The Role of Fairness and Diversity in User Choices and Perceptions of Music Playlists

Karlijn Dinnissen[*]
Utrecht University
Utrecht, The Netherlands
k.dinnissen@uu.nl

Shah Noor Khan[*]
Utrecht University
Utrecht, The Netherlands
s.n.khan1@uu.nl

Hanna Hauptmann
Utrecht University
Utrecht, The Netherlands
h.j.hauptmann@uu.nl

Eelco Herder
Utrecht University
Utrecht, The Netherlands
e.herder@uu.nl

Judith Masthoff
Utrecht University
Utrecht, The Netherlands
j.f.m.masthoff@uu.nl

Abstract

Nowadays, most listeners access music through streaming platforms, which has transformed how recommendations are delivered and received. Notable work has been done on improving fairness for end users or item providers in music recommender systems, often applying pre-processing and re-ranking approaches. However, the positive influence users may exert on fairness through their music selection has not been sufficiently studied. This paper explores whether providing users with insights into the fairness and diversity of recommendations could lead to fairer music selections. To this aim, we conducted a qualitative study with 18 participants involving a think-aloud playlist task and in-depth interviews. The results indicate that while music taste was the deciding factor in participants' choices, insights into fairness did help them reflect on their selections. Our study shows that participants' playlist choices at times conflict with their expressed fairness values, highlighting the need to support users in aligning decisions with their values.

CCS Concepts

• **Information systems → Recommender systems**; • **Human-centered computing**;

Keywords

Music Recommender Systems, Fairness, Transparency, Human-Centered Computing, Interviews, Think-aloud Study

ACM Reference Format:
Karlijn Dinnissen, Shah Noor Khan, Hanna Hauptmann, Eelco Herder, and Judith Masthoff. 2025. The Role of Fairness and Diversity in User Choices and Perceptions of Music Playlists. In *Proceedings of the 36th ACM Conference on Hypertext and Social Media (HT 2025), September 15–18, 2025, Chicago, IL, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3720553.3746671

[*]Both authors contributed equally to this research.

HT 2025, Chicago, IL, USA

ACM ISBN 979-8-4007-1534-1/2025/09
https://doi.org/10.1145/3720553.3746671

1 Introduction

Music is an integral part of many people's lives and identities. Nowadays, they consume music mainly via streaming platforms [26]. As the platforms' recommender systems (RS) help users follow their preferences, those systems greatly impact which music is played [55]. However, RS may amplify filter bubbles [49] and reinforce several biases [52]. Artists that are new, less popular, or beyond-mainstream in terms of, e.g., genre, nationality, or gender may be underexposed [51, 57]. Similarly, users often receive less fitting recommendations when they have a minority background or taste [20, 41]. Such unfairness may be countered algorithmically [14].

While some studies explore how artists want to be fairly represented [16, 24], it remains unclear whether and how music recommender systems' (MRS) *end users* perceive (un)fairness. They may be unsatisfied when they perceive recommendations as less accurate [39]. Also, studies show that users do not always perceive the increased fairness recommendations [25, 66].

Users are often unaware that their choices significantly impact what RS recommend, and, in turn, how fair those recommendations are [62]. Therefore, one way to improve fairness is encouraging users to to choose music in line with their values. Providing fairness-related information about playlists may help them make more informed, fairer choices [6]. As the effectiveness of music choice nudges is person-dependent, personalization of nudges shows promise [21]. However, it is still unknown what users would perceive as a fair MRS, how this impacts their choices, and how they would want to be nudged toward fairness.

We address this research gap and provide insights that could increase fairness for artists and users of MRS. We investigate whether users are aware of unfairness towards artists and themselves in MRS, and how information on fairness influences users' perceptions and choices. To this aim, we conducted a qualitative study with 18 participants that involved a think-aloud playlist listening and selection task, along with in-depth interviews to reflect on the task, the provided descriptions, and fairness as a concept. Our results highlight the importance of users' personal values when counteracting fairness-related issues. We contribute 1) an overview of users' fairness-related values and how those impact playlist selections, 2) insights into decision-making with and without fairness-related descriptions, and 3) users' perceptions on the division of responsibilities concerning fairness-related issues.

2 Related Work

Music streaming platforms serve multiple stakeholders, including end users, music artists (i.e., item providers) and the platforms themselves [2, 48]. Since MRS play a key role in music discovery, it is crucial to consider their impact on all stakeholders when designing and evaluating them [2, 55]. However, MRS evaluation remains mainly quantitative [70], seldom addressing user perception, and rarely considering multiple stakeholders [3, 15, 56]. Below, we further outline the gap in MRS evaluation concerning fairness, transparency, and control.

2.1 Fairness of Music Recommender Systems

Fairness as a normative human concept should be defined for each specific domain and context [14, 19]. Fairness research often focuses on sub-groups of users or item providers who are underrepresented or underserved, aiming to identify and promote fairer outcomes [19]. Translating such goals into practice, and accurately measuring them, remains challenging, as demonstrated for dating apps [61]. For MRS, music artists identify popularity, niches, and imbalanced representations as underlying causes of unfairness [16, 24].

Popularity bias arises when already popular items are recommended more often because of their higher presence in past interactions [1, 11, 42]. MRS are known to suffer from such bias as well [44, 45]. The platforms' responsibility in shaping a fair music ecosystem for artists by integrating more lesser-known items is recognized by both artists [16, 24] and music industry professionals [18]. For end users, popularity bias can lead to homogenization, as RS show a bias towards items preferred by the majority, making it more likely for minority users to receive poor recommendations [42, 68]. Still, algorithmically aligning RS with each user's content consumption may not sufficiently reduce such bias, as many users inherently prefer popular items [40].

Niche or non-mainstream music or genres are generally less well-presented in MRS datasets and recommendations themselves [41, 42], leading to underserved artists and genres. This further contributes to a lack of item diversity in recommendations, which in turn could harm user satisfaction [10, 53]. As such, algorithmic bias and interaction bias amplify existing imbalances in society [8].

Imbalanced representations of groups in the data may also cause unfairness, such as disparities in gender and country or nationality [23, 47, 51, 57]. These imbalances also come up in interviews with artists, but with varying responses on what would constitute a fair system in these dimensions [16, 24].

2.2 Transparency in Recommender Systems

Transparency is widely recognized as a foundational and significant principle in the development of ethical AI [36, 65]. In the domain of RS, it is considered vital to fostering user acceptance and trust [50, 59]. Transparency in a system involves explaining how it works[1] [63], which can increase user confidence and contribute to quicker decision-making [59].

Explanations play an important role in people's reception of recommendations [4, 28, 63]. Zhao et al. [71] show that natural language explanations for why a particular song is recommended to a user encourage higher engagement. Since explanations directly affect user decisions and interactions on the platform, one needs to recognize that not all users have the same goals [63].

[1]To RS developers, experts, item providers, or end users. We focus on end users here.

2.3 Control and Accountability in Decisions

Decision making. RS help with choice overload in decision-making processes by providing personalized suggestions. While much research focuses on algorithms, user traits and decision-making also shape RS effectiveness [12]. For example, personality may influence users' preferences for diverse recommendations [13, 35]. Another work shows that users often consider the entirety of a playlist when evaluating its attractiveness and diversity [30].

Control for RS users can be increased by allowing them to provide inputs [64] or engage in active decision-making [7]. Such increased control can positively affect the perceived quality of recommendations [31]. Some argue that creating awareness and enabling users to support their values could also improve fairness for item providers [17, 69]. However, human decisions are affected by noise and limited insight into the system [37], and more control may also increase cognitive load [34]. Moreover, in a simulation study, re-ranking methods more effectively improved gender fairness in RS over time than user choice models [22].

