User Perceptions and Responses to Minority Fairness in Music Recommendations and Item Labels

Shah Noor Khan
Utrecht University
Utrecht, The Netherlands

Eelco Herder
Utrecht University
Utrecht, The Netherlands

Libio Goncalves Braz
Utrecht University
Utrecht, The Netherlands

Karlijn Dinnissen
Utrecht University
Utrecht, The Netherlands

Judith Masthoff
Utrecht University
Utrecht, The Netherlands

Abstract

Music recommender systems play a central role in how users explore and consume music, yet they are often associated with concerns related to bias and unequal representation. While existing research has largely focused on improving fairness at the algorithmic level, less is known about how users interpret fairness and whether such information affects their choices. This paper examines how users engage with fairness-related information presented alongside songs, and whether it influences their decision-making. To investigate this, we conducted an online mixed method user study with 28 participants, combining a ranking task with follow-up reflections. Participants were asked to rank songs based on their preferences while being shown labels indicating associations with categories such as gender, age, location, and orientation. The findings indicate that these labels had minimal influence on ranking behaviour. At the same time, participants showed low agreement in their rankings, reflecting the highly subjective nature of music preferences. Most participants reported relying on familiarity, personal taste, and emotional responses rather than the provided labels. These results suggest that providing fairness-related information alone is unlikely to change user behaviour. Instead, such information needs to be both meaningful and relevant to users in order to be considered during decision-making. This work highlights the importance of accounting for user perception and taste when designing fairness-oriented recommender systems.

CCS Concepts

• **Human-centered computing → User studies**.

Keywords

Fairness, User Study, Music Recommender system

ACM Reference Format:
Shah Noor Khan, Eelco Herder, Libio Goncalves Braz, Karlijn Dinnissen, and Judith Masthoff. 2026. User Perceptions and Responses to Minority Fairness in Music Recommendations and Item Labels. In *37th ACM Conference on Hypertext (HT '26), September 14–18, 2026, London, United Kingdom.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3800935.3830871

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/10.1145/3800935.3830871

1 Introduction

Music streaming platforms have made music one of the most accessible forms of entertainment for a wide audience [22]. Within these platforms, Music Recommender Systems (MRS) guide users by suggesting content, even when user preferences are not yet fully defined [34]. These systems shape user experience by influencing music discovery and the visibility of artists and genres [42, 44].

Given this influence, fairness has become a key concern in recommender systems. Many approaches prioritize prediction accuracy, unintentionally amplifying biases such as popularity bias and demographic imbalances [1]. Consequently, recommendation algorithms may reinforce inequalities affecting both users and artists, and contribute to phenomena such as filter bubbles [39, 41].

These challenges are particularly relevant in the music domain, where mainstream content, often Western and English language, tends to dominate recommendation outputs [7]. In contrast, ethnic or niche music represents minority user groups and cultural diversity [19]. Users with less common preferences may therefore receive recommendations that do not align well with their tastes [17, 30]. Since such content typically generates fewer interactions, recommender systems have limited data to accurately model these preferences [8, 51]. This can lead to reduced satisfaction, lower engagement, and potential disengagement from the platform [38].

User behaviour further complicates these challenges. The Spiral of Silence theory suggests that individuals may be less likely to express their preferences when they perceive themselves to be in the minority [33, 40]. In online environments, this can underrepresent minority preferences, reinforcing recommender biases and limiting exposure to niche or diverse content

One approach to improving fairness is to make it more transparent to users. Providing fairness-related information, such as diversity or representation indicators, may support informed decision-making and promote more balanced consumption [2]. This aligns with nudging research, which shows that subtle interventions can influence user choices, though effects vary across individuals [18, 29]. Despite increasing attention to fairness-aware recommender systems,the effects of explicitly communicated fairness on user perception and interaction remain unclear.

This paper investigates how users perceive and respond to fairness-related information in music recommender systems, with a particular focus on how such information may influence user choices and engagement. In an earlier study[27], we identified the *commitment principle* as a key factor influencing user choice and engagement. Providing fairness information about songs after users had ranked them did not entice users to change their rankings. The present study was designed to mitigate this effect by directly showing fairness labels instead of afterward. In an online study we captured in-depth quantitative and qualitative data on participants' responses to these labels, and compared participants' perceptions and actions with our previous study.

