jannach et al. [33] identified that there is a lack of research on guidelines to adapt explanations in recommender systems to the needs of users.

To address this research gap, a possible solution to incorporate these differences between users could be to implement a personalized recommender system interface that can adapt the explanations to the needs of the user. To implement such a personalized system, Paramythis et al. [74] proposed a blueprint with five different layers collecting input, interpreting input data, modeling the current state of the world, deciding which adaptation to apply, and applying that adaptation.

In this paper, we start to investigate the fourth layer of this model by evaluating explanations designed for different levels of personal characteristics.

In this fourth layer, given the state of the user model, the system needs to decide whether or not an adaptation is adequate as well as to decide the required type of adaptation [74]. The challenge of deciding upon adaptation is selecting the most useful adaptation out of the numerous options available.

In the field of recommender systems, the system can for example adapt the recommendation strategy itself [49], but it can also adapt different elements in the interface such as changing the number of recommendations [39,79], the diversity of the recommendations [83], the rating scale of the items [91], the level of presentation detail [24], the interaction method [42] or the presentation of the recommendations [34]. In this study, we are interested in providing explanations of a recommended item that are designed for a specific personal characteristic of the user. More details about explanations will be discussed in Subsection 3.2.

3.2 Explanations

Due to the increasing impact of artificial intelligence (AI) on our daily life and the recent requirements arising from the European Union's General Data Protection Regulation (GDPR), users are increasingly aware of the fact that the majority of AI applications still act as black boxes [10,81]. To make these

applications more intelligible to humans, there is an increasing interest in the field of explainable AI (XAI) to investigate methods to explain AI to end-users [29].

Similar to most AI applications, recommender systems often act as black boxes for the end-users [68]. Even as making recommender systems transparent is often not as crucial as transparency in AI applications in the field of medicine, defense, law, and finance, it has been shown that explanations can benefit the user experience in recommender systems [84].

In the next paragraphs, we will go into detail about the purpose, the delivery method, and the risks of explanations in recommender systems.

Purpose Since the study of Herlocker et al. [31] found that providing explanations could increase the acceptance of recommender systems, researchers realized that transparency was not the only advantage of explanations. Tintarev and Masthoff [84] even identified seven different purposes explanations could serve: effectiveness, efficiency, persuasiveness, satisfaction, scrutability, transparency, and trust. Next to these seven, Jannach et al. [33] identified two more purposes, namely debugging and allowing users to learn from the system. Additionally, since the new European Union’s GDPR, explanations could also serve the purpose of complying with legal regulations [62]. The definitions of each of these different purposes are listed in Table 1.

Purpose	Description
Compliance	Compliance with legal regulations
Debugging	Help to identify defects in the system
Education	Allow users to learn something from the system
Effectiveness	Help users make good decisions
Efficiency	Help to make decisions faster
Persuasiveness	Convince users to try, consume, or buy
Satisfaction	Increase the ease of use or enjoyment
Scrutability	Allow users to steer the system
Transparency	Explain how the system works
Trust	Increase the confidence in the system

Table 1: Purposes of explanation (based on [33,84])

Because several studies already argued that it is not enough to provide users with explanations, but that explanations should be accompanied by controls to enable users to correct and steer the recommendation process, our paper focuses on explanations serving the purposes of transparency and scrutability [31,84].

Delivery method Explanations could not only differ in the content they present but also in the way this content is delivered. In general, the delivery methods for explanations in recommender systems can be categorized into three: persistent, on-demand, and autonomous [61].

If explanations are delivered persistently, this means that they are delivered up-front with the recommendations and thus that the users do not need to take action to access them, but also cannot remove them [61]. In contrast, if explanations are delivered on-demand, users need to take action to explicitly indicate that they would like to access the explanations [61]. An example could be a system in which the user can ask a chatbot to explain why an item is recommended [65] or in which they need to push a button to show the explanation [54]. A third option is that the system itself decides autonomously when and for which items the user needs explanations [61].

Risks Despite all advantages of explanations, providing explanations also carries some risks. Naiseh et al. [61] identified six different types of risks linked with explanations: over-trust, under-trust, refusal, perceived loss of control, information overload, and suspicious motivation. As discussed previously, explanations can help to increase the trust of users in the system [84]. However, when users start to trust the system and accept the recommendations rashly, this could lead to undesirable outcomes especially in high-risk domains such as medicine, law, and loans [61]. In contrast, our previous study [54] found that showing the internal reasoning of users could also make clear to users that the system has limitations which lead to under-trust. Another risk of explaining the internal reasoning of the recommender system to users could be that they become aware of a mismatch between their own mental model and the actual system which could lead to refusal to use the system again [46]. It could also be that the explanations overwhelm the user which can lead to information overload and refusal. In case the recommender system is not controllable, only providing explanations could also lead to a perceived loss of control because users are not able to correct wrong assumptions or steering the recommendation process. The last risk of showing explanations could be that users perceive explanations as an attempt to manipulate the users and thus that the system is looking to maximize the profit of the company and not recommending the best items. In this paper, these risks were taken into account in the design process of the explanations as will be discussed in Section 5.3.

3.3 Personal characteristics

One of the requirements to create a successful and usable recommender system is to build a detailed user model that can be used by the system to recommend items or to adapt the interface [27, 66]. This user model can include a large variety of personal characteristics going from demographics to personality traits. To provide an overview of the most used characteristics in recommender systems, we divided them into four categories: personality traits, cognitive styles, cognitive abilities, and domain experience.

3.3.1 Personality traits

Already since antiquity, people have tried to measure and describe the way users behave, think, and feel [3,90]. Since then, there have been multiple attempts to describe the “*personality*” of a person, but the most used approach is the personality trait approach which focuses on different traits that can explain the ways users differ psychologically from one another and on how these traits can be measured [70,90]. Two assumptions on which this approach relies are that traits directly influence behavior and that traits are relatively stable over time [51].

The Big Five After several attempts to identify all relevant traits, it was proven that only 5 factors were replicable [12,25]. These factors led to the Big Five personality model which describes personality using five factors: openness, conscientiousness, extraversion, agreeableness, and neuroticism [25]. A definition of these traits can be found in Table 2.

To assess the Big Five model, a wide range of questionnaires have been proposed going from lengthy questionnaires using more than 130 questions [38,5,75] to a short version using only 10 questions [26]. However, to balance accuracy and time, one of the most used questionnaires is the 44-item Big Five Inventory [26,37].

Trait	Definition
Extraversion	summarizes traits related to activity and energy, dominance, sociability, expressiveness, and positive emotions.
Agreeableness	contrasts a pro-social orientation towards others with antagonism and includes traits such as altruism, tendermindedness, trust, and modesty.
Conscientiousness	describes socially prescribed impulse control that facilitates task-, and goal-directed behavior.
Neuroticism	contrasts emotional stability with a broad range of negative affects, including anxiety, sadness, irritability, and nervous tension.
Openness	describes the breadth, depth, and complexity of an individual’s mental and experiential life.

Table 2: Definitions of the Big Five personality traits according to [6]

Personality in recommender systems The first reason why personality is popular to take into account is its influence on behavior, preferences, decision-making processes, and interests [90,69,?]. As result, the recommender system

would be able to create more accurate recommendations and predict future actions of the user by taking the personality of the user into account [27].

A second reason to use personality is the fact that personality is a relatively stable construct over time and domain. This means that there is almost no risk that the model gets out of date and that the same model can be used to recommend items in different domains [70,90].

Several studies also showed the advantages of using personality in recommender systems. For example, the study of Hu and Pu [32] showed that using personality could help to overcome the cold start problem. Similarly, the study of Fernandez et al. [21] showed that taking personality into account could even alleviate the cold start problem cross-domain.

In the field of music, it has been shown that personality affects music preference [2], the way we browse for music [23], preference for digital program notes [86], and thus that personality-based recommender systems could improve the music recommendation [23].

Effect on explanations Personality has also been shown to influence the way users perceive explanations. In the studies of Kouki et al. [44] and our previous work [55], the effect of personality on the perception of explanations was investigated. Kouki et al. [44] investigated the effect of the Big Five personality traits on seven different explanations and found that users high in neuroticism preferred popularity-based explanations while users low in neuroticism preferred item-based explanations. They also found a positive correlation between openness and the number of explanation styles which they perceived as persuasive, and that they do not prefer any specific explanation style [44].

In our previous study [55], we also investigated the effect of the Big Five personality traits on the presence or absence of explanations. We found a significant interaction effect of openness on the intention to use again the system [55]. Low openness users reported that they prefer a system with explanations over a system without even as they perceive to find fewer novel songs. For high openness users, we did not see this difference in use intention, but they indicated that they find slightly more novel songs when explanations are available. Additionally, a follow-up investigation of the gaze pattern showed that high openness users focus more on recommendations when there are no explanations available. We argue that this could mean that their explanations did not support enough exploration of the recommendations.

3.3.2 Cognitive styles

The second category of personal characteristics that has been used to personalize systems is the cognitive style which indicates the preferred way to process information. It has been shown that cognitive style influences learning performance, learning preferences, subject preferences, and social behavior [78].

In this paper, we choose to use need for cognition (NFC) which was defined by Cacioppo et al. [11] as “*the tendency to engage in and enjoy effortful cog-*

nitive activities". The motivation behind this choice is that NFC is positively correlated to strategic adaptive decision-making behavior [87]. This was also illustrated by Coutinho et al. [19] who showed that users high in NFC asked more for an explanation of the solution in a learning environment than users low in NFC.

Cognitive style in recommender systems Even as the theory shows that cognitive style is a promising characteristic to take into account, the research about the influence of cognitive style in recommender systems is scarce [27]. Notable exceptions are the studies of Tong et al. [87] which showed that NFC influences the willingness of users to rely on a recommender system and the study of and Tam et al. [82] which showed an interaction effect of NFC and the level of preference matching [82].