Accountability. Almeida et al. [5] show how algorithms influence individuals and shape society as a whole. They argue that governance strategies are needed to make organizations accountable for the consequences of their systems. Leerssen [43] critically explore RS accountability in Europe, and what types of accountability these RS adopt. However, shared accountability through added user control and its effect on RS remains underexplored.

In this paper we investigate user perception of control and accountability in MRS, exploring how fairness-focused descriptions influence decision-making [60].

3 Method

We ran a user study with 18 participants. The setup was pretested for clarity, duration, and outliers in the materials with 2 participants. The Ethics and Privacy Quick Scan by the Utrecht University Research Institute of Information and Computing Sciences deemed the research as low-risk, requiring no further assessment.

3.1 Study Setup

The study was held online through Microsoft Teams or in person at Utrecht University. Participants gave informed consent before starting. The study involved four steps, outlined in Figure 1[2] and

[2]All study materials can be accessed at https://doi.org/10.5281/zenodo.15807624.

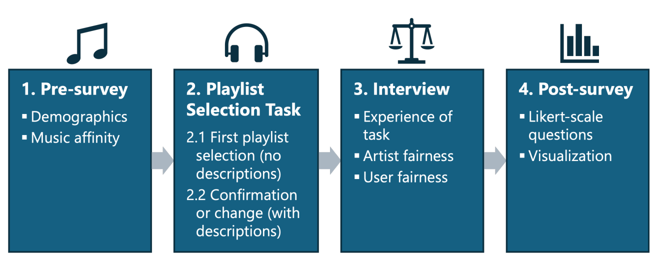

Figure 1: The four-step study procedure.

took 59 minutes on average (Step 1, 2.1, 2.2, 3, and 4, respectively took 2, 26, 5, 20, and 6 minutes on average).

Step 1: Pre-survey. Participants provided their demographics (age, gender, nationality), how much they listen to music per day (in general and attentively), whether they ever released music on a streaming platform, their familiarity with popular and niche music, their affinity to discover new music, and their preferred genres.

Step 2.1: Playlist selection Round 1. During the first round of the playlists selection task, participants were asked to select a playlist from a set of three options for 4 consecutive sets of playlists, while thinking aloud to share their decision-making process. The sets were based on fairness and diversity aspects discussed in previous MRS work (see Section 2.1): artist gender (Set 1), artist nationality (Set 2), song genre (Set 3, also as a proxy for 'mainstreaminess' [54]), and song popularity (Set 4). Within each set, 3 playlists were displayed: one that was not diverse, one moderately diverse, one very diverse. Having participants select one playlist among a couple, as opposed to rating separate songs, allowed us to recreate a more realistic music listening scenario in which we could control for overall diversity. The playlists' order in each set was randomized to prevent middle option bias [58]. Each playlist contained 5 songs. More details on the song selection procedure can be found in Section 3.2. For each song, the title, artist, and album cover were shown. Participants could listen to an audio preview of maximum 30 seconds per song, similar to a study by Fatahi et al. [21].

Step 2.2: Playlist selection Round 2. In the second round of Step 2, participants revisited playlist Sets 1–4—now accompanied by fairness descriptions (see Table 1)—and could confirm or change their initial selections. This round aimed to explore participants' responses to the fairness characteristics of the playlists, and whether they would want to change their choice after seeing those characteristics. Extending beyond previous work [6, 38], participants were asked to think aloud throughout this process.

Step 3: Semi-structured interview. We conducted a semi-structured interview to explore participants' perception of the topics and fairness dimensions that came up in Step 2. The interview covered three topics: 1) *task and description evaluation* inquired about participants' experience of Step 2, the influence of playlist descriptions on their choices, whether having playlist descriptions from the start would have influenced their choices differently, and whether they would appreciate such descriptions on music streaming platforms. 2) *MRS artist fairness* inquired about whether participants felt that artists are currently treated unfairly in MRS, and if so, in what ways—whether related to the dimensions identified in Step 2 or any others they considered relevant. 3) *MRS end user fairness* inquired about participants' perception of MRS (un)fairness towards end users, whether they see their own music taste represented, and whether MRS help them explore new music.

Step 4: Think-aloud and post-survey. Here, we provided several facts illustrating some current imbalances in the music industry, illustrating that such imbalances indeed exist, for each fairness dimension from Step 2 (e.g.: "The US and Canada combined receive a 40,9% market share of the global music industry revenue. 7,5% of the world's global population lives on this continent" [32, 67]). Participants were then asked "How important is it that music from [X] is recommended more on music streaming services?" on a 5-point Likert scale, where [X] was replaced with each of the four respective underrepresented categories. Next, participants viewed their answers plotted against answers from professionals in the music industry in [18]. Participants were asked to think aloud throughout. With these insights, we aimed to measure whether participants would see algorithmic interventions (i.e., disparate treatment [19]) in MRS as a viable way to counter current imbalances.

3.2 Song Selection

All participants received the same playlists in Step 2, which suits the study aims to investigate the influence of playlist diversity and descriptions in a controlled setting, rather than to model realistic recommendation behavior. To further minimize the influence of song characteristics, we purposefully sampled our songs as follows. All songs were released between 2013 and 2024 and available on Spotify. We chose to select songs categorized on Bandcamp[3] under 'pop' and with English lyrics, due to its wide appeal to general audiences [27]. The only exceptions to this were playlists in Set 3 (*Genre*), which contained non-pop songs, and Set 2 (*Nationality*), which contained non-English lyrics.

To minimize the impact of familiarity on participant preference [33], we selected songs from non-Dutch artists that had been played < 100.000 times on Spotify, with no other songs of the artist having over 1 million plays. One exception is the *Popularity* dimension in Set 4, where the popular songs had > 100 million Spotify plays. We used Spotify play counts to gauge familiarity as Spotify is a prominent platform in The Netherlands[4]. We consulted artist profiles on Wikipedia, Bandcamp, Spotify and YouTube to retrieve artist information on *Gender*, *Nationality* and *Genre*. To ensure that playlists in one Set addressed only one fairness dimension, the other dimensions were carefully mixed[5].

We analyzed the initial selection for cohesiveness within playlists. Since playlists perceived as too diverse are generally less well-received [53], we made sure five audio features were similar in each Set[6]. Outliers were identified using a k-nearest neighbors approach (k=4, distance-based threshold=2) applied to a UMAP-generated [46] embedding space of these audio features, and subsequently removed and replaced with different songs. The moderately and very diverse playlists in the *Genre* dimension were excluded from this analysis due to their inherent high variety. After pretesting the study, two songs were replaced due to both participants responding very negatively to them. The final playlists were rechecked to confirm sufficient cohesiveness.

3.3 Data Analysis

We report demographics, music affinity, playlist choices, and survey responses using descriptive statistics. All recordings were transcribed and coded using NVivo 14, applying thematic analysis [9]. Two coders coded two transcripts together, then coded individually and cross-validated codes afterwards.

[3]https://bandcamp.com/
[4]https://www.statista.com/topics/11066/music-streaming-services-worldwide/
[5]E.g., for Set 1 (*Gender*) the artists had varying nationalities, and for Set 3 (*Genre*) and 4 (*Popularity*), the artists were both men and women.
[6]'acousticness', 'danceability', 'energy', 'tempo', and 'valence' were retrieved from https://developer.spotify.com/ (audio feature retrieval is now deprecated).

Table 1: Descriptions used in Step 2.2, focused on the fairness/diversity aspect of the respective Set. Each Set contained a playlist for the majority group, a moderate mix, and the minority group.