2 Related Work

2.1 Music Recommender Systems

Music recommendation differs from other domains due to the subjective and contextual nature of music consumption, as well as users' tendency to revisit familiar tracks [45]. Effective MRS must balance familiarity with discovery while aligning with users' preferences [4]. However, systems that primarily optimize prediction accuracy through historical interaction data [43] may overlook broader user experiences, including how recommendations shape perceptions and listening behaviour [24]. Consequently, they may fail to capture the evolving and diverse nature of musical preferences.

2.2 Popularity Bias and Underrepresented Users

A key challenge in MRS is popularity bias, where frequently consumed tracks are prioritized over long-tail and niche content [1, 8]. Limited interaction data for less popular items reduces their effective modeling, leading to poorer recommendations for users with non-mainstream preferences [30]. This reinforces unequal experiences by favoring mainstream audiences and marginalizing diverse tastes [28]. Furthermore, popularity bias can amplify cultural imbalances, as Western and English-language music often dominates recommendations, limiting exposure to diverse musical expressions [45]. These challenges highlight the need to consider user perceptions of fairness alongside system-level optimization.

2.3 Fairness and User-Centred Perspectives

Fairness in recommender systems has traditionally focused on algorithmic approaches that balance exposure across items or user groups [5]. However, system-level metrics may not reflect users' fairness perceptions [17]. Recent work emphasizes user-centred fairness, where users participate in defining and interacting with fairness criteria [15]. Allowing users to control fairness dimensions, such as diversity or popularity, can improve satisfaction and trust [48]. However, how users interpret explicitly presented fairness information remains underexplored, particularly in music recommender systems.

Fairness in recommender systems is a multi-stakeholder challenge involving trade-offs among users, providers, and items [6]. Prior research shows that fairness has no universal definition, as preferences for popularity, diversity, and representation vary across users and contexts [3, 5]. While algorithmic approaches aim to reduce bias and improve diversity [21, 25], they often focus on system-level objectives and overlook user perceptions. Since user responses are shaped by individual goals, decision-making styles, and behaviours [9, 11], understanding fairness requires incorporating the user perspective.

2.4 Transparency, User Control, and Nudging

Transparency is essential for building user trust and acceptance of recommender systems [50]. Clear explanations and contextual cues can improve understanding and support informed decisions [23]. However, transparency alone may be insufficient; users also need agency. Providing control over recommendation settings, such as diversity or fairness constraints, can enhance autonomy and reduce dissatisfaction [31]. Fairness information presented alongside recommendations may act as a nudge toward more balanced listening behaviours, although individual responses and the mechanisms behind these effects remain unclear [18].

2.5 Cultural and Minority Representation in Music

Music preferences are strongly influenced by cultural background and exposure [3]. Nevertheless, most recommendation systems depend on Western-centric metadata and representations, which can impede the accurate modelling of ethnic or niche music [10]. Furthermore, smaller user bases and limited interaction data pose additional challenges for standard recommendation techniques in these contexts [12], often resulting in the marginalization of minority music and listeners on mainstream platforms.

2.6 SOS in Online Recommender Systems

The Spiral of Silence (SOS) theory states that individuals may suppress opinions perceived to be in the minority due to fear of social isolation [40]. Online, this leads to selective self-expression and conformity to majority views [33, 37]. On music platforms, these behaviours shape users' expressed preferences. For recommender systems, suppressed minority preferences can bias user interaction data, causing algorithms to favor mainstream content, reinforce popularity bias, and reduce exposure to niche items [26, 35, 49]. This feedback loop further marginalizes minority preferences, motivating fairer recommender systems.

Overall, prior work highlights significant challenges related to bias, fairness, and user behaviour in music recommender systems. While existing approaches have primarily focused on algorithmic solutions, less attention has been given to how users perceive and engage with fairness-related information when it is explicitly presented to them. In particular, it remains unclear whether such information is understood, considered relevant, or capable of influencing user decision-making in practice. Addressing this gap is essential for designing fairness-aware recommender systems that are not only technically effective but also meaningful from a user perspective.

3 Methodology

We conducted an online user study to examine whether providing category-based fairness information influences users' decision-making in music ranking tasks. Participants were asked to rank a set of songs according to their personal preferences within an interface where most songs were accompanied by labels indicating their association with specific categories related to fairness, including gender, age, location, and orientation. In addition to the ranking task, the study investigated how participants perceived and interpreted these category labels, as well as the extent to which they considered them relevant to their decisions. To capture this, participants were asked to reflect on the labels after completing the task, with a particular focus on their clarity, usefulness, and influence on the overall user experience. The study was assessed by the university's ethics and privacy board and determined to be low risk, with no further ethical review required.