Effect on explanations in recommender systems There is also a limited number of studies that found that cognitive style influences the user experience of explanations [54, 57, 63].

One of the first studies that investigated the effect of NFC on the perception of explanations was the study of Naveed et al. [63]. They investigated the moderating effect of two thinking styles on explanations in the camera domain: rational ability which is the same as NFC and experiential ability [73]. They found a moderating effect of experiential ability, but no moderating effect of rational ability [63]. However, the results of a follow-up study showed a moderating effect of NFC in the music domain, but not in the camera domain [57].

This effect of NFC in the music domain confirmed the findings of our previous study [54] in which we also found a moderating effect of NFC on the confidence users had in a playlist they created in the presence or absence of explanations. The results of this study showed that users with a low NFC reported higher confidence in having a good playlist when they created the playlist in the presence of explanations. Participants reported that the reason for this was that explanations helped them to gain confidence without the need to put a lot of effort into it. For high NFC, the results were the opposite as these users had more confidence when the explanations were not available. Qualitative analysis suggested that the reason for this was twofold: users who distrust the system because explanations reveal when the recommendations are not a good fit and users who think explanations are redundant because they already know what they want.

In another study, Millecamp et al. [53] analyzed why and when users wanted explanations and they found that the biggest differences between high and low NFC users are: (i) low NFC users reported that they want explanations when they are looking for a specific kind of music, (ii) high NFC users reported they want explanations because it gives them the possibility to steer the recommendation process and (iii) high NFC users report that on a desktop, they would prefer to see explanations upfront for all recommendations and not behind a button. In contrast, a follow-up study [55] with a similar interface did not find

a moderating effect of NFC. Our explanation for this absence of moderating effect is that the delivery method of the explanations changed: in the first study [54], the explanations were visible on-demand while in the follow-up study [55] the explanations were persistent [61]. In the latter, we argue that persistent explanations might reduce the difference between low and high NFC users similar to the effect already described by Jugovac et al. [40]. They described that the presence of a recommender system with persistent explanations diminishes the difference between maximizers and satisficers and it has been shown that maximizers tend to have a higher NFC than satisficers [64].

3.3.3 Cognitive abilities

The third category of personal characteristics is cognitive abilities. Cognitive ability is a construct to describe the differences among individuals in terms of their mental capabilities [71]. The main difference between cognitive style and cognitive ability is that cognitive style influences the user on a general level e.g. positive or negative effect on performance, while there is a more concrete relation between cognitive ability and performance. For example, the higher the ability the better the performance [78].

In the context of processing visual information, different measures such as visual working memory [89], visual literacy [8], spatial memory [20], and perceptual speed [20] have been proposed.

Cognitive abilities in recommender systems Due to this vast amount of research, it has been shown that cognitive abilities affect almost all domains in life, among them the way users interact with visualizations [71]. Jin et al. [35], for instance, found that users with higher cognitive abilities preferred to receive recommendations in a more complex scatter plot than in a more simple bubble chart. Previous work has also found that cognitive abilities can influence cognitive load when interacting with visual information systems [47, 18, 85].

In our previous study [54], we did not find an influence of visual working memory or visualization literacy on the perception of explanations which is the reason we did not include cognitive abilities in this study.

3.3.4 Domain experience

The last category of personal characteristics that is often used to personalize recommender systems is the level of experience of the user. Due to the large variety of different domain aspects, there is not yet a standardized way to measure domain experience, but mostly a continuous scale going from novice to expert is used [4]. To measure domain experience in the music domain, the most used scale is the Goldsmiths Musical Sophistication Index ¹. It is an effective way to measure the music expertise of users, and it has shown a strong correlation with individuals' music preference [60] and listening behavior [22].

¹ <https://www.gold.ac.uk/music-mind-brain/gold-msi/> February 2021

Domain experience in recommender systems Research generally confirms that domain experience has an important influence on the way users interact with and perceive a recommender system [4, 7, 41, 45, 42].

For example, novice users generally lack attribute knowledge and thus prefer to express their preferences in a conversational recommender system by interacting with a natural language avatar while expert users might prefer to use a set of detailed, domain-specific forms [33, 42]. Another example is the study of Knijnenburg et al. [42] who found that users with a higher domain knowledge prefer to get recommendations from a hybrid approach while novice users prefer non-personalized, popular recommendations.

Effect on explanations in recommender systems Both Kouki et al. [44] and our previous work [55] investigated the effect of musical expertise on the perception of explanations, but they found opposite results. Kouki et al. [44] did not find a significant difference between music experts or novice users.

In contrast, our previous work [55] showed that expert users felt more supported with explanations than without to take a decision and that they feel more supported than novice users. Additionally, we investigated the gaze pattern of users and found that low MS users have lower transitional entropy than high MS users. As discussed in Naiseh et al. [61] one of the risks of explanations is that they could cause information overload. The reported difference in gaze and feeling of support between low and high MS users is probably caused by information overload which was more present by lay users than by expert users.

In the broader field of XAI, it is worth mentioning that Ribera and Lapedriza [77] provided guidelines to design explanations for novice and expert users. For domain experts, they argue that the explanations should be provided through interactive visualizations which allow the experts to lead the discovery by themselves [77]. For lay users, the explanations can be briefer and should allow users to select the one argument that is most interesting to their case [77].

4 User studies

As mentioned in Section 2, the goal of the research in this paper is to decide which type of explanation should be provided to different users. As previous research showed that NFC, MS, and openness influence the perception of explanations [44, 54, 55], we investigate how explanations could be designed for the preferences of these personal characteristics. Investigating all three personal characteristics at once would require a large sample size to ensure all combinations of personal characteristics are present [90]. For this reason, we conducted three separate studies in which we investigated whether we could design the best explanation for each of these three different personal characteristics. In each experiment, we designed two interfaces: one personalized for users with lower levels on the personal characteristic and one for users with

higher levels. In this section, we will first describe the experimental design which is the same for the three studies. Afterward, we explain how we recruited the participants, which measurements we used in the studies, and the statistical method we applied.

4.1 Study procedure

After participants filled in an informed consent, they were asked to fill in a questionnaire to measure their NFC, MS, or openness for study 1, 2, and 3 respectively. After this questionnaire participants were given information about the different parts of the interface to make them familiar with the different features of the system (the details of these features are discussed in Section 5). For each experiment, we designed two interfaces: one which we hypothesized would be preferred by the low group and one by the high group. We also identified two different tasks, both involving creating a playlist of eight songs. In one task they were requested to create a playlist of eight songs for a relaxing activity and in the other task for a sports activity². These situations were chosen because previous research has found that users listen the most to music during sports or relaxing [28]. We chose a short playlist of eight songs to prevent fatigue and to stay within the time limit of 20 minutes for the whole experiment. To evaluate the two interfaces, we followed a within-subject, counterbalanced design: when participants read the information about the features, they were randomly assigned to one of the two interfaces and one of the two tasks. After creating a playlist with the first interface-task, users would then repeat these steps with the second interface-task pair. At the end of the study, we asked users which of the two interfaces, and hence associated explanations, they preferred and also to motivate this choice.

4.2 Participants

For all three experiments, we conducted a crowd-sourced study on Amazon Mechanical Turk³. We recruited 90 participants for each experiment because this is considered an appropriate sample size to detect medium to small effects in a within-subject study [43].

To ensure ethical research on Amazon Mechanical Turk, we followed the guidelines of Moss et al. [58] by opening multiple tasks to people with different experience on the platform and by testing whether the experiment was user-friendly. We tested the user-friendliness of the experiment by conducting several pilot studies. These pilot studies were also used to estimate the time needed to perform the tasks which was communicated to the participants before accepting the task. For all three studies, we estimated that it would take

² The exact phrasing for the second situation: *Please create a playlist of 8 songs to which you would listen during a relaxing activity*

³ <https://www.mturk.com/>

20 minutes on average to complete the experiment and for this reason we offered \$4 for completing the HIT as \$12 per hour is the average wage offered on Amazon Mechanical Turk [30]. To safeguard high-quality answers, we added three attention checks in different questionnaires and removed users (N=47) who missed one or more of these checks.

To divide the participants into balanced groups with respect to each of the tested user characteristics, we used a median split resulting in a group with low and a group with high scores for the personal characteristic. The demographics of the participants as well as the personal characteristics of the low and high groups can be found in Table 3.

Table 3: Demographics for participants in the three studies

PC	N (F)	Age	Time	Range	Mean	Med	Low		High	
							N	Mean	N	Mean
NFC	91 (34)	33.47	17m55s	[0-72]	42.36	40	52	33.5	39	55.1
MS	90 (26)	32.5	18m08s	[18-126]	72.66	78.5	45	59.5	45	85.7
Openness	90 (33)	29.3	17m58s	[0-42]	27.12	27	48	24.0	42	30.7

4.3 Measurements

4.3.1 Personal characteristics

In each study, the relevant personal characteristic was measured at the beginning of the study using a standard psychology test.

To measure NFC, we used the well-established, 18 items questionnaire developed by Cacioppo et al. [11] which measures NFC on a scale from 0 to 72. A sample question is *"I prefer complex to simple problems."* which is rated on a 5-point Likert scale going from *Extremely uncharacteristic for me* to *Extremely characteristic for me*.

For measuring general MS, we used the Goldsmiths Musical Sophistication Index which measures MS on a scale between 18 and 126. This questionnaire also has 18 items, and each item is rated on a 7-point Likert scale. A sample question is *"I am able to judge whether someone is a good singer or not"*.

To measure openness which is one of the Big Five personality traits, we used the subset of 10 items which relate to openness from the 44-item Big Five Inventory. For each item, participants needed to indicate on a 5-point Likert scale to which extent they think the item applied to them. An example of such an item is *"I am original, I come up with new ideas"*.