Set	Dimension	Description: not diverse	Description: moderately diverse	Description: addressing imbalance
1	*Artist gender*[7]	Songs by men	Songs by men & women	Songs primarily by women
2	*Artist nationality*	Artists from the US	Moderately diverse nationalities	Highly diverse nationalities
3	*Music genre*	Single genre	Moderately diverse genres	Highly diverse genres
4	*Song popularity*	Popular artists	Popular & lesser-known artists	Lesser-known artists

Table 2: Reasons for Playlist Selection

Category	# Participants	# Mentions
Taste	17	63
Vibe	12	29
Specific song	11	22
Familiarity	9	20
Cohesiveness	9	21
Novelty	8	18
Variation	8	13

Codes from previous work were used as a starting point (deductive) [16]. Other codes were added or adapted inductively. Thematic analysis revealed five main themes, discussed in Section 4. While we note frequency for each theme, our focus remains primarily on the qualitative insights.

4 Results

Here, we present key findings organized by the main themes from our thematic analysis. For each theme, we note the study phase(s) in which it emerged, and analyze participants' verbal expressions, playlist selections, and survey responses.

4.1 Participants

Participants were recruited through convenience sampling until thematic saturation was reached. 10 were Dutch, 6 European, and 2 non-European. 9 participants identified as women and 9 as men. The average age was 30.5 years (*Med*=29.5, *SD*=7.71). On average, participants listened to music for 3.1 hours daily, of which 2.9 via a streaming service, and 1.3 attentively. 3 participants indicated having released music on a streaming platform. Our survey also included 3 Likert scale questions and a genre affinity rating:

- *Familiarity with popular music/artists:* Agree (7), Disagree (4), Neither (6).
- *Perception of their own taste as niche:* Strongly Agree (1), Agree (4), Neither (5), Disagree (6), Strongly Disagree (1).
- *Curiosity for unfamiliar music:* Strongly Agree (5), Agree (9), Neither (2), Disagree (2).
- *Avg. Genre affinity:* Pop (3.6), Jazz (3.5), Rock (3.4), Classical (3.1), Metal (3.0), Folk (2.8), Electronic (2.6), Hip-Hop (2.2).

4.2 Decision Making

Here, we report 1) what participants were thinking aloud during Steps 2.1 and 2.2, and 2) what they mentioned in the interviews in Step 3, when asked about what influenced their playlist choices.

[7]We acknowledge that these sets do not reflect the full spectrum of gender identities. Full representation was not possible due to limitations in the study's scope.

In Step 2.1, participants selected one playlist each from four sets of three playlists while thinking aloud. They discussed various aspects of the songs that they paid attention to whilst listening, including melodies (all participants), vibe (16), album covers (6), lyrics (5), and artist names (4). Then, when selecting their favorite playlist, participants mentioned similar factors (i.e., taste, familiarity, vibe, and diversity) to be important in their decision. Table 2 gives an overview of these themes. The left side of Figure 2 shows the choices users made between the three playlist of each set, with most people choosing mixed gender, moderately diverse genres, highly diverse nationalities, and highly popular artists.

Musical taste was the main factor in users' choices (16 participants) including the playlists' vibe: *P10: "When I'm choosing a playlist, I would probably go for, like, overall vibe."* Preferences included genre and style, and occasionally individual songs: *P3: "From these two playlists, I like the first two songs a little bit more, because it's like more electronic dance music, which I do tend to listen to."*.

(Un)familiarity with music was frequently noted in Sets 1–3 (with unfamiliar songs by design), by 12 participants. In Set 4 (*Song popularity*), participants often expressed joy at recognizing songs. In that Set, participants often showed a preference for known songs: *P13: "The main reason I chose this playlist is because I recognize the songs."* They spent more time on unfamiliar songs than familiar ones, sometimes replaying them: *P7: "I don't know [this song] [...] So then I'm listening longer to it. But, yeah, a lot of these songs I know, so then I don't have to listen to them very long to remember the song."* They often shared opinions about artists familiar to them. Some participants also tried to identify the playlists' themes unprompted: *P5: "This seems like an international playlist, right? [...] All the songs are different languages."*

The language of song lyrics was mentioned by 4 participants as an important aspect. For Set 2 (*Artist nationality*), some expressed a preference or dislike for music in a particular language: *P10: "I do like French music a lot"*. Overall, the playlist containing the most diverse nationalities was selected by the majority of participants, as they found it refreshing to explore songs outside of their usual rotation: *P8: "I think I'll go for this one with all the different languages. I think that would be fun."* However, they preferred to listen to diverse playlists when actively exploring new music.

In Step 2.2, participants were asked to change or confirm their selected playlists, now with each dimension explicitly labeled. This process was much quicker, as most participants (15) felt confident in their initial choices. Some re-listened to certain songs when they were uncertain about their initial selections, but nearly all participants maintained those selections, taking into account several aspects outlined in Table 2. At times, the descriptions made them

Table 3: Perception of Playlist Descriptions

Category		# Participants	# Mentions
Influenced by description	Yes	8	9
	No	16	26
Want more description	Yes	15	26
	No	5	6
Care about description	Yes	11	28
	No	16	49
Insight into ethical aspects		7	11
Surprise		6	7
Would influence others		5	5

feel happier with their choices: *P15: "Yeah, some things were good to see, that I tend to move towards playlists with diverse genres."*. At other times, it led to disappointment: *P11: "With the [popularity-based playlists] I was like: "oh, bummer". [...] I would like to listen to new music, and it's annoying [my choice is] so mainstream."* While the descriptions influenced participants' feelings about their choices, they did not affect their final selection (see Section 4.3 for details).

The perception of playlist descriptions is summarized in Table 3. 8 participants indicated feeling influenced by the descriptions, though often not for all fairness dimensions, or only for ones where they had already made the choice that corresponded to their values. Conversely, in 16 interviews, participants (also) expressed not being influenced by the description for at least one dimension. Furthermore, some stated that the descriptions did not provide information about the music itself: *P2: "Just the description didn't tell me much about if I was going to like a playlist or not. [...]'Diverse music' can have literally any music in it."* Still, 15 participants did want streaming platforms to include such descriptions more, e.g., to speed up their decision-making process: *P7: "[The descriptions] would make it easier for me to choose, because it kind of explains what I'm going to listen to."* 7 participants also mentioned ethical implications: *P6: "They made me think more ethically about the choices I make. I do think it's also just important to listen to music you like, you shouldn't just listen to music purely because it's from a certain group or not. But [it's] good to be mindful of it."* Lastly, 5 participants expressed that such descriptions would help other users to be more mindful about diversity in their own listening behavior: *P11: "I think it would be more eye opening for [some people]."*

Showing descriptions before choosing a playlist would supposedly influence some participants to choose a fairer and more diverse option from the start: *P1: "I would have had an incentive because of curiosity to look at the diverse nationality playlist"*. Table 4 shows the influence the participants estimated for each dimension. Finally, some participants only wanted descriptions in certain contexts: *P15: "If I want to have a review sometimes of my playlists, how I listen to music, how diverse it is, then those statistics would help. But [...] I wouldn't want that in my face [laughing] all the time."*

In summary, the playlist selection process was driven by diverse considerations outside the fairness dimensions. For some participants, the descriptions were helpful, but not for all participants or for all Sets. These conflicting values will be discussed next.

4.3 Conflicting Values

Here, we explore the fairness values that participants expressed in Step 3 and 4, and discuss whether those values were reflected in the choices they made in the playlist selection task (Step 2).

Table 4: Participant responses on whether a diversity-related description would have influenced their choices in Round 1.