We refer to the current study as Study 2. Our previous study[27] (which we refer to as Study 1) had a similar setup and used the same song collection and the same fairness categories – which will be introduced in the next subsection. In Study 1[27], participants were first presented with the songs without fairness labels. After having ranked the songs, they were asked to identify which songs belonged to which fairness label. Finally, they were allowed to rerank the songs and to freely comment on the songs and their assigned categories. Even though the participants of Study 1 expressed many opinions, almost no reranking took place, which we attributed to the *commitment principle*, where people adhere to their first choice.

3.1 Selection of Fairness Categories

Prior work has identified several dimensions along which bias can emerge in music recommender systems. One prominent issue is *popularity bias*, where widely consumed mainstream tracks are recommended more frequently, often at the expense of niche or less popular music [1].

Another important dimension is *gender representation*. Studies have shown that female and non-binary artists are often underrepresented in recommendations and may receive less accurate exposure compared to male artists [20, 36].

Nationality also shapes recommendation outcomes. Artist visibility varies across regions, with local and region-specific music often receiving less exposure due to limited data and the Western-centric design of many recommender systems [3, 10, 12]. Consequently, recommender systems may underrepresent local music relative to users' listening behaviour, reinforcing the dominance of English-language, particularly U.S., music [32].

Age also influences music preferences, as music encountered during late adolescence and early adulthood often has a lasting impact on listening behaviour [13].

Finally, *LGBTQ-themed* music and playlists represent another relevant dimension.Such content reflects and shapes LGBTQ identities and cultural expression, emphasizing the need for inclusive representation in recommender systems [14].

3.2 Study Setup

The study was conducted online. The procedure was as follows:

Step 1: Pre-study survey. Following informed consent, participants completed a demographic questionnaire that included age, nationality, and gender before starting the main task. A total of 28 participants completed the study, with a mean age of 25.8 years (range: 19–49). Although recruitment occurred in the Netherlands, the sample was internationally diverse. Twenty participants were European, with 14 from the Netherlands. The gender distribution included 20 male participants, 6 female participants, and 2 participants who selected "prefer not to say."

Step 2: Category Description. Prior to the ranking task, participants were provided with brief descriptions of the categories (Figure 1 to clarify their meaning and ensure a shared understanding of the definitions).

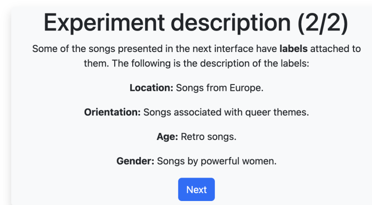

Figure 1: Category description : Descriptions of the categories were presented.

Step 3: Ranking task. Participants completed a ranking task with 20 songs displayed in a grid-based interface (Figure 2). To avoid middle-option bias [47], song order was randomized for each participant. Each song was presented as a card showing the title and artist, and participants had the option to listen to a short preview [16, 18]. The songs were chosen to represent different fairness-related categories (see Section 3.1 and 3.3). To help participants recognize these categories, the interface showed which categories each song belonged to and highlighted them using different background colours. Category labels were displayed continuously during ranking. Participants were asked to arrange the songs according to their personal preferences by dragging and dropping them into their preferred order. Once they were satisfied with their ranking, they submitted their responses and moved on to the next part of the study.

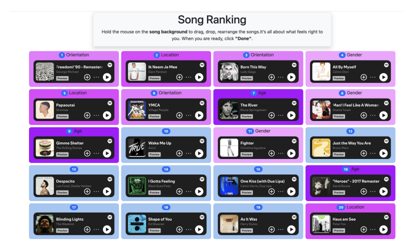

Figure 2: Ranking Task: Songs were labeled from 1 to 20, with additional information provided, and participants ranked them from most (1) to least (20) preferred by dragging and dropping.

Step 4: Reflection. Finally, participants were asked to answer two open-ended questions reflecting on their perception of the labels and the extent to which these labels influenced their ranking decisions . This step was included to gain deeper insights into the reasoning behind participants' choices. To ensure clarity, descriptions of the fairness categories were made available during this stage as well, allowing participants to consult them while responding to the questions.

3.3 Song Selection

A total of 20 songs were selected to provide a manageable yet varied set of options while avoiding dominance by any single song. Three songs represented each fairness dimension, and eight highly popular mainstream songs served as a baseline, reflecting the prominence of mainstream content on music streaming platforms.