4.4 Analysis

To analyze whether users with a certain level of personal characteristic had a significant difference in preference for a type of explanation, we used the

chi-square goodness-of-fit test. If users would not have such a difference in preference, we expect an equal distribution of preference between the two interfaces which motivates the hypothesized distribution of 50% for the chi-square test.

To analyze the answers to the question of why users preferred a type of explanation, we first divided the answers into two groups: users who preferred the first and users who preferred the second type of explanation. Afterward, we performed a thematic analysis on both groups using an inductive approach to identify patterns in this qualitative data following the six steps of Braun and Clarke [9]: familiarization with the data, coding, searching for themes, reviewing themes, naming and defining themes, and writing it all up. This thematic analysis was performed by two researchers separately and the results were synthesized during a discussion to enhance credibility [67, 72].

5 Interface and explanations

To investigate the personalization of explanations, we designed a modular music recommender system interface in which we could plug in different explanations. In the following paragraphs, we will explain in detail the interface and the different explanations that can be plugged in.

5.1 Spotify API

To generate recommendations, we used the Spotify API which takes two arguments as input: a source song and different target values for audio features⁴. As output, this API generates a list of songs that are similar to the provided source song and which are sorted based on the distance to the provided target values for audio features.

Through this API, we are also able to search for a song, to get a 30s preview of a song, and to get the different audio features of a song. There are fifteen different audio features of which we choose four: danceability, energy, happiness, and popularity. These audio features were selected based on their popularity and their uniform distribution [56].

5.2 Interface

As shown in part A of Figure 1, the left side of the interface is dedicated to the playlist of the user. Each song of the user’s playlist is represented by the title, the artist, and a circular picture of the album cover. Additionally, users were able to play a 30 seconds preview of the songs (play button on the cover), to delete the song from the playlist (red cross), and to select whether

⁴ <https://developer.spotify.com/documentation/web-api/reference/reference-index>

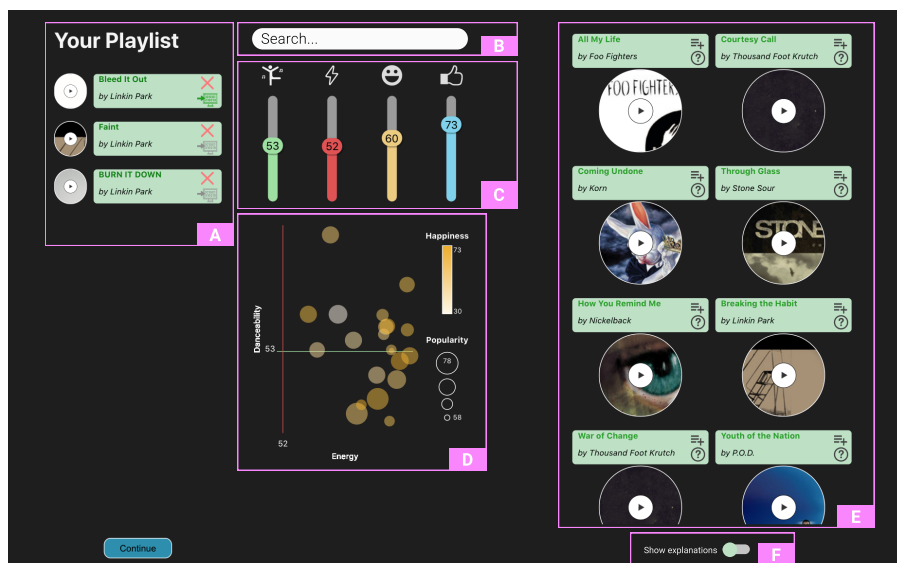


Fig. 1: Interface with the different parts highlighted in purple. A: Playlist, B: Search box, C: Preference of the user, D: Scatter plot, E: Recommendations, F: Switch for explanations

or not they want to get recommendations based on this song (icon in the lower right corner). For each song that was selected to serve as input, we generated recommendations using the Spotify API. In total, we displayed a list of 40 recommended songs.

Central in the interface, users could use the search box to look up a song that they want in their playlist as shown in Part B of Figure 1. As shown in Part C of this figure, four different sliders allowed users to indicate their preferred value for danceability, energy, happiness, and popularity. In part D of Figure 1, users could then explore these recommendations in a scatter plot which will be discussed in more detail in Section 5.3.

As soon as the user added one song they like, a list of recommendations appeared on the right side of the interface, as shown in part E of this figure. Similar to the songs in the playlist, the recommendations were presented by a circular picture of the album cover, the title, and the artist. Users were also able to play a 30s preview of the song, to add the song to their playlist, and to demand an explanation of why this song was recommended by clicking on the . Under the list of recommendations, users were able to switch on or off the explanations for all songs at once (part F). The recommendations were sorted based on how well they fitted the preferred audio features of the user.

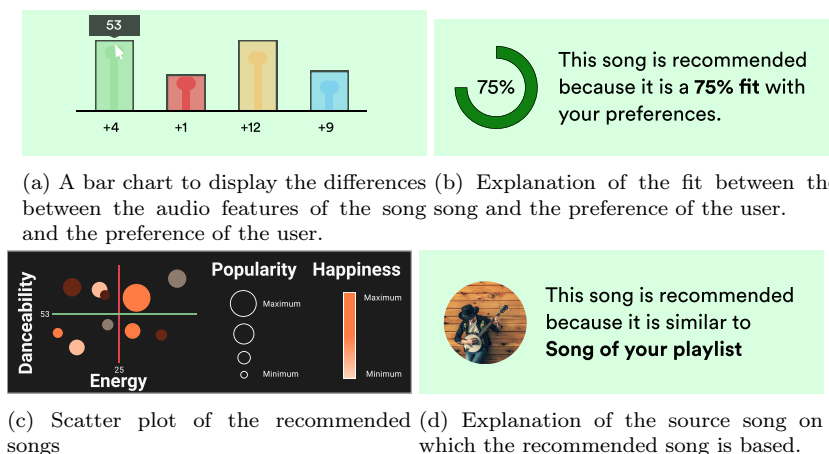


Fig. 2: Different possible explanations

5.3 Explanations

As mentioned in Section 3, we aim to design explanations that support transparency and scrutability. A possible way to design these scrutable explanations is by providing users with information that explains the relation between the provided inputs and the recommendations [68]. As we used the opaque Spotify API to generate recommendations, we were limited to three different information sources to explain the recommendations: the source song on which the recommendation is based, the audio features the user requested, and the audio features of the recommended song. Based on this information and explanations used in our earlier work [54], we produced four different kinds of explanations: the fit between the song and the preference of the user, a bar chart to display the different audio features of the song and the difference with the preference of the user, the source song on which the recommendation is based and a scatter plot to explore the recommendations in four dimensions. Due to the opaque character of the Spotify API, we cannot extract the influence of the provided inputs on the generated recommendations. These different explanations are discussed in detail in the following paragraphs.

5.3.1 Bar charts

As shown in part B of Figure 1, users were able to steer the recommendation process by changing the colored sliders which represent different audio features. To explain to the user that a song is recommended because the audio features of the recommended song are similar to their preferences, we implemented a bar chart as shown in Figure 2a. This bar chart shows the different audio features of the recommended song as colored bars of which the color is consistent with the colors of the sliders in Part B. In the background of each bar, there is a

silhouette of the slider of the preference of the users. By comparing the height of the slider and the height of the bar, users could see how well the song fitted each audio feature and which feature they need to change if they want other recommendations. To help users compare these two values, the numeric value of this difference was also shown under the bar. As also shown in Figure 2a, when the users hover over a bar, the exact value of the bar is shown in a tooltip.

5.3.2 The fit between the song and the preference of the user

To aggregate this information shown in the bar chart about the differences in audio features, we calculated the total fit between the audio features of the song and the audio features of the user. To do so, we used the function $\sum_{i=1}^4 \|F_{s,i} - F_{u,i}\|/4$, where $F_{s,i}$ and $F_{u,i}$ are the normalized values of the i^{th} audio feature of the song and the user respectively. As discussed, we used this fit to sort the generated recommendations from best fit to worst fit, so showing this brief score to the user externalizes this core criterion underlying the recommendations. Showing this fit to the users can be done visually as a radial progress bar but also textually as shown in Figure 2b.

5.3.3 Scatter plot

Another way to show the features of the recommended songs and to support the exploration of these songs is through a scatter plot [88,36]. As shown in part D of Figure 1 and in Figure 2c, the scatter plot showed the four audio features of recommendations in four different dimensions. The danceability and the energy value of a recommended song were presented by the position on the y- and x-axis of the circle respectively. To represent the popularity of a song, we used the size of the radius with a higher radius representing a more popular song. To represent the happiness of a song, the color of the circle was used with a color scale between yellow and orange, with orange representing a higher happiness value. When hovering over a circle, the corresponding recommendation in part E of Figure 1 was highlighted.

5.3.4 Source song

Next to the audio features, the recommendations are also generated based on a source song. As such, another kind of explanation we could provide to the user is the song in their playlist on which the recommended song is based and thus is similar to. As shown in Figure 2d, this is done by showing the cover of the song in the playlist and by a textual description that this song was based on a source song in their playlist. Due to the opaque nature of the Spotify API we can only show that the song is similar to the source song, but not the exact rationale for this similarity.

5.4 Overcoming risks of explanations

As discussed in Section 3, Naiseh et al. [61] identified several risks and side-effects that could arise while users are receiving explanations. The first risk that they identified was under-trust. In this study, we avoid under-trust as much as possible by limiting the system to recommend only items with a good enough fit (>80%). A limitation of this approach might be that the list of recommendations was sometimes shorter than 40 songs. To address the risk of refusal and perceived loss of control, all explanations show the relation between the provided input features of the user and the recommended song. As such, users can infer from this information how they should change their input features to get different recommendations (e.g. remove a song in their playlist as a source, change the slider to get more energy, etc.). To avoid information overload, it was possible to hide all the explanations by clicking on the toggle button as shown in part F of Figure 1. As creating a playlist is not a high-risk task and as we do not ask the users to buy specific songs we do not address the risk of over-trust or suspicious motivation in this study.