Dimension	Would have influenced	Would not have influenced	Unknown
Gender	8	10	-
Nationality	11	6	1
Genre	11	4	3
Popularity	9	8	1

Table 5: Expressed Values of Participants

Category	# Participants	# Mentions
More less-popular artists	18	63
Gender diversity	18	45
Location diversity	16	32
Genre diversity	16	30
More outside of bubble	14	27
More less-popular songs	5	5

Expressed values were collected when interviewing participants on what they consider to be fair for artists and for users. Table 5 shows the values mentioned. Participants focused on wanting to hear more music from artists of certain underrepresented or niche groups, which they related to artist fairness concerns (calling for better representation) and touches on user fairness issues (e.g., feeling that current recommendations do not meet their needs). One key value expressed in the interviews was discovery: participants sought more recommendations beyond their usual bubble, valuing greater diversity overall. This is further addressed in Section 4.5.

Supporting lesser-known artists was mentioned by all participants. Some related this to users not being able to find them on streaming platforms easily (see Section 4.5). Others emphasize the artist unfairness perspective, mentioning that either algorithms, or how the music industry as a whole functions, cause unfair advantages for already popular artists: *P2: "If it's too focused on what everyone likes, then it's very hard for the lesser known people or lesser known genres to become more known. [...] And that's kind of unfair to the lesser known people, I think."* Some did note that recommending more popular artists is understandable from a business perspective: *P15: "It's more related to how [streaming platforms] make the money and business out of the big artists. So from their business perspective, [...] I understand, but it is not fair, obviously."*

Artists' gender was also discussed in all interviews, but participants here were less unanimous on whether they perceived the current platforms as unfair towards artists from underrepresented genders, and on whether action should be taken to address historical gender imbalances. Supporting minority genders was a priority for some: *P2: "I am all for having more female voices in music. So especially in, for example, rap is a very male dominated genre."* While for others, gender should not be a factor: *P4: "What difference should it make? [...] What does the gender have to do with anything?"*

Artists' nationality was mentioned in 16 interviews in a fairness context. 6 participants would like to be able to better discover artists from other countries. Another 7 specifically mentioned that artists outside the US (and sometimes UK and Western Europe) should be promoted: *P8: "I think we are doing the world a disservice by only selecting big pop stars from America. [...] [For example], a Dutch artist will only be pushed into the Netherlands and nowhere else. [...] We're all missing out on a lot of good music because of this policy."* Participants recognized that song lyrics affects an artist's reach, noting that the prevalence of the English language provides an advantage to artists that write their lyrics in English: *P11: "I don't like the fact that [the US and Canada have] such a big share of the market, but they do have the luxury that English is such a commonly spoken language, so people can resonate with it a bit better."*

Music genre was also mentioned in a fairness context by 16 participants. Here, the main sentiment was more nuanced, with participants expressing that they did not see it as a problem if music listeners have one preferred genre and only listen to that. They suggested giving the choice for more genre diversity to the user: *P4: "If I ask for it, then yes. Otherwise no. [...] I'd assume that the people know what they roughly want to listen to. So if that happens to be pop, yeah."* However, they noted that if one wants to discover a new genre, streaming platforms do not sufficiently cater to that need: *P2: "I also like it, but then somehow I end up only listening to pop if I only click on [recommended] stuff, because it doesn't recommend me any other stuff. And that's kind of unfair to the lesser known people."*

As previously mentioned, nearly all participants confirmed their initial playlist choices after seeing the descriptions, with only 3 exceptions. 2 changed their choice in the Set 3 (*Artist nationality*) from 'US artists' to 'highly diverse nationalities', both indicating that they were triggered to explore new content. The participant stated that their third change was unrelated to the descriptions.

The conflict between values and actions is illustrated in Figure 2, which shows the relation between the playlists selected after seeing the fairness descriptions (on the left), and participants' Step 4 responses on their level of concern regarding the fairness dimensions (on the right). Interestingly, participants' playlist choices were often not coherent with their responses in Step 4.

The Popularity dimension shows most notable difference. Here, 11 participants chose the 'Popular' playlist, even though they identified the lack of recommendations of lesser-known artists as an issue during Step 3, and in Step 4, agreed to the statement that more lesser-known artists should be promoted on streaming platforms.

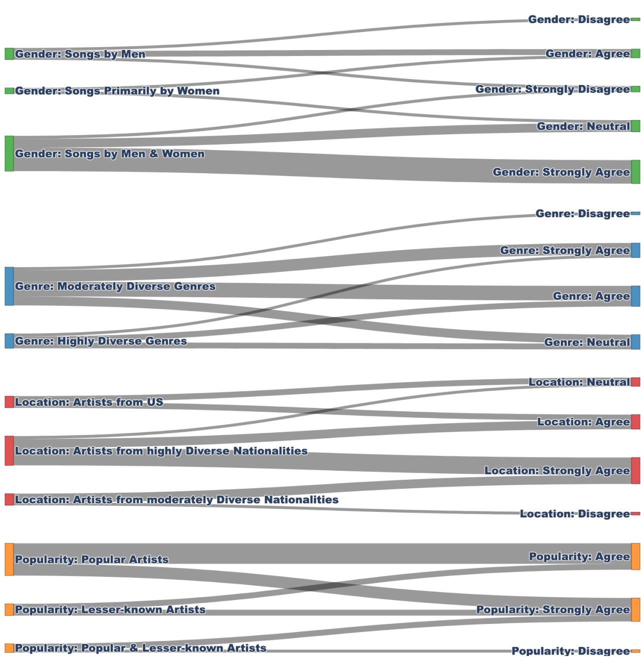

Figure 2: Sankey Diagram of Playlist Selection in Step 2.2 and Survey Responses on Disparate Treatment in Step 4

For **Gender**, Figure 2 shows that almost all who (strongly) agreed that gender balance should be improved, still preferred the mixed playlist (songs by men & women) over the playlist with songs primarily by women. Notably, the average response from the 9 male participants (avg = 3.2) was lower than that of the 9 female ones (avg = 4.3), whereas for other fairness dimensions the differences between participant genders were negligible. Regardless of their opinion, most participants chose the mixed playlist.

Lastly, regarding **Nationality**, 16 participants expressed a preference to promote non-US artists, which is in line with the majority choice for artists from diverse nationalities.

In summary, even though participants indicated to care about one or more fairness dimensions, this often did not influence their final playlist selection. Their decision process was largely driven by other considerations: *P6: "I do want more diverse types of playlists. But I still went with the one I liked."*

4.4 Causes of (Un)fairness

In addition to discovering participants' specific values (see Section 4.3), in Step 3 (interviews) we asked participants whether they had ever considered the fairness of streaming platforms and MRS towards artists and towards users. If so, they were asked about their perception of what makes such a platform fair, and whether platforms are currently fair or not. As some participants mentioned ideas on how to improve fairness, we discuss those here, too.

