Perceptions of Fairness and its Impact on User Choices in Music Recommendation*

Shah Noor Khan[1,*], Eelco Herder[1], Libio Goncalves Braz[1], Karlijn Dinnissen[1] and Judith Masthoff[1]

[1]Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands

Abstract

Music recommender systems (MRS) have significantly changed how listeners discover and interact with music. However, these systems often prioritize mainstream content and hence marginalize niche music, as well as users with specific tastes or backgrounds. This imbalance stems from demographic biases as well as algorithmic design, and it may reinforce filter bubbles and limit exposure to diverse music. While previous research mostly addressed fairness in music recommendation from a system-centric perspective, little attention has been given to user perceptions of fairness and how these perceptions affect decision-making. In this paper, we aim to explore participants' initial preferences, how additional information influenced their choices, and the factors motivating changes in decision-making. To address these questions, a mixed-methods user study was conducted with 28 participants, combining quantitative song ranking and categorization tasks with qualitative reflections on fairness categories. The results show that users are generally reluctant to revise their initial choices, even when given more positive or negative information, showing the influence of the commitment principle. By viewing fairness as a user-centered experience rather than merely a system attribute, this study offers new insights for designing recommender systems that better promote diversity and fairness.

Keywords

Fairness, User Study, Music recommender system

1. Introduction

Music plays an important role in our lives and is often associated with particular life events, locations and emotions. Particularly, music allows individuals to convey their identities [1, 2, 3]. Music recommender systems (MRS) provide guidance by suggesting music to users [4], therewith influencing what is being played [5]. However, despite tailoring recommendations to users' preferences, these systems ironically tend to favor mainstream music [6], resulting in unsatisfactory suggestions for users with niche or non-mainstream tastes [7, 8].

Effective RS should not only align with established user preferences but also introduce novel music selections that extend prior listening history [9]. Existing methodologies primarily focus on accuracy and preference alignment, yet often neglect the importance of joint item selection and recommendation timing, contributing to suboptimal recommendation strategies [10]. Furthermore, fairness and inclusion of minority, non-mainstream users and user groups is becoming increasingly important in RS. However, studies [11, 12] indicate that increased algorithmic fairness does not consistently improve users' perceived fairness.

As empirical studies reveal a gap between perceived and metric-based fairness, this poses a core challenge for fairness-aware design [7] and requires exploring user perceptions and influence on fairness. While prior research has examined artists' perceptions on fair representation [13, 14], little is known about how users' fairness perceptions affect their decision-making. Providing fairness-related information may promote more equitable choices [15], but its effectiveness varies by user, as shown in

Joint Proceedings of the ACM UMAP Workshops 2026, UMAP 2026, June 8–11, 2026, Gothenburg, Sweden

*Corresponding author.
s.n.khan1@uu.nl (S. N. Khan); e.herder@uu.nl (E. Herder); l.b.goncalvesbraz@uu.nl (L. Goncalves Braz); k.dinnissen@uu.nl (K. Dinnissen); j.f.m.masthoff@uu.nl (J. Masthoff)
0009-0006-4899-2987 (S. N. Khan); 0000–0003–0107–299X (E. Herder); 0009-0007-2906-6811 (L. Goncalves Braz); 0000-0003-2498-2881 (K. Dinnissen); 0000-0001-8099-0515 (J. Masthoff)

nudging studies on music selection [16].

Fairness in terms of representation of user groups and their respective music choices may be perceived neutral, positively or negatively by users not belonging to these groups. This is expected to depend on a user's perception of different minority groups. However, it is unknown whether fairness explanations will affect a user's music preferences. Prior research [17] provided indications that this may not be the case. For instance, one may be supportive of the LGBTQ+ movement, but not necessarily enjoy quirky queer songs. In our study, we explore participants' initial preferences and whether additional information on the songs' cultural and societal associations alters their preferred ordering. We also examine whether and how participants relate songs to minority groups. Most importantly, we aim to understand the factors that motivate participants to change or maintain their choices.

2. Related Work

In cultural areas like music, fairness and diversity matter greatly [6, 7]; nevertheless, most research still focuses on the technical side. Far less is known about how users perceive fairness, or how those perceptions affect their satisfaction.

2.1. Fairness in Recommender System

Popularity bias remains a persistent obstacle: users tend to gravitate toward mainstream music, causing long-tail or niche items to be underrepresented [18, 19, 8, 20]. As a result, this bias disproportionately amplifies dominant artists and genres, which are primarily Western and English-language, similar to the origins of the dominant, mainstream user groups [6, 18, 21, 22].

Increasing diversity in recommendations can improve the appeal and user satisfaction of recommendation lists. However, greater diversity may also increase the difficulty of making decisions [23, 24]. Systems that do not account for user-specific preferences in the decision process may undermine perceived fairness, particularly for users with niche interests [21]. For instance, **gender bias** amplifies these challenges, as female artists are frequently underrepresented.