Songs were selected from publicly available rankings and curated lists relevant to each dimension, prioritizing Dutch sources to match the participant population and supplementing them with international sources when needed. Familiarity was approximated using Spotify popularity scores[1], with selected tracks scoring above 50 for fairness categories and above 80 for mainstream songs.

Gender dimension. Here, songs were selected to represent underrepresented gender groups (see Section 3.1). To ensure clear attribution, only solo artists were included. The selection was based on Billboard's ranking of top women artists of the 21st century.[2]

Age dimension. This category included classic rock songs released between 1970 and 1980. Despite participants not belonging to the target age group, these tracks were selected for their enduring popularity and familiarity. Selections were based on the 2024 Classic Rock Top 100 by Dutch radio station Arrow Classic Rock.[3]

LGBTQ+ dimension. Songs in this category were selected based on their recognition within LGBTQ communities through their lyrical themes or the identity and influence of the performing artists.Sources included Billboard's lists of LGBTQ anthems and a Dutch "Gay Top 100" chart list[4].

Nationality dimension. To capture linguistic and regional diversity, songs with non-English lyrics were selected from countries geographically close to the Netherlands, including Belgium (French), Germany (German), and the Netherlands (Dutch). Tracks were sourced from Top40.nl to ensure sufficient familiarity[5].

Mainstream baseline. The remaining songs represented popular mainstream music outside the defined fairness categories. These were primarily English-language tracks by U.S., UK, or Canadian artists, released after 2009 and selected based on strong chart performance, including extended runs at number one.

4 Results

We first present the quantitative results, followed by the qualitative findings in this section.

4.1 Quantitative Results

This section reports the average rankings of songs across the different categories, reflecting participants' preferences. In addition, we compute Kendall's Tau "τ" coefficients to assess the similarity between participants' rankings.

4.1.1 Average Song Ranking. To investigate category preferences, we calculated the average song ranking for each category by aggregating them within each category. Lower averages indicated higher preference (where 1.0 is the best possible ranking). Mainstream songs consistently received the lowest average rankings, indicating stronger preference. This aligns with Study 1 [27], which used the same songs with a different participant group. In that study, participants ranked songs first without labels and then with labels.

To check if the differences between studies were meaningful (Study 1, Round 1, where participants evaluated the same songs without category labels, and Study 2, where different participant groups evaluated the same songs including labels), we ran both independent t-tests and Mann–Whitney U tests[6] for all categories. The results in Table 1 showed no statistically significant differences between the two studies (all $p > .05$), which means participants' ranking behaviour was consistent. There were some small changes in mean rankings, like slightly lower rankings for gender and location-based songs in Study 2, but these were not statistically significant. Effect sizes were small to moderate, so the practical impact was limited. Overall, these results suggest that by showing fairness-related information earlier we did not observe statistically significant differences on how participants ranked the songs.

Table 1: Average song rankings per category across the two studies. Lower values indicate higher preference. No statistically significant differences were observed (all $p > .05$).

Category	Study 1	Study 2	t-test (p)	Mann–Whitney (p)
Mainstream	9.34	8.52	.167	.131
Gender	11.13	12.30	.115	.151
Location	11.50	12.82	.154	.140
Age	12.05	11.40	.618	.385
Orientation	10.42	10.76	.679	.566

To capture the ordinal nature of the ranking data, we visualised the distribution of category rankings using violin plots (Figure 3). The distributions are highly similar across studies, indicating consistent ranking behaviour. Lower values indicate higher preference, and violin width reflects the density of observations at each rank.

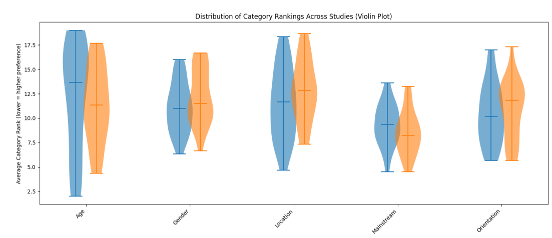

Figure 3: Violin plots of category rankings in Study 1 (blue) and Study 2 (orange).

4.1.2 Kendall's Tau "τ" Correlation. To determine whether consistent preference patterns emerged among participants, Kendall's Tau "τ" was computed. This statistic quantifies the degree of agreement between rankings, enabling assessment of the similarity in how participants ranked the songs.