Fairness for artists was considered by 11 participants before participating; 6 had never considered it, and for 1 it was unclear. 15 participants said they considered the MRS integrated in streaming platforms to currently be unfair. Table 6 shows the frequency of themes coded on this topic. The most frequently mentioned cause of unfairness was popularity bias (10 participants): *P4: "Popularity is so bloody influenced by these algorithmic things apparently nowadays, it's just like: "oh, hey, look, this guy has passed a certain threshold, so we instantly recommend him to many more people." Like, I don't know, is that fair?"* Secondly, 6 participants noted that artists and labels with good connections and money get an unfair advantage: *P1: "How much money you have when you start out, this is also something that's connected to a lot of social factors, right? And whether you have access to strong support of music labels, or have to do it on your own."* Some even suspected that some artists or labels are paying streaming platforms to be recommended more: *P3: "I think music labels will probably pay Spotify money in order for them to put certain songs in certain playlists, which feels weird. Like I'm sort of getting tricked to listen to a certain song."* The third cause for unfairness, mentioned by 4, was that they felt artists get paid (too) little: *P5: "They're not paid well enough. And that's also probably because we pay too little for the streaming. And most of the money goes to popular artists or*

Table 6: Fairness for Artists

Category	# Participants	# Mentions
Profit / pay-to-play / renumeration	15	27
Algorithmic fairness	13	26
Music labels	6	8

Table 7: Fairness for Users

Category		# Participants	# Mentions
Transparency		11	23
Control		9	19
Taste recognized	Yes	8	10
	Undecided	4	4
	No	6	8
Representation	Good	13	22
	Undecided	3	4
	Bad	2	2

to the company." Only one considered MRS to be fair: *P18: "I think unfairness is removed from such a platform because anybody can post anything as long as it's within the guidelines of the service."*

Fairness towards users of music streaming platforms had never been considered by 4 participants. 5 were dissatisfied with it, while 9 found platforms mostly fair as they meet expectations and offer convenience: *P10: "All of my preferences or the types of things that I could think about for playlist, I feel like I get those."*

Table 7 shows the frequency of themes coded regarding user fairness. In terms of feeling that their taste was recognized in MRS, 8 participants agreed, 4 were undecided, and 6 disagreed, one mentioning: *P7: "I don't always feel that [the system] knows what I want to listen to at the moment. [...] But I understand [that if] the history is not representative, then it can't make a good decision for the present."*

Further, 6 participants viewed the lack of transparency or control as unfair: *P1: "I don't know why things are recommended to me. I don't think I have the option to calibrate this a lot."* 5 participants saw lack of diversity as unfair: *P11: "I'm seeking more variety and diversity and different angles, different songs [...]. It's difficult to stray away from the music norm or a music genre if you're just being fed this the whole time."* Finally, as with artists, some considered monetary interests: *P16: "What I find frustrating is that sometimes I feel like because of a business model [...] certain platforms will sort of try to force-feed a certain type of [profitable] content to you."*

Finally, 13 participants expressed being able to find music that resonated with their taste and associated community. Others mentioned being undecided about representation: *P1: "[I personally feel represented], because I'm in a relatively privileged position. [...] but I wouldn't say it's all good for everybody."* while the rest felt they were not well-represented: *P4: "I felt caged. But maybe that's because I did like too much of the same thing. And then it overfits to like: "Oh! More of this stuff, here we go." [...] It was horrible."*

In summary, participants were more aware of artist- than user-related fairness issues, focusing for the latter on limited taste capture and a desire for more discovery. We expand on this next.

4.5 Discovery and Filter Bubbles

Here, we show how participants discover music and feel supported by their streaming platform, as expressed in Step 3 (interviews).

Lack of support in discovering new music was mentioned by all participants: *P15: "They don't encourage you to listen to something different. [...] To explore something new."* They reported that streaming platforms perform better in helping them discover music close to their preferences than for music that is further away. Participants did mention that platforms recognizing their usual taste offers convenience, despite the possible drawbacks. However, to find niche music, participants indicated the need for more specific queries: *P15: "I think it's more of our own effort. And then once they recognize that this is what you're interested in, then they push it in, instead of [they] themselves taking some initiative to push something."* They believed that this is caused by RS algorithms, or the workings of the music industry, mentioning: *P2: "It knows that you like one thing, it will only recommend you that one thing. [...] it's difficult to tell the system like: "no, I don't only want this, I want some of this as well"."* They also felt like this narrow focus on their preferences gets amplified over time: *P15: "It's pushing the same thing again and again and now it's been to the point when I feel like they're just repeating the same stuff [...]. So that is kind of frustrating"*

More control by giving users the ability to set their own selection criteria was seen as ideal, as participants felt that currently the algorithm is setting the criteria for them: *P1: "At the very least, the user should have the option of choosing certain criteria that they want to not impact the recommendations [...]. So, not only being transparent [...], but also allowing the user to completely influence them."*

In summary, while participants felt streaming platforms help them discover music and recognize their taste, downsides of such platforms are that they lack diversity and control.

4.6 Who Is Responsible

Finally, we examine where participants feel the responsibility lies to make music streaming platforms and MRS fairer, using data from Step 3 (interviews) and Step 4 (survey and think-aloud).

The role of streaming platforms in potentially improving fairness was mentioned by 11 participants. 5 saw them as key stakeholders responsible for implementing changes to enhance fairness: *P5: "It should be also the job of, like, Spotify, to [...] not amplify popular artists, but to give a chance for smaller artists."* 3 did not consider them responsible, and 3 were unsure.

Regarding the role of other stakeholders, some viewed the music industry as responsible, especially for countering inequalities between popular and lesser-known artists: *P16: "I think it's more the music industry as a whole that needs to do something about it. The streaming [platform] is at the very end of the chain."* P12 noted that inequalities should be solved at their source, mentioning both algorithms and society as a whole: *P12: "Is [gender imbalance] a result of the recommendation system, or of people's individual biases? [...] If it is the former, then [women] should be recommended more, but if it's not then I don't think that is the job of the recommender system to fix, and it's more something that society should fix."* As previously described, participants also often identified financial gains as causes of unfairness, recognizing that the status quo may be more profitable for the platforms. In this case, broader societal tendencies, such as capitalism, were blamed for unfairness and viewed as inevitable: *P6: "It feels a little bit silly to say, but I think the capitalism behind it. [...] In the end, they all want to make money."*

Suggested improvements to achieve artist or user fairness were shared by 12 participants. Responses mainly concerned changing the recommendation algorithms or increasing transparency into, and control of, recommendations in the user interface. On transparency, P1 mentioned: *P1: "I think [it would be fair] to be transparent. Make me understand as best possible why I'm being recommended certain things, [and others] are being hidden from* *me."* More control was considered to increase fairness as well: *P11: "[Streaming platforms] want you to stick to the product, and I think if it were to really be fair, they would let that choice be upon you, and they would also design it in such a way."* A certain overall fairness preference could be set in a user's profile: *P7: "It would be good if users could indicate if they want to, like, support lesser known artists"*

Regarding disparate treatment for underrepresented artists in MRS, participants were divided (see also Figure 2). In Step 4, some further reflected on whether disparate treatment was a solution, sometimes seeing it as potentially viable: *P2: "It sounds at least somewhat fair that the [percentages of men and women in recommendations] are similar."* In generated playlists, fairness goals could be directly integrated: *P17: "If you put on a popular artists radio [...] and you put in lesser known artists, it would be really nice to get to know more artists that have the same kind of genre or something."* Two participants were unsure, one noting: *P2: "I think it's strange to say: "Oh, this group is underrepresented, so we should [...] give them more attention." Because that doesn't sound like the right reason somehow. But it's just to kind of make it fair - I don't know."*

In short, there was no consensus on who is responsible to improve fairness, with participants mentioning streaming platforms, the music industry, users themselves, and even society as a whole.

5 Discussion

Here, we discuss our insights into our participants' playlist selection process with and without fairness-related descriptions, how fair participants perceive current MRS, and and with whom they see the responsibility for counteracting unfairness.

Regarding the decision process, participants rarely changed their playlist choices, even though they indicated that certain descriptions could have influenced their choices if provided on first listen. This discrepancy may be explained with insights from decision-making theory, such as the commitment principle [37]. This principle states that people tend to stick with their initial choices even when presented with new or better information. We thus need to provide explanations of fairness *before* a decision takes place.