6 Experiment 1: need for cognition

6.1 Hypotheses

Based on the related work discussed in Subsection 3.3, we came up with two different designs which we hypothesized would fit best the needs of users with a low NFC and users with a high NFC.

We hypothesize that:

- *H1: users with lower levels of NFC prefer explanations accessible on-demand per recommendation over explanations provided all at once upfront.*
- *H2: users with higher levels of NFC prefer explanations provided all at once upfront or not at all over explanations accessible on-demand per recommendation.*

The reasoning behind these hypotheses is twofold:

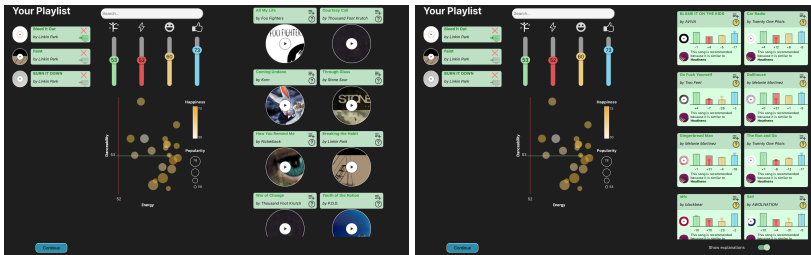
First, previous studies suggest that whether or not there is an effect of NFC on how users perceive explanations depends on the delivery method [55]. Specifically, in previous studies where the explanation condition were delivered on-demand, there was evidence of an interaction effect between NFC and having access to explanations on confidence in the playlist [54], perceived decision support, choice satisfaction [57], and explanation attention [17]. However, in a user study in which the explanation interface had all explanations available up-front, no such interaction effect was detected [55]. Additionally, in a follow-up study [53] almost half of the users with higher levels of NFC reported that they would prefer to see explanations all at once and up-front instead of on-demand. In contrast, for low NFC users, this was only reported by one user.

Second, Martin et al. [50] found that users with low NFC are less motivated to study information in depth and another study found that low NFC users put less effort into finding all information before deciding [53]. Additionally, in a previous study [54], a minority of users with higher levels of NFC reported that they did not need explanations because they can decide what they want without the information in the explanations.

We hypothesized that low NFC users prefer explanations accessible on-demand per recommendation based on (i) the results of our previous work that suggest that the presence or absence of interaction effect between NFC and perception depends on the delivery method, (ii) the result that in our previous work only one low NFC user reported that they would prefer to see explanations all at once and upfront, and (iii) that result of our previous work indicated that low NFC users are less motivated to study information in depth and to find all information before deciding.

We hypothesized that high NFC users prefer explanations all at once and upfront based on (i) the results of our previous work that suggests that the presence or absence of interaction effect between NFC and perception depends on the delivery method, (ii) the results in our previous work that show that almost half of high NFC users reported that they would prefer to see explanations all at once and up-front, and (iii) the result of our previous work in which a minority of high NFC users reported that they do not need explanations.


6.2 Design



(a) The interface designed for low NFC (b) The interface designed for high NFC.

Fig. 3: Interfaces designed for low and high NFC

As shown in Figure 3, we designed two interfaces, one designed for low NFC and one designed for high NFC.

To design the interface for low NFC, we disabled the option to see all explanations at once by hiding the toggle switch (Part F of Figure 1). As a consequence, explanations were only accessible by clicking on the . We will further refer to this interface as *On-demand*.

To design the interface for users with higher levels of NFC, explanations were shown upfront for all songs as shown in Figure 3b. To design the explanations for the needs of the minority of high NFC who reported that they do not need explanations, this overview of explanations could be turned off and back on by clicking on the toggle switch below the recommendations. We will refer to this interface as *Up-front*.

As shown in Table 4, only the delivery method of explanations was different between the two interfaces and not the explanation elements. Similar to our previous work in which there was an effect of NFC on the perception of explanations [54], we provided a bar chart, a scatter plot, and the source song as explanation elements. The explanation for each song is shown in Figure 4.

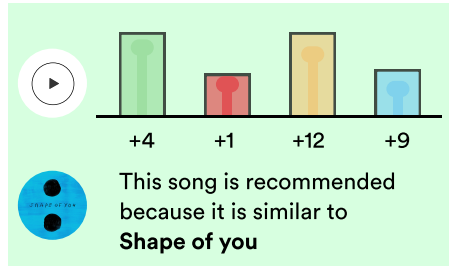


Fig. 4: Explanation for a recommended song

Table 4: Overview of the explanation elements and the delivery methods for the first user study

		On-demand	Up-front
Explanation elements	Bar chart	•	•
	Fit		
	Scatter plot	•	•
	Source song	•	•
Delivery method	One at a time	•	
	All at once		•

6.3 Participants

After filtering out the users that did not answer correctly all attention checks, 91 (34F) valid users remained. The average age of these participants was 33.47 and on average it took 17min 55s to finish the experiment. To divide the users into two groups, we performed a median split resulting in two groups with 52 and 39 users for low and high NFC respectively. The median value of NFC

was 40 which is slightly lower than the reported means in previous studies (44 [54] and 47 [55]). A histogram of the NFC scores of participants is shown in Figure 5.

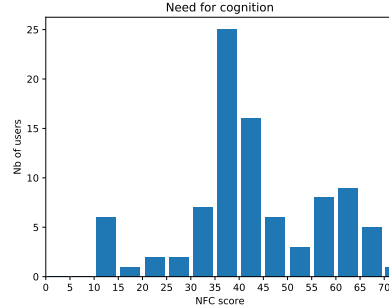


Fig. 5: Histogram of NFC score of participants

6.4 Results

As shown in Figure 6, we cannot confirm H1 because low NFC users prefer to see all explanations upfront. To test H2, we conducted a chi-square goodness of fit test which revealed that we could confirm H2 as the preference for Overview is significant ($\chi(1) = 24.03, p < .001$). As a follow-up, we verified whether the preference for On-demand was stronger by low NFC users than by high NFC users and a one-sample proportion test revealed that this was the case. Significantly more low NFC users preferred On-demand than high NFC users ($p = .029$).

As shown in Table 5, we also logged the number of users who opened the explanations in On-demand. For both the low and high NFC group, 15.4 % of the users opened explanations of at least one recommendation. For the low NFC group, these users opened on average the explanations for 5.38 songs. For the high NFC group, this was only for 1.67 recommendations. Thus, there seems to be a minor trend that low NFC make more use of the functionality to access explanations on-demand than high NFC users.

For Up-front, we logged the interaction with the toggle button to see who opened and closed the explanations all at once. For low NFC 17.3% used this functionality, for high NFC, this was 15.4% as shown in Table 5. If we look at the number of times participants used the functionality to close or open all explanations at once, there does only seem to be a small difference between the low and high NFC group.

Additionally, we also asked participants to report why they preferred one of the two explanations. After dividing the answers into answers of participants

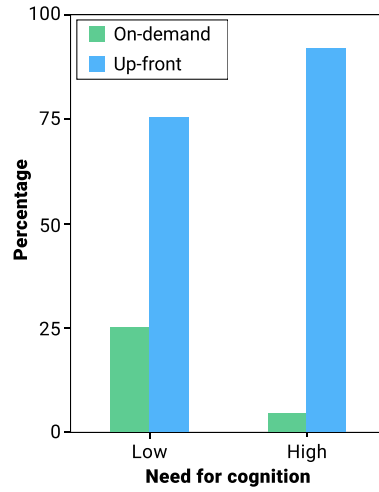


Fig. 6: Preference for explanations for users with low and high NFC

Table 5: Logging information for low and high NFC users in both interfaces with N the percentage of users who interacted with the explanations

	On-demand		Up-front	
	N	Nb of opened explanations	N	Nb of closing opening all explanations at once
Low NFC	15.4%	5.38 ± 7.44	17.3%	2.88 ± 3.1
High NFC	15.4%	1.67 ± 0.81	15.4%	1.83 ± 0.8

who preferred On-demand and Up-front, we categorized the answers for each group into themes. In Table 6, we show the identified themes for On-demand and Up-front together with the participants in the low of high NFC group who reported the theme.

On-demand We identified two themes in the answers about why participants preferred On-demand. A first theme is that they do not want to see information in which they are not interested: “If it’s not a song I’m interested in, I don’t care to see the information.” (P24). A second reason could be that they feel overwhelmed by the overview of explanations: “I want minimal information shown to me - I don’t want a wealth of choice overwhelming me.” (P34).

Up-front We identified five different themes in the answers why users preferred Up-front: Information, Decision support, Comparison, Conditionally, and Transparency.

The first reason users preferred the Up-front was because they just want to see all the **information** at once. For example, P31 reported *“It’s fairly useful information, so I’d be interested in it for any and all songs.”* A second theme we identified is related to the information, but in these answers, participants reported specifically that the overview **supported** them to make a **decision**: *“Because it helps me decide whether or not I’ll like a song.”* (p51) Some participants also mentioned why Up-front helped to make a decision, namely because it helped them to **compare** the different recommendations. An example is the answer of P58: *“To quickly compare songs without needing to specifically open each song.”*

The fourth theme was that even as users might like Up-front, they specifically mention that they only want to see the overview **conditionally** and thus need the option to switch off the overview of explanations. For example, P20 reported *“I think it’s cool info, but sometimes I may not care so I wanna turn it off or on”*.

The last theme we identified was that some participants preferred Up-front because it increased the **transparency** as it helps to understand the reasoning of the recommender system: *“I think it’s very neat to see why they recommended one song over another.”* (P50).