Table 2 shows low agreement among participants, with mean and median correlations close to zero, indicating diverse musical preferences. The highest correlation reflects only moderate agreement, while the lowest represents opposing ranking patterns. Overall,

[1]Retrieved from https://developer.spotify.com/documentation/web-api/reference/get-track
[2]Billboard Top Women Artists of the 21st Century.
[3]Arrow Classic Rock 500 (2024).
[4]Sources: Billboard LGBTQ+ Anthems and Hitzound Homo Top 100.
[5]Sources: Top40 French-language hits, Top40 German-language hits, and Top40 Dutch-language hits.

[6]We avoid relying exclusively on normality tests, as they can be underpowered in small samples. Consistency between both tests confirms that our findings are not driven by distributional artifacts or outliers.

these findings demonstrate substantial variability and subjectivity in song ranking behaviour.

Table 2: Overall Kendall's Tau "τ" Statistics

Statistic	Avg. τ	Min. τ	Max. τ	Median τ
Value	0.141	-0.389	0.526	0.136

4.2 Qualitative Analysis

We conducted a thematic analysis of participants' open-ended responses to understand how they perceived the category labels and whether these influenced their ranking decisions. Five main themes emerged (see Table 4), providing insights into participants' engagement with the labels and their impact on the ranking task.

As shown in Table 3, the most frequently mentioned themes relate to the limited influence of labels and participants' lack of attention to them. In contrast, preference-based decision making was also commonly reported, indicating that participants primarily relied on personal taste rather than external labels during ranking.

Table 3: Frequency of themes across participant responses (N = 28)

Theme	Count
Label clarity and understanding	12
Perceived relevance of labels	14
Attention to labels	18
Influence on ranking	20
Preference-based decision making	16

Table 4: Coding framework for the qualitative data

Theme	Definition
Label clarity and understanding	How clearly participants understood and interpreted the labels.
Perceived relevance of labels	Whether participants felt the labels meaningfully described the songs.
Attention to labels	The extent to which participants considered labels during the ranking task.
Influence on ranking	Whether and how labels affected participants' rankings.
Preference-based decision making	Whether participants relied on personal taste, familiarity, or feelings rather than labels.

4.2.1 Label Clarity and Understanding. Many participants described the category labels as vague, broad, or difficult to interpret. Some expressed uncertainty about their meaning or application. For example, one participant noted that "the category labels are very broad and could refer to many different things" (P2), while another stated, "I didn't quite understand why or for what purpose those tags were added to the song" (P11).

Some participants needed to revisit the descriptions several times to understand them, with one explaining, "I had to go back and read all of the descriptions again" (P9). These responses suggest that the labels were not always intuitive, and a lack of clarity may have hindered engagement.

4.2.2 Perceived Relevance of Labels. Participants also questioned whether the labels accurately represented the songs. A few found the labels suitable; "the labels fit with the songs" (P4), but others saw little connection between labels and music. For instance, one participant stated, "I honestly didn't notice any correlation between the songs and the labels they were assigned" (P19).

Several participants felt the labels oversimplified the complexity of music or were too generic, as in the comment, "the message of a whole song cannot be captured in one category label" (P20). Even when understood, labels were not always seen as meaningful or representative, reducing their usefulness.

4.2.3 Attention to Labels. A key finding was that many participants paid little attention to the labels during the ranking process. Several forgot about the labels soon after starting the task or did not engage with them at all. One participant said, "I forgot what the description of the labels was as soon as I started ranking the songs" (P3). Others prioritized the music itself, with one stating, "I only care about the music itself and the lyrics" (P20). Overall, labels were not considered central or essential to the ranking activity.

4.2.4 Influence of Labels on Ranking. Reflecting this limited attention, most participants reported that labels did not influence their ranking decisions. Many explicitly stated that labels had little or no effect, such as "they did not influence my ranking" (P22). Some participants used labels to group songs or noted a subtle or subconscious influence, as in "they probably influenced me, but not consciously" (P24). A few observed patterns, such as placing songs with a particular label higher in their ranking. "I noticed I was putting at the top mainly songs from the label 'Location'" (P10). While explicit influence was limited, labels may have shaped perception in more subtle ways.

4.2.5 Preference-Based Decision Making. Participants consistently reported using personal criteria to rank songs. Familiarity with the music, emotional response, personal memories, and artist or genre preferences were common factors. "I ordered them according to my personal experiences." (P18) For example, participants ranked songs based on whether they knew them or how much they enjoyed them. "I ranked them based on my familiarity with the song." (P8) This indicates that participants relied primarily on internal, experience-based factors rather than external metadata such as labels.

4.2.6 Pattern Formation: High vs. Low average Kendall's Tau "τ" Values. To examine label engagement, we compared the five participants with the highest average Kendall's Tau (τ) values (P5, P28, P13, P8, P10) and the five with the lowest (P3, P6, P26, P21, P18).