In line with prior work [6], we further observe that users consistently prioritized their musical preferences over fairness-related values, leading to discrepancies between expressed support for disparate treatment and their own choices. Extending previous findings [40], this most prominently impacts the popularity dimension; our participants' wish for familiarity counters their wish to support lesser-known artists. This trade-off is also visible in previous studies [33, 66]. Finally, participants had diverse values and showed varying support for fairness countermeasures.

Regarding users' perceptions of fairness-related issues, most users had considered artist fairness before, but fewer of them user fairness. For artist fairness, the focus was on impact of the music industry and renumeration, whereas participants saw little connection to user choices in the system. In line with artists [16, 24], users expressed support for reducing disparities. For user fairness, participants mainly experienced filter bubbles with popular or mainstream content. They also perceived a lack of options to discover new and diverse content. Still, participants generally did not relate the lack of diversity to misrepresentation; most felt privileged and did not want to 'complain' about their representation in the system. However, they did suspect that the system optimizes for financial profits, thereby sometimes providing 'worse' recommendations.

Regarding the responsibility for counteracting fairness-related issues, some users saw music streaming platforms as the cause *and* solution, via either curation or algorithmic processing. Previous work even proposed that algorithms are perceived to have similar roles as institutions [5]. Most participants did not see their own role in improving fairness in the system, even though some did mention society as a whole. The think-aloud protocols and interview results indicated that participants reflected more after seeing fairness information on the playlists, even though they did not change their choices. Another way to shift responsibility to the users while keeping the convenience of MRS is to let users provide preferences on fairness dimensions *before* receiving recommendations.

Limitations. First, the participant pool has limited representativeness, as common in qualitative research. Second, some methodological choices may have influenced our results. Our songs, though carefully selected, were not personalized and might not have fit users' preferences. Also, seeing the dimensions in Step 2 may have primed participants to consider them in Step 3. Still, participants also raised fairness concerns beyond the Step 2 ones. Also, user choices and perceptions may differ in real-life contexts. Since users often engage with platforms like Spotify passively and may overlook fairness information or playlist text, the practical impact could be limited. Lastly, social desirability bias [29] may have affected participants' dedication to the playlist selection (e.g., more depth of interaction) and replies to the ethics-related questions, potentially causing a mismatch between stated values and actual preferences.

As future work, we suggest extending our insights to lesser-known fairness dimensions and and examining how transparency and control affect real-life decisions with personalized content, in MRS and beyond. The ideal timing and framing of fairness descriptions should especially be studied. Further, how to model users' individual fairness-related values should be better defined. Finally, helping users discover new music on streaming platforms—especially in a fair and engaging way—remains an unsolved challenge.

6 Conclusion

In this study, we explored the role that fairness-focused playlist descriptions may play in helping music listeners make choices that correspond to their values. Eighteen participants took part in think-aloud music selection tasks and semi-structured interviews. Although participants indicated that certain fairness dimensions are important to them, this was not necessarily reflected in their playlist choices, and giving more (personalized) insight into these dimensions may help to address this discrepancy, and could be implemented besides algorithmic interventions to improve fairness. Therefore, we conclude that giving end users more insights into fairness aspects of the music they are playing is worthwhile to explore further. While it will be challenging to personalize such insights to fit users' goals and needs, it may help them find music faster, gain insight into their listening behavior, and increase their agency on aligning their choices with their values. Based on the users' expressed fairness-related values, such agency could lead to less biased consumption overall, benefiting both users and artists.

References

[1] Himan Abdollahpouri. 2020. *Popularity bias in recommendation: A multi-stakeholder perspective*. Ph. D. Dissertation. University of Colorado at Boulder.

[2] Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnodebski, and Luiz Pizzato. 2020. Multi-stakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction* 30, 1 (2020), 127–158. doi:10.1007/s11257-019-09256-1

[3] Himan Abdollahpouri and Robin Burke. 2022. *Multistakeholder Recommender Systems*. 647–677. doi:10.1007/978-1-0716-2197-4_17

[4] Darius Afchar, Alessandro Melchiorre, Markus Schedl, Romain Hennequin, Elena Epure, and Manuel Moussallam. 2022. Explainability in music recommender systems. *AI Magazine* 43, 2 (2022), 190–208. https://doi.org/10.1002/aaai.12056

[5] Virgílio Almeida, Ricardo Fabrino Mendonça, and Fernando Filgueiras. 2024. Thinking of Algorithms as Institutions. *Commun. ACM* 68, 1 (Dec. 2024), 20–23. doi:10.1145/3680411

[6] Gabrielle Alves, Dietmar Jannach, Rodrigo Ferrari De Souza, and Marcelo Garcia Manzato. 2024. User Perception of Fairness-Calibrated Recommendations. In *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP '24)*. 78–88. doi:10.1145/3627043.3659558

[7] Claus Atzenbeck, Eelco Herder, and Daniel Roßner. 2023. Breaking the routine: spatial hypertext concepts for active decision making in recommender systems. *New Review of Hypermedia and Multimedia* 29, 1 (2023), 1–35. https://doi.org/10.1080/13614568.2023.2170474

[8] Ricardo Baeza-Yates. 2018. Bias on the web. *Commun. ACM* 61, 6 (May 2018), 54–61. doi:10.1145/3209581

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa

[10] Pablo Castells, Neil Hurley, and Saúl Vargas. 2022. *Novelty and Diversity in Recommender Systems*. 603–646. doi:10.1007/978-1-0716-2197-4_16

[11] Òscar Celma. 2010. *Music Recommendation*. Chapter 3, 43–85. doi:10.1007/978-3-642-13287-2_3

[12] Li Chen, Marco de Gemmis, Alexander Felfernig, Pasquale Lops, Francesco Ricci, and Giovanni Semeraro. 2013. Human Decision Making and Recommender Systems. *ACM Trans. Interact. Intell. Syst.* 3, 3, Article 17 (Oct. 2013), 7 pages. doi:10.1145/2533670.2533675

[13] Li Chen, Wen Wu, and Liang He. 2013. How personality influences users' needs for recommendation diversity? In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, USA, 829–834. doi:10.1145/2468356.2468505

[14] Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogin, Alessandro Difonzo, and Dario Zanzonelli. 2024. Fairness in recommender systems: research landscape and future directions. *User Modeling and User-Adapted Interaction* 34, 1 (2024), 59–108. https://doi.org/10.1007/s11257-023-09364-z

[15] Karlijn Dinnissen and Christine Bauer. 2022. Fairness in Music Recommender Systems: a Stakeholder-centered Mini Review. *Frontiers in Big Data* 5, Article 913608 (2022), 9 pages. doi:10.3389/fdata.2022.913608

[16] Karlijn Dinnissen and Christine Bauer. 2023. Amplifying Artists' Voices: Item Provider Perspectives on Influence and Fairness of Music Streaming Platforms. In *Proceedings of the 31st Conference on User Modeling, Adaptation and Personalization (UMAP '23)*. ACM, New York, NY, USA, 12 pages. doi:10.1145/3565472.3592960

[17] Karlijn Dinnissen and Christine Bauer. 2023. How Control and Transparency for Users Could Improve Artist Fairness in Music Recommender Systems. In *Proceedings of the 24th International Society for Music Information Retrieval Conference (ISMIR '23)*. 482–491. doi:10.5281/zenodo.10265331