Table 6: Themes identified in the question why users preferred On-demand or Up-front

Preference	Theme	Low NFC	High NFC
On-demand	Not interested	P24,P60,P82	P40
	Overwhelming	P34	P17
Up-front	Information	P7,P61,P79	P31,P52,P57,P68,P70,P72,P77
	Decision support	P11,P28,P51,P54,P73,P83	P23,P26,P85
	Comparison		P18,P32,P47,P58
	Conditionally	P61,P67,P90	P21,P22,P63
	Transparency	P14	P18,P49,P50,P53,P63,P70

6.5 Discussion

Our results show that we needed to reject our hypothesis that low NFC would prefer explanations on-demand per recommendation over explanations delivered up-front for all recommendations at once (H1). As our hypothesis was based on two arguments, there might also be two reasons why we needed to reject H1. A first reason could be that previous studies argued that the difference between low and high NFC was due to the persistent or on-demand way of delivering explanations and that low NFC users would prefer explanations on-demand. In this study, even as Up-front does provide all explanations at once, explanations were not delivered in a purely persistent way, because users could turn the explanations off when they wanted. As shown in Table 6, low

NFC users also reported this functionality as motivation for the reason why they preferred Up-front.

A second reason could be that even as previous studies found that low NFC users are less motivated to seek and study information, they do not dislike the presence of the information. In Table 6, we can see that only one person with low NFC reported that he/she felt overwhelmed by the information, and there are even a few low NFC users who liked that the information was available in Overview.

For users with high NFC, there is a strong preference for Up-front and only a small minority choose On-demand over Up-front. This confirms the qualitative results of our previous study [53] that high NFC users prefer to see all the information at once without the need to seek explanations individually. However, as low NFC users also preferred these explanations, this preference seems not to be linked to NFC. Still, our results suggest that the preference for Up-front of high NFC users is stronger than the preference of low NFC users.

6.5.1 Design suggestions

As our results show, both low and high NFC users preferred to see explanations in an overview. As a consequence, we recommend providing an overview of explanations for all recommendations. However, it is important that this overview of explanations should also be delivered in an on-demand way to allow especially low NFC users to turn off the overview. Additionally, we would also recommend providing the option to seek explanations individually on-demand to meet the needs of a minority of users who feel overwhelmed by the overview.

7 Experiment 2: musical sophistication

7.1 Hypotheses

To design explanations for the needs of users with low and high MS, we put forward two different hypotheses:

- *H3: users with lower levels of MS prefer brief explanations which do not require domain knowledge over explanations with an interactive visualization, and which do require domain knowledge.*
- *H4: users with higher levels of MS prefer explanations with interactive visualizations and which require domain knowledge over brief explanations which do not require domain knowledge.*

The reasoning behind H3 is twofold:

First, previous studies showed that users with lower domain experience typically lack attribute knowledge [1, 16] which might prohibit these users to steer the recommendation process effectively when the controls rely on this attribute

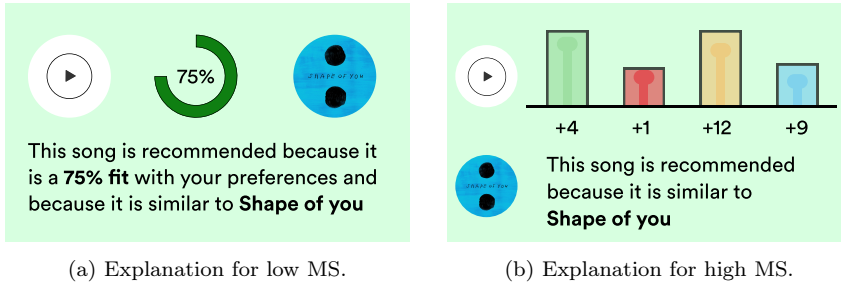


Fig. 7: Explanations for low and high MS

knowledge [42]. Second, Ribera et al. [77] provide the guideline to provide brief explanations to novice users.

The reasoning behind H4 is twofold:

First, Ribera et al. [77] suggest tailoring explanations to domain experts by providing interactive visualizations, allowing experts to explore and letting them decide when and how to question the explanations. Second, there is also evidence that high MS users feel well supported with domain-specific explanations [55].

7.2 Design

As shown in Table 7, there are three differences in the explanations designed for low and high MS users.

As we hypothesize that low MS users prefer brief explanations which do not require domain knowledge, we choose to hide the scatter plot and the bar chart as these require knowledge of the attributes. To still explain why songs were recommended, we show the goodness of fit and the source song as shown in Figure 7a. We will refer to this explanation as *Brief*.

Table 7: Overview of the explanation elements and the delivery methods for the second user study

		Low MS	High MS
Explanation elements	Bar chart		•
	Fit	•	
	Scatter plot		•
	Source song	•	•
Delivery method	One at a time	•	•
	All at once	•	•

Ribera et al. [77] recommend providing an interactive visualization as an explanation for expert users, so we choose to show the scatter plot for high MS

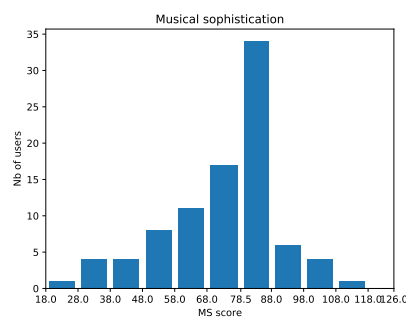


Fig. 8: Histogram of musical sophistication score of participants

users. As discussed before, users could hover over this scatter plot to highlight the corresponding recommended song. Additionally, we choose to show the bar chart and not the goodness of fit to allow domain experts to control the recommendation process based on the audio features which require domain knowledge. We also explained that a song was recommended because it was similar to a source song as shown in Figure 7b. This explanation is the same as the explanation used in the previous study. We will call this explanation *Combination*.

7.3 Participants

In total, 90 valid participants completed this study of which 26 were female. On average, users were 32.5 years old and completed the experiment in 18min08s. The distribution of MS scores is shown in Figure 8 and ranged from 21 to 112 with an average of 72.66 and a median of 78.5. This median is slightly lower than the 82 reported by Mullensiefen et al. [60] but higher than the MS reported in previous studies on Amazon Mechanical Turk [54,55]. Based on this median, we divided the participants into two different groups resulting in a group of low MS users (45) and a group of high MS users (45).

7.4 Results

As shown in Figure 9, low MS users prefer Brief over Combination. A chi-square goodness of fit test revealed that this preference for Brief is significant and thus that H3 could be confirmed $\chi(1) = 18.69, p < .001$.

Our fourth hypothesis was that users with a high MS would prefer Combination over Brief, but that hypothesis could not be confirmed $\chi(1) = 1.09, p = .297$. As a follow-up, we investigated whether there is a difference in interactions with the scatter plot in Combinations. As shown in Table 8, 53% of high MS users interacted with the scatter plot which is the same for low MS users.

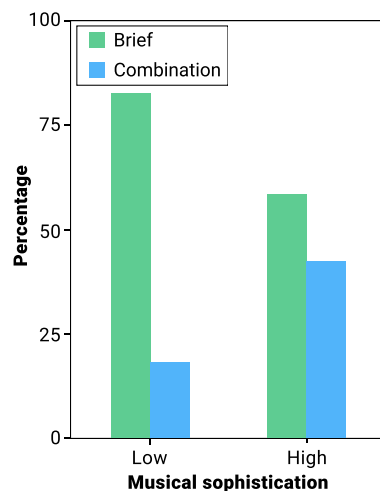


Fig. 9: Preference for simple and complex explanations of users with low and high MS

But as shown in this table, high MS users seem to use the hover function in the scatter plot more often than low MS users.

Table 8: Number (nb) of interactions with the scatter plot for low and high MS users in the Combination interface

	Nb of users that interacted with scatter plot	Avg. nb of interactions with scatter plot
Low MS	53%	4.83 ± 3.0
High MS	53%	8.46 ± 16.8

Additionally, we also analyzed why users preferred one of the two interfaces. The identified themes in these answers are listed in Table 9.

Brief The most emerging theme in the answer why users preferred Brief over Combination was that the explanation was **easier** and simpler to use and to understand. This is illustrated by the answer of P74: “[Brief] gives me the most information that I can understand easily.” We identified this theme both by users with low MS (N=19) as users with high MS (N=10). Another reason users preferred Brief, was because they **did not like the scatter plot** in Combination. This theme was identified by five low MS and one high MS user. P3 has formulated it as “The scatter plot is a little overwhelming to look at while the total fit is a one-stop graphic that is easy to understand.”

Combination As shown in Table 9, we identified three different themes in the answers why users reported to prefer Combination. A first theme that was identified is that users like the additional **transparency** provided by Combination: *“It is easier for me to understand why a song is being recommended” (P66)*. Another reason could be that they just more liked to see the **mix of information**. For example, P83 reported *“because I used a combination of scatter plot and bars”*. A last theme that was identified was that Combination is more visual. P69 reported *“I like this visual display of information as it gives me the most information.”*

Table 9: Themes identified in the question why users preferred Simple or Complex

Interface	Theme	Low MS	High MS
Brief	Easiest	19 participants	10 participants
	Dislike scatter plot	P3,P26,P38,P44,P76	P63
Combination	Transparent	P81	P29,P66
	Mix of information	P5,P11,P33,P41	P2,P23,P40,P43,P69,P77,P83
	Visual	P51,P81,P85	P4,P40,P69

7.5 Discussion

As hypothesized, users with a low MS prefer a brief explanation which does not require domain knowledge. Low MS users reported that they prefer Brief because this explanation is easier to understand and that they dislike that scatter plot in Combination. These results show empirical proof for the guidelines of Ribera et al. [77] and are most likely attributed to the lack of domain knowledge which makes the scatter plot and bar chart more difficult to understand at a glance than the goodness of fit.