Participants with higher τ values showed more structured engagement, sometimes identifying patterns or grouping songs. For example, one participant noted, "while ranking I noticed I was putting at the top mainly songs from the label 'Location'" (P10), while another acknowledged that labels had "probably" influenced them, though "not consciously" (P24). However, decisions were still primarily guided by familiarity and personal preference (P5, P8), suggesting that higher τ values reflected consistent decision criteria rather than direct reliance on labels.

Participants with lower τ values more often reported confusion or disengagement, with some stating they "forgot what the description of the labels was" (P3), "don't care about labels at all" (P6), or "didn't look much at the labels" (P21). Others questioned their usefulness or relevance (P18, P26). These findings suggest that unclear labels may reduce engagement and increase ranking variability.

Overall, higher τ values were associated with greater engagement with the task structure, whereas lower values reflected greater label disengagement. Across both groups, personal preference, familiarity, and emotional response remained the main drivers of ranking decisions. This aligns with Study 1 [27], where agreement with fairness labels was associated with higher τ values and more consistent choices. Although causality cannot be established, the findings suggest that agreement with label assignments may relate to more consistent decision-making.

5 Discussion

This study aimed to investigate how users perceive category-based labels in music recommendations and whether these labels influence their choices. Previous research indicates that providing additional context or fairness information can guide users [2, 18]. However, our results show that these labels had little impact, with participants relying primarily on their own preferences.

Further,quantitative analysis revealed low Kendall's Tau (τ) values, indicating limited agreement in song rankings and reflecting diverse personal preferences. This aligns with prior research showing that users often favor familiar or popular songs [8, 30], while genre preferences do not necessarily imply agreement on specific songs [46]. Such variability highlights the importance of personalization in music recommender systems [44]. The variation may also stem from the study design, as participants ranked songs based on personal preferences without predefined criteria, with differences in song familiarity further contributing to diverse rankings.

Qualitative findings further explain this variation. Many participants found the labels vague, broad, or difficult to interpret, leading them to disregard or forget them during the ranking task and focus primarily on the music itself. Thus, category cues alone are insufficient without user engagement.

Most participants reported that labels did not influence their rankings, instead relying on familiarity, personal taste, and emotional responses. This aligns with prior research showing that users tend to prioritize their own preferences when interacting with recommender systems, even when additional information is provided [25]. Although some participants noted minor or subconscious label effects, these did not alter their choices.

Based on prior work (Section 2) and participant remarks from Study 1[27], we expected fairness labels to influence song rankings, as individuals may hold conscious or unconscious associations with minority groups. However, our results suggest otherwise: musical preference remained the primary driver of ranking decisions.

Further agreement level analysis showed that participants with higher Kendall's Tau (τ) values approached the task more systematically, identifying patterns or grouping strategies, whereas those with lower values often reported confusion and disengagement from the labels. This suggests that label comprehension and acceptance may improve ranking consistency, while confusion increases variability. Consistent with Study 1, participants aligned with system categorization made more consistent selections, whereas those who disagreed showed greater diversity. Although causality cannot be established, these findings suggest that agreement with system categories may support more consistent decision-making, with implications for fairness-oriented music recommender systems.

However, our findings suggest that fairness labels are ineffective if users do not understand, value, or use them. As strong personal preferences limit their impact,relevant and comprehensible labels are needed [18]. As their acceptance depends on users' preferences and worldviews, personalized labelling is a promising direction.

In conclusion, this study highlights the difficulty of shaping user choices in music recommendations. To achieve both personalization and fairness, system designs must be user-friendly, meaningful, and perceived as valuable by users.

6 Conclusion

This study examined how users interpret category-based labels in music recommender systems and whether these labels influence their ranking decisions. A comparison of ranking distributions between two studies – one without category labels and one with label – demonstrated that this is not the case. Further, we used both statistical analysis, via Kendall's Tau "τ", and qualitative data from participant responses to assess the consistency of user choices and the reasoning behind them.

Our findings indicate low agreement in rankings, highlighting the subjective nature of musical preferences and the need for personalized recommender systems. Qualitative analysis further showed that vague category labels reduced their usefulness, suggesting that transparency-based fairness interventions depend on users' ability to understand and engage with the provided information.

In general, category labels exerted minimal direct influence on how users ranked songs. Most participants relied on their own preferences, such as familiarity, emotional response, and prior experience, rather than on the provided labels. Although some participants noted subtle label effects, these were insufficient to alter overall ranking distributions.