[18] Karlijn Dinnissen, Isabella Saccardi, Marloes Vredenborg, and Christine Bauer. 2023. Looking at the FAccTs: Exploring Music Industry Professionals' Perspectives on Music Streaming Services and Recommendations. In *Proceedings of the 2nd International Conference of the ACM Greek SIGCHI Chapter (CHIGREECE '23)*. ACM, New York, NY, USA, Article 25, 5 pages. doi:10.1145/3609987.3610011

[19] Michael D Ekstrand, Anubrata Das, Robin Burke, and Fernando Diaz. 2022. Fairness in Information Access Systems. *Foundations and Trends® in Information Retrieval* 16, 1–2 (2022), 177 pages. doi:10.1561/1500000079

[20] Michael D. Ekstrand, Mucun Tian, Ion Madrazo Azpiazu, Jennifer D. Ekstrand, Oghenemaro Anuyah, David McNeill, and Maria Soledad Pera. 2018. All The Cool Kids, How Do They Fit In?: Popularity and Demographic Biases in Recommender Evaluation and Effectiveness. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (FAT* '18, Vol. 81)*, Sorelle A. Friedler and Christo Wilson (Eds.). 172–186. https://proceedings.mlr.press/v81/ekstrand18b.html

[21] Somayeh Fatahi, Mina Mousavifar, and Julita Vassileva. 2023. Investigating the effectiveness of persuasive justification messages in fair music recommender systems for users with different personality traits. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization (UMAP '23)*. ACM, New York, NY, USA, 66–77. doi:10.1145/3565472.3592958

[22] Andres Ferraro, Michael D. Ekstrand, and Christine Bauer. 2024. It's Not You, It's Me: The Impact of Choice Models and Ranking Strategies on Gender Imbalance in Music Recommendation. In *Proceedings of the 18th ACM Conference on Recommender Systems* (Bari, Italy) *(RecSys '24)*. ACM, New York, NY, USA, 884–889. doi:10.1145/3640457.3688163

[23] Andrés Ferraro, Xavier Serra, and Christine Bauer. 2021. Break the Loop: Gender Imbalance in Music Recommenders. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval (CHIIR '21)*. 249–254. doi:10.1145/3406522.3446033

[24] Andrés Ferraro, Xavier Serra, and Christine Bauer. 2021. What Is Fair? Exploring the Artists' Perspective on the Fairness of Music Streaming Platforms. In *Human-Computer Interaction – INTERACT 2021: 18th IFIP TC 13 International Conference (INTERACT '21, Vol. 12933)*, C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, and K. Inkpen (Eds.). 562–584. doi:10.1007/978-3-030-85616-8_33

[25] Bruce Ferwerda, Eveline Ingesson, Michaela Berndl, and Markus Schedl. 2023. I Don't Care How Popular You Are! Investigating Popularity Bias in Music Recommendations from a User's Perspective. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval (CHIIR '23)*. 357–361. doi:10.1145/3576840.3578287

[26] World Economic Forum. 2023. How the World Consumes Music. https://www.weforum.org/stories/2023/02/world-consume-music-infographic/

[27] Simon Frith, Richard Leppert, and Susan McClary. 2004. Towards an aesthetic of popular music. *Popular music: Critical concepts in media and cultural studies* 4 (2004), 32–47.

[28] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. 2014. How should I explain? A comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies* 72, 4 (2014), 367–382. https://doi.org/10.1016/j.ijhcs.2013.12.007

[29] Pamela Grimm. 2010. Social desirability bias. *Wiley international encyclopedia of marketing* (2010). https://doi.org/10.1002/9781444316568.wiem02057

[30] Sophia Hadash, Yu Liang, and Martijn C. Willemsen. 2019. How playlist evaluation compares to track evaluations in music recommender systems. In *IntRS 2019 Interfaces and Human Decision Making for Recommender Systems 2019 (CEUR Workshop Proceedings)*, P. Brusilovsky, M. de Gemmis, A. Felfernig, P. Lops, J. O'Donovan, G. Semeraro, and M.C. Willemsen (Eds.). CEUR-WS.org, 1–9. https://research.tue.nl/files/146414140/paper1.pdf 6th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, IntRS 2019 ; Conference date: 19-09-2019.

[31] F. Maxwell Harper, Funing Xu, Harmanpreet Kaur, Kyle Condiff, Shuo Chang, and Loren Terveen. 2015. Putting Users in Control of their Recommendations. In *Proceedings of the 9th ACM Conference on Recommender Systems* (Vienna, Austria) *(RecSys '15)*. ACM, New York, NY, USA, 3–10. doi:10.1145/2792838.2800179

[32] IFPI. 2024. Global Music Report 2024. (2024). https://globalmusicreport.ifpi.org/

[33] Dietmar Jannach, Lukas Lerche, and Michael Jugovac. 2015. Item Familiarity as a Possible Confounding Factor in User-Centric Recommender Systems Evaluation. *i-com* 14, 1 (2015), 29–39. doi:doi:10.1515/icom-2015-0018

[34] Yucheng Jin, Bruno De Lemos Ribeiro Pinto Cardoso, and Katrien Verbert. 2017. How do different levels of user control affect cognitive load and acceptance of recommendations?. In *IntRS@ RecSys*. 35–42. http://ceur-ws.org/Vol-1884/paper7.pdf

[35] Yucheng Jin, Nava Tintarev, and Katrien Verbert. 2018. Effects of Individual Traits on Diversity-Aware Music Recommender User Interfaces. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization (UMAP '18)*. 291–299. doi:10.1145/3209219.3209225

[36] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature machine intelligence* 1, 9 (2019), 389–399. https://doi.org/10.1038/s42256-019-0088-2

[37] Daniel Kahneman, Olivier Sibony, and Cass R Sunstein. 2021. *Noise: a flaw in human judgment*. Hachette UK.

[38] Shah Noor Khan, Eelco Herder, and Diba Kaya. 2025. Effects of Representation Nudges on the Perception of Playlist Recommendations. In *Recommender Systems for Sustainability and Social Good*, Ludovico Boratto, Allegra De Filippo, Elisabeth Lex, and Francesco Ricci (Eds.). Springer Nature Switzerland, Cham, 151–160.

[39] Jaekyeong Kim, Ilyoung Choi, and Qinglong Li. 2021. Customer satisfaction of recommender system: Examining accuracy and diversity in several types of recommendation approaches. *Sustainability* 13, 11 (2021), 6165. https://doi.org/10.3390/su13116165

[40] Anastasiia Klimashevskaia, Mehdi Elahi, Dietmar Jannach, Christoph Trattner, and Lars Skjærven. 2022. Mitigating Popularity Bias in Recommendation: Potential and Limits of Calibration Approaches. In *Advances in Bias and Fairness in Information Retrieval*, Ludovico Boratto, Stefano Faralli, Mirko Marras, and Giovanni Stilo (Eds.). 82–90. https://doi.org/10.1007/978-3-031-09316-6_8

[41] Dominik Kowald, Peter Müllner, Eva Zangerle, Christine Bauer, Markus Schedl, and Elisabeth Lex. 2021. Support the underground: Characteristics of beyond-mainstream music listeners. *EPJ Data Science* 10, 1, Article 14 (2021), 28 pages. doi:10.1140/epjds/s13688-021-00268-9

[42] Dominik Kowald, Markus Schedl, and Elisabeth Lex. 2020. The unfairness of popularity bias in music recommendation: A reproducibility study. In *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020,*

Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II 42. Springer, 35–42. https://doi.org/10.1007/978-3-030-45442-5_5

[43] Paddy Leerssen. 2020. The soap box as a black box: regulating transparency in social media recommender systems. *European Journal of Law and Technology* 11, 2 (2020). http://dx.doi.org/10.2139/ssrn.3544009