These issues underscore the necessity for equity and systemic change that extends beyond algorithmic performance metrics [25, 26, 27]. Most fairness-aware systems are evaluated using offline metrics, which fail to capture user perceptions or interactive experiences [28, 7]. While top-down interventions may satisfy statistical targets, they can still appear unfair or irrelevant to users.

To address these gaps, **user-centered approaches** are necessary. In these approaches, fairness is co-defined with users to ensure that algorithmic outcomes align with perceived fairness [29].

2.2. Transparency, User control and Perceived fairness

Transparency is critical for fostering user acceptance and trust in recommender systems (RS) [30]. Providing explanations can encourage faster decision-making [31], and promote sustained engagement [32, 30, 33]. However, transparency is not a one-size-fits-all principle, as different stakeholders have different goals [30].

The application of fairness constraints without adequate explanation can diminish user trust, as users may experience alienation if outcomes differ from their expectations, even when these outcomes are statistically fair [25]. [25] indicate that providing contextual cues, such as clarifying the prioritization of diversity, assists users in interpreting fairness interventions and sustaining trust.

User control is key to fairness in recommender systems. Allowing users to set fairness preferences aligns recommendations with individual needs [34] and helps balance relevance and diversity. This reduces dissatisfaction when system-imposed fairness differs from user expectations [35].

Perceived fairness is influenced by the degree of the user's familiarity with the recommended items. Excessive exposure to either familiar or novel items can decrease user satisfaction. This highlights the importance of calibrated diversity that maintains a balance between user comfort and exploration [36].

Interface design choices, such as modifying genre visibility or implementing nudging mechanisms, can promote exploration while preventing user overload [37].

2.3. Bias Mitigation and Personalization

Research on online platforms highlights the challenge of applying fairness in personalization and recommendation systems due to their multi-sided nature involving users and creators [38]. Researchers have recently explored rich sub-group fairness, aiming for fairness across multiple intersecting sub-group categories [39]. In another study, Previous research explored how the visibility of minority groups could be amplified or mitigated by varying levels of homophily within each subgroup [40], therewith making other users more aware of what aspects define the subgroup's identity.

Fairness encompasses multiple dimensions, such as popularity, genre, gender, and nationality [41]. Allowing users to customize fairness objectives increases satisfaction and reduces cognitive dissonance resulting from unexpected or misaligned outcomes [42].

Hybrid approaches integrate personalization with fairness constraints, demonstrating that relevance and bias reduction can be compatible objectives [43]. Dynamic adjustment of fairness parameters enables systems to accommodate changes in user intent and context. Consequently, fairness is conceptualized as a continuous negotiation rather than a static configuration [44].

Biases in data and model objectives influence not only what is recommended but also how recommendations are perceived. Demographic, cultural, and popularity biases embedded in historical data can lead to unequal exposure, even when systems appear accurate [7]. Music preferences for diversity and popularity vary widely across countries and cultural groups [45], underscoring the need for fairness definitions that adapt to context. Unfortunately, popularity bias prioritizes tracks with high play counts, marginalizing niche artists and reinforcing mainstream hierarchies, ultimately leaving little to no space for music and artists that differentiate too much from what is considered 'the norm' [22, 18].

To combat this effect, algorithms designed to enhance diversity by introducing users to various genres or employing multi-objective optimization can reveal novel yet pertinent content [23, 36], thereby supporting more informed decision-making and exploratory behavior.

In summary, fairness becomes an increasingly important topic in RS. Negative effects of biases and the role of transparency and explanations have received quite some attention. However, the effect of fairness explanations on a user's music preferences is an under-researched area, and therefore the focus of our study.

3. Methodology

We conducted an online user study to investigate whether providing additional information about the fairness attributes of songs influences users' decision-making in ranking tasks. More concretely, we asked them to first rank songs based on their individual taste and then provided labels on how some of these songs were associated with specific groups defined by gender, age, orientation and nationality. The study also examined how participants reflect on these fairness categories themselves and which of these they considered important, with a particular focus on user experience. The university's ethics and privacy board classified the study as low risk, requiring no further assessment.

3.1. Selection of Fairness Categories

In this study, we address several key dimensions that have been identified as sources of bias in music recommendation systems, as discussed in Section 2.1. *Popularity bias* often arises when mainstream music dominates recommendations, overshadowing niche works [46]. *Gender representation* is another persistent concern, with female and non-binary artists frequently under-represented compared to male artists [25, 47]. *Nationality* also influences visibility, as local artists experience varying levels of exposure across different regions [45]. Recommending ethnic or region-specific music is particularly challenging, given its limited compatibility with Western-centered recommendation methods and smaller user bases [48, 49]. Further, studies [50] indicate that late adolescence and early adulthood significantly influence individuals' music preferences. Music released during periods targeting specific *age demographics* is more likely to be favoured by them. Finally, research has shown that *LGBTQ-themed* playlists both reflect and shape LGBTQ identities and cultures, highlighting an additional important aspect of fairness in music recommendation [51].