For users with a high MS, we could not confirm our hypothesis that they like an explanation with an interactive visualization and which requires domain knowledge. Even as the thematic analysis showed that ten high MS users reported that they like the visualization and the mix of information, there was an equal amount of high MS users reporting that they liked Brief because it is easier. From the logging data, it is clear that the same amount of high and low MS users interacted with the scatter plot. The only difference is that high MS users interacted slightly more. A possible reason might be that the need for explanations with an interactive visualization and which contain domain information to steer the recommendation process is dependent on the recommendations. It might be that high MS users only want to interact with the scatter plot when the recommendations do not match the preference of the users or when they receive unexpected recommendations. Future re-

search should investigate the interaction effect of recommendation quality and serendipity with the preference of explanation for high MS users.

7.6 Design suggestions

Our results show that low MS users prefer Brief over Combination and report that they prefer these explanations because it is easy to understand and because they dislike the scatter plot, and as such we recommend providing low MS users with a brief explanation which does not require music domain knowledge because. As the preference of high MS users seems to be divided, we would recommend allowing these users to switch between two kinds of explanations: (i) a brief explanation that does not require domain knowledge because they are easy to understand and (ii) an interactive explanation that relies on domain knowledge such as a scatter plot because this explanation provides a mix of information and is more visual.

8 Experiment 3: openness

8.1 Hypotheses

For openness, we have two hypotheses to design explanations for the needs of low and high openness users:

- *H5: users with lower levels of openness prefer an explanation with a lower number of explanation elements over explanations with a higher number of explanation elements that support exploration.*
- *H6: users with higher levels of openness prefer an explanation with a higher number of explanation elements that also support exploration over an explanation with a lower number of explanation elements that does not support exploration.*

The reasoning behind H5 and H6 is mainly based on the results of two studies: First, Kouki et al. [44] provided evidence of a positive correlation between openness and the number of explanation styles perceived as persuasive. As a consequence, we hypothesize that users with low openness prefer a low number of explanation elements while high openness users prefer a high number of explanation elements. Second, a study of Chen et al. [14] found a positive correlation between openness and the preferred diversity which can be achieved by exploring the recommendations. Additionally, there is empirical evidence that low openness users have a higher intention to use again a system with explanations than without, even as they do not find more novel songs [55]. As a consequence, we hypothesize that users with high openness prefer an explanation that supports diversity while low openness users do not.

Table 10: Overview of the explanation elements and the delivery methods for the third user study

		Low MS	High MS
Explanation elements	Bar chart	•	•
	Fit		•
	Scatter plot		•
	Source song		•
Delivery method	One at a time	•	•
	All at once	•	•

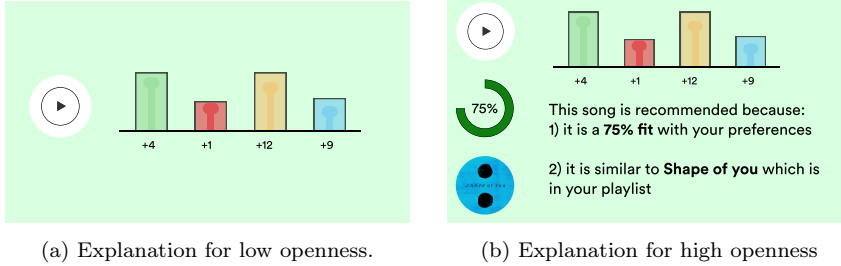


Fig. 10: Explanations for low and high openness

8.2 Design

As shown in Table 10, we designed explanations for low openness users by providing only one explanation element, namely the bar chart. We choose the bar chart as this would still allow users to control the audio features and thus limit the risk of perceived loss of control. This explanation is shown in Figure 10a and is called *Bar chart*.

For high openness users, we needed an explanation that supported exploration and that provided many different explanation sources. To support the exploration, we provided the scatter plot as this visualization can serve the purpose of a diversity-oriented recommendation explanation [88]. Additionally, we included all explanation elements to explain an individual song as shown in Table 10 and Figure 10b. We will call this explanation *All*.

8.3 Participants

Ninety participants (33F), with an average age of 29.3, completed this experiment without missing one of the three attention checks in an average time of 17min58s. The median openness score of participants in this experiment was 27 which is the same as reported by [6]. Based on this score, we performed a median split resulting in a group of 48 and 42 participants for low and high openness respectively. The distribution of the openness score of participants is shown in Figure 11.

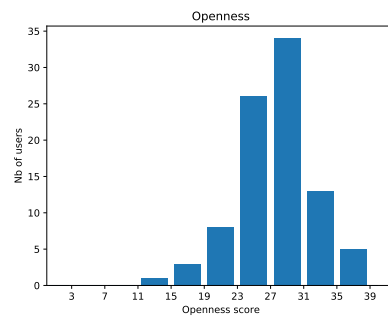


Fig. 11: Histogram of openness score of participants

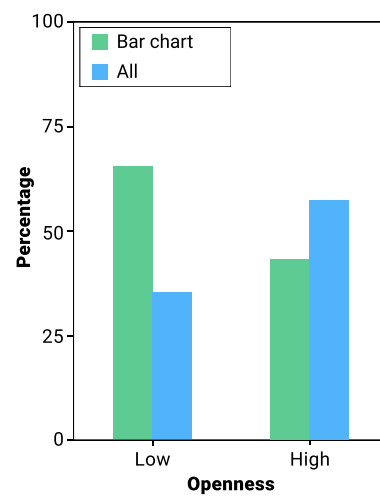


Fig. 12: Preference for Bar chart or All for low and high openness

8.4 Results

As shown in Figure 12, users with low openness preferred Bar chart over All, while this was the opposite for users with high openness. A chi-square goodness of fit test revealed we could confirm H5 as this preference was significant for low openness users $\chi(1) = 4.08, p < .043$, but that we could not confirm H6 as for high openness users this difference was not significant $\chi(1) = 0.86, p < .350$.

As a follow-up, we analyzed the logging data of the use of the scatter plot in All. As shown in Table 11, a lower number of high openness users interacted with the scatter plot than the low openness users and the number of interactions with the scatter plot was similar for both groups.

To analyze why users preferred Bar chart or All, we first analyzed and categorized the answers of the participants per preferred explanation and listed afterward which of these participants belonged to the low or high openness group. The results of this analysis are shown in Table 12.

Bar chart The most reported reason for preferring Bar chart is that it is the **easiest** explanation to **understand**. This was both reported by low as high openness users. For example, P34 reported “*I feel like the bars were easiest to understand by far [...]*”. Participants also reported that the bar chart is not only easy to understand, but also contains all the **essential information** and that the other information in All is not useful. For example P5 mentions “*The Bar chart gave me the most push to try to even preview the song. The other options felt like useless information.*” and P14 reports “*Because it [Bar chart] is easier to understand, sometimes more information makes it difficult to make decisions*”.

All We identified three different themes in the reported reasons why participants preferred All. The first theme was that All provides **more information** than just the bar charts. “*I like several details and information*” (P13). A second theme we could identify was that participants preferred the different **visual representations** of the explanations: “*I tend to read charts and graphs better than text. They make more sense to me.*” (P7). Similar to this theme, some participants such as P18 mentioned specifically that they like the **scatter plot** “*The cover is not necessary but the scatter plot adds useful information.*”

Table 11: Number (nb) of interactions with the scatter plot for low and high openness users in All

	Nb of users that interacted with scatter plot	Avg. nb of interactions with scatter plot
Low openness	60%	4.33 ± 2.0
High openness	55%	4.39 ± 6.3

Table 12: Themes identified in the question why users preferred Bar chart or All

Interface	Theme	Low openness	High openness
Bar chart	Easy to understand	P9,P21,P34	P4,P14,P38,P39
	Essential info	P25,P36	P5,P6,P14,P90
All	More information	P20,P33	P12,P13,P29,P32,P65,P89
	Visual processing	P11	P7,P31,P37
	Scatter plot	P29	P18,P19,P28

8.5 Discussion

Our results show that low openness users indeed prefer an explanation with a low number of explanation elements such as Bar chart. As the reason for this preference, they reported that Bar chart is easy to understand and does only show essential information. This confirms the finding of Kouki et al. [44] that users with lower levels of openness indeed prefer a lower number of explanation styles.

Even as our results show that there is a trend that high openness users prefer a higher number of explanations because these contain more information and because they liked the visual encoding of the information in the scatter plot, this trend is not significant and thus we needed to reject H6. This is in contrast with the finding of Kouki et al. [44] who found that users high in openness were more persuaded by many explanation styles. A possible reason for this might be that users perceive many explanations as more persuasive, but that users do not necessarily prefer the most persuasive explanation. Another reason might be that All was too complex, required too much cognitive effort, or caused information overload.

From the analysis of the logging data, it seems that high openness users did not interact more with the scatter plot than low openness users. A possible reason for this might be that that high openness users want more diverse recommendations [14] and like the scatter plot to visualize this diversity, but that they do not want to do the effort to look for more diverse songs by exploring the recommendations.

8.6 Design suggestions

As this experiment showed that low openness users preferred Bar chart because it is easy to understand and because it contains the essential information, we recommend providing explanations with a single explanation element.

For high openness users, we identified a trend that they prefer explanations with more explanation elements, especially if the information is presented visually. However, this trend was not significant, so we recommend providing high openness users the choice between explanations with a low and explanations with a high number of explanation elements. For the explanation with multiple explanation elements, we also recommend providing visual explanation elements such as a scatter plot.

9 General discussion and conclusion

9.1 Need for cognition

In our first experiment, we designed explanations for low and high NFC. For low NFC users, we provided explanations on-demand that could only be ac-

cessed for each recommendation individually. For high NFC users, explanations were provided up-front for all recommendations at once but they could be turned off. Our results show that both low and high NFC users prefer the explanations up-front for all recommendations at once. As a consequence, there is no need of adapting the explanations to NFC. Instead, we recommend providing explanations up-front for all recommendations at once that can be turned on or off to help users compare different recommendations or to understand the recommender system. To avoid information overload, we also recommend providing the option to access an explanation for a recommended song individually.