[44] Zhongzhou Liu, Yuan Fang, and Min Wu. 2023. Mitigating Popularity Bias for Users and Items with Fairness-centric Adaptive Recommendation. *ACM Trans. Inf. Syst.* 41, 3, Article 55 (Feb. 2023), 27 pages. doi:10.1145/3564286

[45] Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. 2020. Feedback Loop and Bias Amplification in Recommender Systems. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (Virtual Event, Ireland) *(CIKM '20)*. ACM, New York, NY, USA, 2145–2148. doi:10.1145/3340531.3412152

[46] Leland McInnes, John Healy, and James Melville. 2020. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. https://arxiv.org/abs/1802.03426

[47] Alessandro B. Melchiorre, Navid Rekabsaz, Emilia Parada-Cabaleiro, Stefan Brandl, Oleg Lesota, and Markus Schedl. 2021. Investigating gender fairness of recommendation algorithms in the music domain. *Information Processing & Management* 58, 5, Article 102666 (2021), 27 pages. doi:10.1016/j.ipm.2021.102666

[48] Silvia Milano, Mariarosaria Taddeo, and Luciano Floridi. 2021. Ethical aspects of multi-stakeholder recommendation systems. *The information society* 37, 1 (2021), 35–45. https://doi.org/10.1080/01972243.2020.1832636

[49] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd International Conference on World Wide Web* (Seoul, Korea) *(WWW '14)*. ACM, New York, NY, USA, 677–686. doi:10.1145/2566486.2568012

[50] Ingrid Nunes and Dietmar Jannach. 2017. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction* 27 (2017), 393–444. https://doi.org/10.1007/s11257-017-9195-0

[51] Ricardo S Oliveira, Caio Nóbrega, Leandro Balby Marinho, and Nazareno Andrade. 2017. A Multiobjective Music Recommendation Approach for Aspect-Based Diversification. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR '17)*. 414–420. doi:10.5281/zenodo.1416999

[52] Yesid Ospitia-Medina, Sandra Baldassarri, Cecilia Sanz, and José Ramón Beltrán. 2022. Music Recommender Systems: A Review Centered on Biases. *Advances in Speech and Music Technology: Computational Aspects and Applications* (2022), 71–90. https://doi.org/10.1007/978-3-031-18444-4_4

[53] Kyle Robinson, Dan Brown, and Markus Schedl. 2020. User Insights on Diversity in Music Recommendation Lists. In *Proceedings of the 21st International Society for Music Information Retrieval Conference* (October 11–16) *(ISMIR '20)*. 446–453. doi:10.5281/zenodo.4245464

[54] Markus Schedl and David Hauger. 2015. Tailoring Music Recommendations to Users by Considering Diversity, Mainstreaminess, and Novelty. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '15)*. 947–950. doi:10.1145/2766462.2767763

[55] Markus Schedl, Peter Knees, Brian McFee, and Dmitry Bogdanov. 2022. Music Recommendation Systems: Techniques, Use Cases, and Challenges. In *Recommender Systems Handbook* (3rd ed.), Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). 927–971. doi:10.1007/978-1-0716-2197-4_24

[56] Markus Schedl, Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo, and Mehdi Elahi. 2018. Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval* 7 (2018), 95–116. doi:10.1007/s13735-018-0154-2

[57] Dougal Shakespeare, Lorenzo Porcaro, Emilia Gómez, and Carlos Castillo. 2020. Exploring Artist Gender Bias in Music Recommendation. In *Proceedings of the Workshops on Recommendation in Complex Scenarios and the Impact of Recommender Systems co-located with 14th ACM Conference on Recommender Systems (RecSys 2020)* (2020) *(ComplexRec-ImpactRS 2020, Vol. 2697)*. CEUR Workshop Proceedings, Article 1, 9 pages. https://arxiv.org/pdf/2009.01715

[58] Alexander Simons, Markus Weinmann, Matthias Tietz, and Jan vom Brocke. 2017. Which reward should I choose? Preliminary evidence for the middle-option bias in reward-based crowdfunding. *Proceedings of the 50th Hawaii International Conference on System Sciences* (2017). http://hdl.handle.net/10125/41687

[59] Rashmi Sinha and Kirsten Swearingen. 2002. The role of transparency in recommender systems. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI EA '02)*. ACM, New York, NY, USA, 830–831. doi:10.1145/506443.506619

[60] Paul Slovic, Baruch Fischhoff, and Sarah Lichtenstein. 1977. Behavioral decision theory. *Annual review of psychology* (1977). https://doi.org/10.1146/annurev.ps.28.020177.000245

[61] Jessie J. Smith, Aishwarya Satwani, Robin Burke, and Casey Fiesler. 2024. Recommend Me? Designing Fairness Metrics with Providers. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. 2389–2399. doi:10.1145/3630106.3659044

[62] Nasim Sonboli, Jessie J. Smith, Florencia Cabral Berenfus, Robin Burke, and Casey Fiesler. 2021. Fairness and Transparency in Recommendation: The Users' Perspective. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21)*. 274–279. doi:10.1145/3450613.3456835

[63] Nava Tintarev and Judith Masthoff. 2015. Explaining recommendations: Design and evaluation. In *Recommender Systems Handbook*. 353–382. https://link.springer.com/content/pdf/10.1007/978-1-4899-7637-6.pdf

[64] Chun-Hua Tsai and Peter Brusilovsky. 2021. The effects of controllability and explainability in a social recommender system. *User Modeling and User-Adapted Interaction* 31 (2021), 591–627. https://doi.org/10.1007/s11257-020-09281-5

[65] Matteo Turilli and Luciano Floridi. 2009. The ethics of information transparency. *Ethics and Information Technology* 11 (2009), 105–112. https://doi.org/10.1007/s10676-009-9187-9

[66] Robin Ungruh, Karlijn Dinnissen, Anja Volk, Maria Soledad Pera, and Hanna Hauptmann. 2024. Putting Popularity Bias Mitigation to the Test: A User-Centric Evaluation in Music Recommenders. In *Proceedings of the Eighteenth ACM Conference on Recommender Systems (RecSys '24)*. 10 pages. https://doi.org/10.1145/3640457.3688102

[67] United Nations. 2024. 2024 Revision of World Population Prospects. (2024).

[68] Wenjie Wang, Fuli Feng, Xiangnan He, Xiang Wang, and Tat-Seng Chua. 2021. Deconfounded Recommendation for Alleviating Bias Amplification. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (Virtual Event, Singapore) *(KDD '21)*. ACM, New York, NY, USA, 1717–1725. doi:10.1145/3447548.3467249

[69] Mireia Yurrita, Tim Draws, Agathe Balayn, Dave Murray-Rust, Nava Tintarev, and Alessandro Bozzon. 2023. Disentangling Fairness Perceptions in Algorithmic Decision-Making: the Effects of Explanations, Human Oversight, and Contestability. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. ACM, New York, NY, USA, Article 134, 21 pages. doi:10.1145/3544548.3581161

[70] Eva Zangerle and Christine Bauer. 2022. Evaluating Recommender Systems: Survey and Framework. *ACM Comput. Surv.* 55, 8, Article 170 (Dec. 2022), 38 pages. doi:10.1145/3556536

[71] Guoshuai Zhao, Hao Fu, Ruihua Song, Tetsuya Sakai, Zhongxia Chen, Xing Xie, and Xueming Qian. 2019. Personalized Reason Generation for Explainable Song Recommendation. *ACM Trans. Intell. Syst. Technol.* 10, 4, Article 41 (July 2019), 21 pages. doi:10.1145/3337967