3.2. Study Setup

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

Step 1: Pre-study survey. After providing informed consent, participants provided demographic information (age, nationality, gender) before proceeding to the next task. We only included participants residing in the Netherlands. The study was completed by 28 participants with a mean age of 37 years, ranging from 25 to 61. Nineteen participants were European, with a majority of 15 from the Netherlands. The gender distribution included 14 women, 13 men, and 1 Non-Binary/Other.

Step 2: Initial ranking. Participants were presented with 20 songs in a grid-style format(Figure 1), each assigned a randomly defined rank number, to avoid middle-option bias [52]. Each song was displayed as a card containing the title and artist name, with the option to listen to a short preview [16, 17]. The songs were selected based on various fairness categories, as explained earlier in Section 3.1. At this point in time, no labels regarding the fairness categories were presented to the participants.

Participants ranked the songs based on personal preference by dragging and dropping them to define their desired order before submitting their responses and proceeding to the next task. The purpose of this task was to capture the participants' baseline music preferences. The details on song selection are in the next Section 3.3.

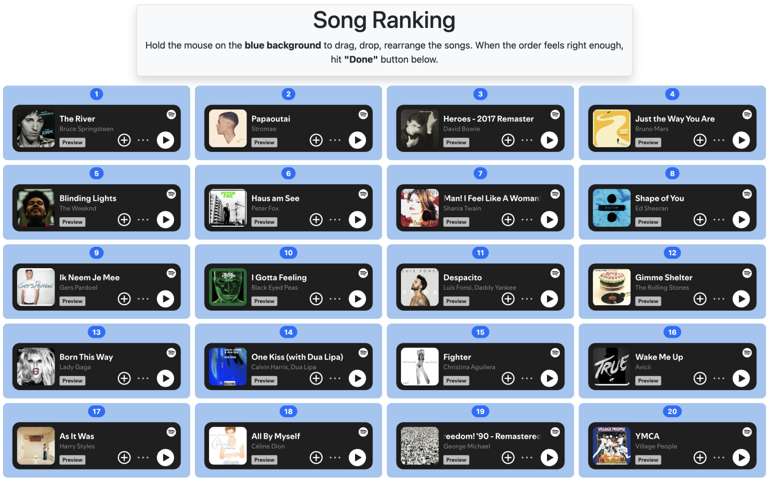

Figure 1: First Ranking Task : Song locations are labeled from 1 to 20, and participants ranked the songs from most (1) to least (20) preferred by dragging and dropping.

Step 3: Categorization task. In this task, participants were asked to place the 20 songs from the previous task into the four predefined fairness categories related to gender, age, orientation, nationality and mainstreamness, as described in Section 3.1, with each category having three slots(Figure 2). Each category included a short description to help guide their choices. Songs appeared in the same rank order as the participant had assigned earlier in Step 2. Each song could be placed in only one category, but participants were free to move songs between categories. This task aimed to understand how participants associated songs with the different fairness categories.

Step 4: Comparison task. For each fairness category, participants saw how their own song categorizations compared to the system's (Figure 3). The interface included open-ended questions for participants to reflect on the categories and the system's choices. The task aimed to gather perspectives on the categories, assess agreement with the system's interpretation, and understand how participants perceived the comparison.

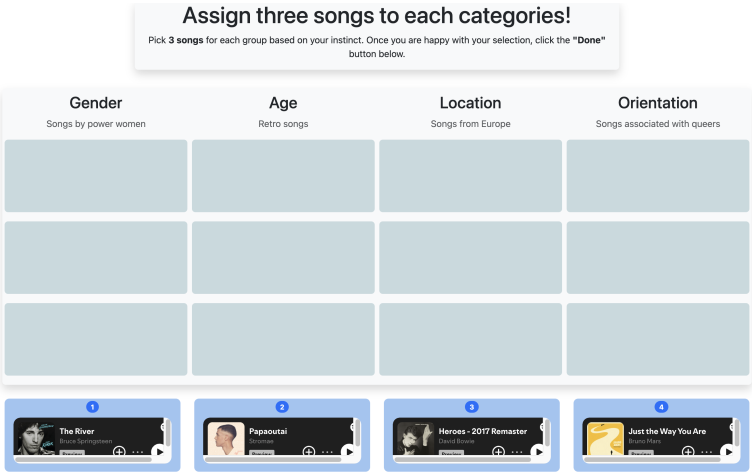

Figure 2: Categorization Task : Participants were asked to assign the presented songs into categories by dragging and dropping.