9.2 Musical sophistication

For MS, we designed explanations for different levels by providing a brief explanation that does not require domain knowledge for low MS users (Brief) and an interactive explanation that relies on domain knowledge for high MS users (Combination). Our results show that most users with low MS prefer Brief and that one of the reasons for this is that it was easy to understand. For users with high MS, the difference in preference between Brief and Combination is not significant. Some participants prefer Brief because they are easy to understand, while others prefer Combination because of the mix of information available. Further research is needed to investigate whether this preference is moderated by the quality of recommendations or other factors.

9.3 Openness

We designed an explanation for low openness users that does only contain one explanation element (Bar chart). For high openness users, we designed explanations with different explanation elements and that support exploration (All). Our results found that low openness users prefer the Bar chart over All and report that this explanation is easy to understand, and contains all essential information. For high openness users, our results show that slightly more than half of the users prefer All over Bar chart. As the reason for this preference, they report that it provides more information or that they prefer to process visual explanations. However, the others prefer explanations with only one explanation element for the same reasons as low openness users. As a consequence, we recommend providing explanations with only one explanation element for low openness users.

For high openness users, we recommend providing users with the choice between explanations with more explanation elements that contain visual information and explanations with only one explanation element.

9.4 Limitations and further research

One of the limitations of this study is that due to the recruiting through Amazon Mechanical Turk, there might be a bias towards lower-income, American and Indian users [30]. Additionally, due to the use of a median split our results even more dependent on the characteristics of the recruited participants. As a consequence, this might limit the extent to which our results can be generalized. Additionally, our results are obtained in the specific context of creating a playlist in a music recommender interface which also limits the extent to which our results can be generalized. Further research should replicate this study in different domains with different user groups to validate the design suggestions provided in this study.

A second limitation is that we only designed explanations for personal characteristics separately and thus do not investigate the interaction effect of different personal characteristics on preference for different explanations. The investigation of these interaction effects might be an interesting direction for future research.

A third limitation of this study is that our results only show a correlation and do not prove causality between the preference and the personal characteristics. However, we argue that personal characteristics are at least one of the explaining factors for the different preferences as our hypotheses have fundamentals in the theory behind the personal characteristics. Additionally, another limitation regarding our results is fact that H3-H6 contain multiple elements (i.e. interactivity and music domain knowledge) which makes it difficult to disentangle the individual influence of these elements on the results. Moreover, a more in-depth analysis of the preference might have been possible if we would have asked the users to rate both interfaces instead of forcing them to choose one of the two interfaces. Our results might also been biased because we administered the questionnaires before the main task which might have influenced the participants.

A fourth limitations of this study is that it only reports the results of three short-term experiments which investigate the perception of explanations in a first visit of the system. A long-term study could shed light into the evolution of the preference of personalized explanations in music recommender systems.

Another limitation that is inherent to the problem of designing and evaluating an element of an interface is the possible influence of different other interface elements. For example, we focused on scrutable explanations, but did not investigate the effect of the control elements on the preference of the explanations. Additionally, even as we conducted a pilot study to eliminate usability issues, design choices such as the size or the color of some elements might have influenced the preference of explanations. Moreover, we limited our study to four different explanation elements so further research should investigate if other explanation elements confirm or reject our results.

References

1. Alba, J.W., Hutchinson, J.W.: Dimensions of consumer expertise. *Journal of consumer research* **13**(4), 411–454 (1987)
2. Anderson, I., Gil, S., Gibson, C., Wolf, S., Shapiro, W., Semerci, O., Greenberg, D.M.: “just the way you are”: Linking music listening on spotify and personality. *Social Psychological and Personality Science* p. 1948550620923228 (2020)
3. Asendorpf, J.B., Neyer, F.J.: *Psychologie der Persönlichkeit*. Springer-Verlag (2012)
4. Aykin, N.M., Aykin, T.: Individual differences in human-computer interaction. *Computers & industrial engineering* **20**(3), 373–379 (1991)
5. Barbaranelli, C., Caprara, G.V.: Studies of the big five questionnaire. *Big five assessment* pp. 109–124 (2002)
6. Benet-Martinez, V., John, O.P.: Los cinco grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the big five in spanish and english. *Journal of personality and social psychology* **75**(3), 729 (1998)
7. Bettman, J.R., Luce, M.F., Payne, J.W.: Constructive consumer choice processes. *Journal of consumer research* **25**(3), 187–217 (1998)
8. Boy, J., Rensink, R.A., Bertini, E., Fekete, J.D.: A principled way of assessing visualization literacy. *IEEE transactions on visualization and computer graphics* **20**(12), 1963–1972 (2014)
9. Braun, V., Clarke, V.: Using thematic analysis in psychology. *Qualitative research in psychology* **3**(2), 77–101 (2006)
10. Burnett, M.: Explaining ai: fairly? well? In: *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pp. 1–2 (2020)
11. Cacioppo, J.T., Petty, R.E., Feng Kao, C.: The efficient assessment of need for cognition. *Journal of personality assessment* **48**(3), 306–307 (1984)
12. Cattell, R.B.: The description of personality: Principles and findings in a factor analysis. *The American journal of psychology* **58**(1), 69–90 (1945)
13. Chang, S., Harper, F.M., Terveen, L.G.: Crowd-based personalized natural language explanations for recommendations. In: *Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 175–182 (2016)
14. Chen, L., Wu, W., He, L.: How personality influences users’ needs for recommendation diversity? In: *CHI’13 extended abstracts on human factors in computing systems*, pp. 829–834 (2013)
15. Chen, X., Chen, H., Xu, H., Zhang, Y., Cao, Y., Qin, Z., Zha, H.: Personalized fashion recommendation with visual explanations based on multimodal attention network: Towards visually explainable recommendation. In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 765–774 (2019)
16. Chernev, A.: When more is less and less is more: The role of ideal point availability and assortment in consumer choice. *Journal of consumer Research* **30**(2), 170–183 (2003)
17. Conati, C., Barral, O., Putnam, V., Rieger, L.: Toward personalized xai: A case study in intelligent tutoring systems. *Artificial Intelligence Journal* (2021)
18. Conati, C., Carenini, G., Hoque, E., Steichen, B., Toker, D.: Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. In: *Computer Graphics Forum*, vol. 33, pp. 371–380. Wiley Online Library (2014)
19. Coutinho, S., Wiemer-Hastings, K., Skowronski, J.J., Britt, M.A.: Metacognition, need for cognition and use of explanations during ongoing learning and problem solving. *Learning and Individual Differences* **15**(4), 321–337 (2005)
20. Ekstrom, R.B., Dermen, D., Harman, H.H.: *Manual for kit of factor-referenced cognitive tests*, vol. 102. Educational testing service Princeton, NJ (1976)
21. Fernández-Tobías, I., Braunhofer, M., Elahi, M., Ricci, F., Cantador, I.: Alleviating the new user problem in collaborative filtering by exploiting personality information. *User Modeling and User-Adapted Interaction* **26**(2-3), 221–255 (2016)
22. Ferwerda, B., Graus, M.: Predicting musical sophistication from music listening behaviors: a preliminary study. *arXiv preprint arXiv:1808.07314* (2018)

23. Ferwerda, B., Schedl, M.: Personality-based user modeling for music recommender systems. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 254–257. Springer (2016)
24. Goker, M., Thompson, C.: The adaptive place advisor: A conversational recommendation system. In: Proceedings of the 8th German Workshop on Case Based Reasoning, pp. 187–198. Citeseer (2000)
25. Goldberg, L.R.: An alternative" description of personality": the big-five factor structure. *Journal of personality and social psychology* **59**(6), 1216 (1990)
26. Gosling, S.D., Rentfrow, P.J., Swann Jr, W.B.: A very brief measure of the big-five personality domains. *Journal of Research in personality* **37**(6), 504–528 (2003)
27. Graus, M., Ferwerda, B.: 1 theory-grounded user modeling for personalized hci. *Personalized Human-Computer Interaction* (2019)
28. Greb, F., Schlotz, W., Steffens, J.: Personal and situational influences on the functions of music listening. *Psychology of Music* **46**(6), 763–794 (2018)
29. Gunning, D.: Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA), nd Web **2**(2) (2017)
30. Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., Bigham, J.P.: A data-driven analysis of workers' earnings on amazon mechanical turk. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2018)
31. Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining collaborative filtering recommendations. In: Proceedings of the 2000 ACM conference on Computer supported cooperative work, pp. 241–250. ACM (2000)
32. Hu, R., Pu, P.: Using personality information in collaborative filtering for new users. *Recommender Systems and the Social Web* **17** (2010)
33. Jannach, D., Jugovac, M., Nunes, I.: 5 explanations and user control in recommender systems. *Personalized Human-Computer Interaction* p. 32 (2019)
34. Jin, Y., Tintarev, N., Htun, N.N., Verbert, K.: Effects of personal characteristics in control-oriented user interfaces for music recommender systems. *User Modeling and User-Adapted Interaction* **30**(2), 199–249 (2020)
35. Jin, Y., Tintarev, N., Verbert, K.: Effects of individual traits on diversity-aware music recommender user interfaces. In: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, pp. 291–299 (2018)
36. Jin, Y., Tintarev, N., Verbert, K.: Effects of personal characteristics on music recommender systems with different levels of controllability. In: Proceedings of the 12th ACM Conference on Recommender Systems, pp. 13–21 (2018)
37. John, O.P., Donahue, E.M., Kentle, R.L.: The big five inventory—versions 4a and 54 (1991)
38. Johnson, J.A.: Web-based personality assessment. In: 71st annual meeting of the eastern psychological association, Baltimore, MD (2000)
39. Jugovac, M., Jannach, D.: Interacting with recommenders—overview and research directions. *ACM Transactions on Interactive Intelligent Systems (TiiS)* **7**(3), 1–46 (2017)
40. Jugovac, M., Nunes, I., Jannach, D.: Investigating the decision-making behavior of maximizers and satisficers in the presence of recommendations. In: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, pp. 279–283 (2018)
41. Kamis, A., Davern, M.J.: Personalizing to product category knowledge: exploring the mediating effect of shopping tools on decision confidence. In: 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the, pp. 10–pp. IEEE (2004)
42. Knijnenburg, B.P., Reijmer, N.J., Willemsen, M.C.: Each to his own: how different users call for different interaction methods in recommender systems. In: Proceedings of the fifth ACM conference on Recommender systems, pp. 141–148 (2011)
43. Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., Newell, C.: Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* **22**(4-5), 441–504 (2012)
44. Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J., Getoor, L.: Personalized explanations for hybrid recommender systems. In: Proceedings of the 24th International Conference on Intelligent User Interfaces, pp. 379–390 (2019)