Participants with higher Kendall's Tau (τ) values showed greater task engagement, whereas those with lower values more often reported confusion or disengagement with the labels.

These findings highlight that fixed category labels or fairness information alone are insufficient to change user behaviour. Effective interventions must be clear, meaningful, and aligned with user expectations. Future research should explore personalized labelling strategies and alternative fairness presentations, including their long-term effects on user choices.

Declaration on Generative AI

Grammarly was used for grammar and spelling checks. The authors reviewed all content and take full responsibility for the publication.

References

[1] Himan Abdollahpouri. 2019. Popularity Bias in Ranking and Recommendation. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (Honolulu, HI, USA) *(AIES '19)*. Association for Computing Machinery, New York, NY, USA, 529–530. doi:10.1145/3306618.3314309

[2] 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

[3] Young Min Baek. 2014. Relationship between cultural distance and Cross-Cultural music video consumption on YouTube. *Social Science Computer Review* 33, 6 (Dec. 2014), 730–748. doi:10.1177/0894439314562184

[4] Rodrigo Carvalho Borges and Marcelo Gomes de Queiroz. 2018. Automatic music recommendation based on acoustic content and implicit listening feedback. *Revista Música Hodie* 18, 1 (2018), 31–43.

[5] Robin Burke. 2017. Multisided Fairness for Recommendation. *FATML'17* (07 2017). arXiv:1707.00093 doi:10.48550/arXiv.1707.00093

[6] Robin Burke, Nasim Sonboli, and Aldo Ordonez-Gauger. 2018. Balanced neighborhoods for multi-sided fairness in recommendation. In *Conference on fairness, accountability and transparency*. PMLR, 202–214.

[7] Michael A Casey, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney. 2008. Content-based music information retrieval: Current directions and future challenges. *Proc. IEEE* 96, 4 (2008), 668–696.

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

[9] 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

[10] Olmo Cornelis, Micheline Lesaffre, Dirk Moelants, and Marc Leman. 2010. Access to ethnic music: Advances and perspectives in content-based music information retrieval. *Signal Processing* 90, 4 (2010), 1008–1031. https://doi.org/10.1016/j.sigpro.2009.06.020

[11] Paolo Cremonesi, Antonio Donatacci, Franca Garzotto, Roberto Turrin, et al. 2012. Decision-Making in Recommender Systems: The Role of User's Goals and Bounded Resources.. In *Decisions@ recsys*. 1–7. https://www.researchgate.net/profile/Paolo-Cremonesi-2/publication/281705736_Decision-making_in_recommender_systems_The_role_of_user's_goals_and_bounded_resources/links/584de12c08aeb9892526488d/Decision-making-in-recommender-systems-The-role-of-users-goals-and-bounded-resources.pdf

[12] James Davidson, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. 2010. The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*. 293–296.

[13] Callum Davies, Bill Page, Carl Driesener, Zac Anesbury, Song Yang, and Johan Bruwer. 2022. The power of nostalgia: Age and preference for popular music. *Marketing Letters* 33, 4 (2022), 681–692.

[14] Frederik Dhaenens and Jean Burgess. 2019. 'Press play for pride': The cultural logics of LGBTQ-themed playlists on Spotify. *new media & society* 21, 6 (2019), 1192–1211.

[15] Karlijn Dinnissen and Christine Bauer. 2022. Fairness in music recommender systems: a stakeholder-centered mini review. *Frontiers in Big Data* 5 (22 July 2022), 1–9. doi:10.3389/fdata.2022.913608

[16] 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 '25)*. Association for Computing Machinery, New York, NY, USA, 48–57. doi:10.1145/3720553.3746671

[17] 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

[18] 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

[19] Tiago Fernandes Tavares and Leandro Collares. 2020. Ethnic music exploration guided by personalized recommendations: system design and evaluation. *SN Applied Sciences* 2, 4 (2020), 549. https://doi.org/10.1007/s42452-020-2318-y

[20] 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

[21] Bruce Ferwerda, Mark P Graus, Andreu Vall, Marko Tkalcic, and Markus Schedl. 2017. How item discovery enabled by diversity leads to increased recommendation list attractiveness. In *Proceedings of the Symposium on Applied Computing*. 1693–1696.

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

[23] 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

[24] Dietmar Jannach, Iman Kamehkhosh, and Geoffray Bonnin. 2016. Biases in automated music playlist generation: A comparison of next-track recommending techniques. In *Proceedings of the 2016 conference on user modeling adaptation and personalization*. 281–285.

[25] Dietmar Jannach, Lukas Lerche, and Markus Zanker. 2018. *Recommending Based on Implicit Feedback*. Springer International Publishing, Cham, 510–569. doi:10.1007/978-3-319-90092-6_14

[26] Shah Noor Khan and Eelco Herder. 2023. Effects of the spiral of silence on minority groups in recommender systems. In *Proceedings of the 34th ACM Conference on Hypertext and Social Media* (Rome, Italy) *(HT '23)*. Association for Computing Machinery, New York, NY, USA, Article 31, 5 pages. doi:10.1145/3603163.3609041

[27] Shah Noor Khan, Eelco Herder, Libio Goncalves Braz, Karlijn Dinnissen, and Judith Masthoff. 2026. Perceptions of Fairness and its Impact on User Choices in Music Recommendation. (2026).

[28] Shah Noor Khan, Eelco Herder, and Diba Kaya. 2024. Experiences of Non-Mainstream and Minority Users With Music Recommendation Systems. Mensch und Computer 2024 - Workshopband. doi:10.18420/muc2024-mci-ws11-141

[29] 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.

[30] 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

[31] Jurek Leonhardt, Avishek Anand, and Megha Khosla. 2018. User Fairness in Recommender Systems. In *Companion Proceedings of the The Web Conference 2018* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 101–102. doi:10.1145/3184558.3186949

[32] Oleg Lesota, Emilia Parada-Cabaleiro, Stefan Brandl, Elisabeth Lex, Navid Rekabsaz, and Markus Schedl. 2022. Traces of Globalization in Online Music Consumption Patterns and Results of Recommendation Algorithms.. In *ISMIR*. 291–297.

[33] Dugang Liu, Chen Lin, Zhilin Zhang, Yanghua Xiao, and Hanghang Tong. 2019. Spiral of silence in recommender systems. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. 222–230.

[34] Meijun Liu, Xiao Hu, and Markus Schedl. 2018. The relation of culture, socio-economics, and friendship to music preferences: A large-scale, cross-country study. *PloS one* 13, 12 (2018), e0208186.

[35] 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*. 2145–2148.

[36] Masoud Mansoury, Himan Abdollahpouri, Jessie Smith, Arman Dehpanah, Mykola Pechenizkiy, and Bamshad Mobasher. 2020. Investigating potential factors associated with gender discrimination in collaborative recommender systems. In *The thirty-third international flairs conference*.

[37] Jörg Matthes, Johannes Knoll, and Christian von Sikorski. 2018. The "spiral of silence" revisited: A meta-analysis on the relationship between perceptions of opinion support and political opinion expression. *Communication Research* 45, 1 (2018), 3–33.

[38] Danaë Metaxa-Kakavouli, Kelly Wang, James A Landay, and Jeff Hancock. 2018. Gender-inclusive design: Sense of belonging and bias in web interfaces. In *Proceedings of the 2018 CHI Conference on human factors in computing systems*. 1–6.

[39] 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

[40] Elisabeth Noelle-Neumann. 1974. The spiral of silence a theory of public opinion. *Journal of communication* 24, 2 (1974), 43–51.

[41] 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

[42] Lorenzo Porcaro, Carlos Castillo, and Emilia Gómez Gutiérrez. 2021. Diversity by design in music recommender systems. *Transactions of the International Society for Music Information Retrieval. 2021; 4 (1)*. (2021).

[43] Markus Schedl, Peter Knees, and Fabien Gouyon. 2017. New paths in music recommender systems research. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. 392–393.

[44] 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

[45] 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

[46] Anne Siebrasse and Melanie Wald-Fuhrmann. 2023. You don’t know a person (’s taste) when you only know which genre they like: taste differences within five popular music genres based on sub-genres and sub-styles. *Frontiers in Psychology* 14 (2023), 1062146.

[47] 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

[48] Nasim Sonboli, Farzad Eskandanian, Robin Burke, Weiwen Liu, and Bamshad Mobasher. 2020. Opportunistic Multi-aspect Fairness through Personalized Re-ranking. arXiv:2005.12974 [cs.IR] https://arxiv.org/abs/2005.12974

[49] Harald Steck. 2018. Calibrated recommendations. In *Proceedings of the 12th ACM conference on recommender systems*. 154–162.

[50] 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

[51] Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep content-based music recommendation. *Advances in neural information processing systems* 26 (2013).