45. Kramer, T.: The effect of measurement task transparency on preference construction and evaluations of personalized recommendations. *Journal of Marketing Research* **44**(2), 224–233 (2007)
46. Kunkel, J., Donkers, T., Michael, L., Barbu, C.M., Ziegler, J.: Let me explain: Impact of personal and impersonal explanations on trust in recommender systems. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12 (2019)
47. Lallé, S., Conati, C., Carenini, G.: Impact of individual differences on user experience with a visualization interface for public engagement. In: *Proc. of UMAP '17*, pp. 247–252. ACM (2017)
48. Lu, Y., Dong, R., Smyth, B.: Why i like it: multi-task learning for recommendation and explanation. In: *Proceedings of the 12th ACM Conference on Recommender Systems*, pp. 4–12 (2018)
49. Mahmood, T., Ricci, F., Venturini, A., Höpken, W.: Adaptive recommender systems for travel planning. *Information and Communication Technologies in Tourism* **8**, 1–11 (2008)
50. Martin, B.A., Lang, B., Wong, S., Martin, B.A.: Conclusion explicitness in advertising: The moderating role of need for cognition (nfc) and argument quality (aq) on persuasion. *Journal of Advertising* **32**(4), 57–66 (2003)
51. Matthews, G., Deary, I.J., Whiteman, M.C.: *Personality traits*. Cambridge University Press (2003)
52. McInerney, J., Lacker, B., Hansen, S., Higley, K., Bouchard, H., Gruson, A., Mehrotra, R.: Explore, exploit, and explain: personalizing explainable recommendations with bandits. In: *Proceedings of the 12th ACM Conference on Recommender Systems*, pp. 31–39 (2018)
53. Millecamp, M., Haveneers, R., Verbert, K.: Cogito ergo quid? the effect of cognitive style in a transparent mobile music recommender system. In: *Proceedings of the 28th Conference on User Modeling, Adaptation and Personalization* (2020)
54. Millecamp, M., Htun, N.N., Conati, C., Verbert, K.: To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 397–407 (2019)
55. Millecamp, M., Htun, N.N., Conati, C., Verbert, K.: What’s in a user? towards personalising explanations for music recommender interfaces. In: *Proceedings of the 28th Conference on User Modeling, Adaptation and Personalization* (2020)
56. Millecamp, M., Htun, N.N., Jin, Y., Verbert, K.: Controlling spotify recommendations: effects of personal characteristics on music recommender user interfaces. In: *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, pp. 101–109 (2018)
57. Millecamp, M., Naveed, S., Verbert, K., Ziegler, J.: To explain or not to explain: the effects of personal characteristics when explaining feature-based recommendations in different domains. In: *CEUR workshop proceedings*. CEUR (2019)
58. Moss, A.J., Rosenzweig, C., Robinson, J., Litman, L.: Is it ethical to use mechanical turk for behavioral research? relevant data from a representative survey of mturk participants and wages. *PsyArXiv* (2020)
59. Muhammad, K., Lawlor, A., Rafter, R., Smyth, B.: Great explanations: Opinionated explanations for recommendations. In: *International Conference on Case-Based Reasoning*, pp. 244–258. Springer (2015)
60. Müllensiefen, D., Gingras, B., Musil, J., Stewart, L.: The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS one* **9**(2), e89642 (2014)
61. Naiseh, M., Jiang, N., Ma, J., Ali, R.: Explainable recommendations in intelligent systems: delivery methods, modalities and risks. In: *International Conference on Research Challenges in Information Science*, pp. 212–228. Springer (2020)
62. Naiseh, M., Jiang, N., Ma, J., Ali, R.: Personalising explainable recommendations: Literature and conceptualisation. In: *World Conference on Information Systems and Technologies*, pp. 518–533. Springer (2020)
63. Naveed, S., Donkers, T., Ziegler, J.: Argumentation-based explanations in recommender systems: Conceptual framework and empirical results. In: *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, pp. 293–298 (2018)

64. Nenkov, G.Y., Morrin, M., Schwartz, B., Ward, A., Hulland, J.: A short form of the maximization scale: Factor structure, reliability and validity studies. *Judgment and Decision making* **3**(5), 371–388 (2008)
65. Nguyen, T.N., Ricci, F.: A chat-based group recommender system for tourism. *Information Technology & Tourism* **18**(1-4), 5–28 (2018)
66. Norman, D.A., Draper, S.W.: *User centered system design: New perspectives on human-computer interaction*. CRC Press (1986)
67. Nowell, L.S., Norris, J.M., White, D.E., Moules, N.J.: Thematic analysis: Striving to meet the trustworthiness criteria. *International journal of qualitative methods* **16**(1), 1609406917733847 (2017)
68. Nunes, I., Jannach, D.: A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction* **27**(3-5), 393–444 (2017)
69. Nunes, M.A.S., Hu, R.: Personality-based recommender systems: an overview. In: *Proceedings of the sixth ACM conference on Recommender systems*, pp. 5–6 (2012)
70. Nunes, M.A.S.N.: *Recommender systems based on personality traits*. Ph.D. thesis, Universite Montpellier II-Sciences et Techniques du Languedoc (2008)
71. Ones, D.S., Dilchert, S., Viswesvaran, C., Salgado, J.F.: Cognitive abilities. *Handbook of employee selection* pp. 255–275 (2010)
72. O'Connor, C., Joffe, H.: Intercoder reliability in qualitative research: debates and practical guidelines. *International Journal of Qualitative Methods* **19**, 1609406919899220 (2020)
73. Pacini, R., Epstein, S.: The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of personality and social psychology* **76**(6), 972 (1999)
74. Paramythis, A., Weibelzahl, S., Masthoff, J.: Layered evaluation of interactive adaptive systems: framework and formative methods. *User Modeling and User-Adapted Interaction* **20**(5), 383–453 (2010)
75. Paunonen, S.V., Ashton, M.C.: The nonverbal assessment of personality; the npq and the ff-npq. *Big five assessment* pp. 171–190 (2002)
76. Ribeiro, M.T., Singh, S., Guestrin, C.: " why should i trust you?" explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144 (2016)
77. Ribera, M., Lapedriza, A.: Can we do better explanations? a proposal of user-centered explainable ai. In: *IUI Workshops* (2019)
78. Riding, R.J.: On the nature of cognitive style. *Educational psychology* **17**(1-2), 29–49 (1997)
79. Scholz, M., Dorner, V.: Estimating optimal recommendation set sizes for individual consumers. In: *Proceedings of the International Conference on Information Systems*, pp. 2440–2459 (2012)
80. Springer, A., Whittaker, S.: Making transparency clear. In: *Algorithmic Transparency for Emerging Technologies Workshop*, p. 5 (2019)
81. Springer, A., Whittaker, S.: Progressive disclosure: empirically motivated approaches to designing effective transparency. In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pp. 107–120 (2019)
82. Tam, K.Y., Ho, S.Y.: Web personalization as a persuasion strategy: An elaboration likelihood model perspective. *Information systems research* **16**(3), 271–291 (2005)
83. Tintarev, N., Dennis, M., Masthoff, J.: Adapting recommendation diversity to openness to experience: A study of human behaviour. In: *International Conference on User Modeling, Adaptation, and Personalization*, pp. 190–202. Springer (2013)
84. Tintarev, N., Masthoff, J.: A survey of explanations in recommender systems. In: *2007 IEEE 23rd international conference on data engineering workshop*, pp. 801–810. IEEE (2007)
85. Tintarev, N., Masthoff, J.: Effects of individual differences in working memory on plan presentational choices. *Frontiers in psychology* **7** (2016)
86. Tkalčič, M., Ferwerda, B., Hauger, D., Schedl, M.: Personality correlates for digital concert program notes. In: *International Conference on User Modeling, Adaptation, and Personalization*, pp. 364–369. Springer (2015)

87. Tong, S.T., Corriero, E.F., Matheny, R.G., Hancock, J.T.: Online daters' willingness to use recommender technology for mate selection decisions. In: *IntrRS@ RecSys*, pp. 45–52 (2018)
88. Tsai, C.H., Brusilovsky, P.: Beyond the ranked list: User-driven exploration and diversification of social recommendation. In: *23rd International Conference on Intelligent User Interfaces*, pp. 239–250. ACM (2018)
89. Vogel, E.K., Woodman, G.F., Luck, S.J.: Storage of features, conjunctions, and objects in visual working memory. *Journal of experimental psychology: human perception and performance* **27**(1), 92 (2001)
90. Völkel, S.T., Schödel, R., Buschek, D., Stachl, C., Au, Q., Bischl, B., Bühner, M., Hussmann, H.: 2 opportunities and challenges of utilizing personality traits for personalization in hci. *Personalized Human-Computer Interaction* p. 31 (2019)
91. Wörndl, W., Lamche, B.: User interaction with context-aware recommender systems on smartphones. *i-com* **14**(1), 19–28 (2015)