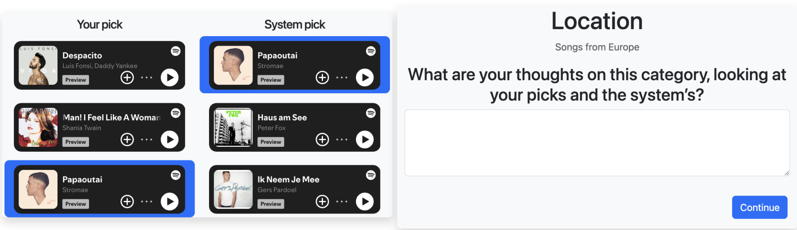

Figure 3: Comparison Task :Participants compared their own selections with the system's and provided feedback on the categories.

Step 5: Final ranking. In this task, participants were presented with the same 20 songs again, displayed in the ranking order they had previously created (Figure 4). This time, the interface included explicit labels, indicating the fairness categories to which each song belonged. Participants were given the option to adjust their rankings if they wished, or to retain their original order.

Step 6: User experience. At the end of the study, participants were presented with an optional open-ended question to share their reflections on the overall user experience.

3.3. Song Selection

We included 20 familiar songs to ensure that participants had enough options without any one dominating song. We selected three songs per fairness dimension, adding up to twelve song. In addition, eight mainstream songs were included as a baseline, because we wanted our interface to resemble the homepage of a music platform, which primarily features mainstream songs.

As our task aimed to ensure participants' familiarity with as many songs as possible, we selected tracks from websites listing top artists relevant to our study dimensions. These websites originated primarily from the Netherlands, where our participants resided, or from Billboard.com when local sources were insufficient. Spotify popularity scores[1] were used as a proxy for familiarity.

Gender dimension. The artists in this dimension were selected to represent female artists, an underrepresented gender category, as discussed in Section 3.1. To ensure clarity in gender attribution, only solo artists were included. The final artist selection was based on Billboard's overview of the top

[1]Retrieved from https://developer.spotify.com/documentation/web-api/reference/get-track

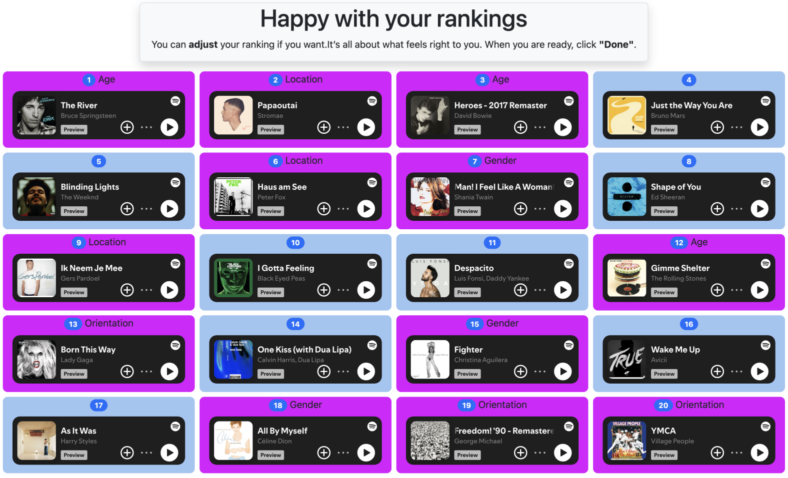

Figure 4: Final Ranking Task : Songs were labeled from 1 to 20, with additional information provided, and participants ranked them from most (1) to least (20) preferred by dragging and dropping.

(power) women artists of the 21st century[2]

Age dimension. Here, we selected classic rock songs released in the 1970—1980 that are commonly associated with older generations, as our participants were either very young or not yet born when the song was released, but they would most likely still be familiar with the songs. Artists and songs were selected from the 2024 Classic Rock Top 100, published by prominent Dutch radio station Arrow Classic Rock[3]

LGBTQ+ dimension. Here, we included songs recognized within the queer community for their thematic content or the values represented by the performing artists. Sources included Billboard's overview of queer anthems and a Dutch 'Gay Top 100' hit list[4]

Nationality dimension. To reflect music with non-English lyrics from countries geographically close to the Netherlands, we selected songs from Belgium (in French), Germany (in German), and the Netherlands (in Dutch). To ensure familiarity among our Netherlands-based participants, we used Top40.nl, the main Dutch music chart website, as a source for major hits in each language[5]

Super-mainstream songs. Lastly, the task required a baseline of highly popular mainstream songs that clearly did not fall under any fairness category. To fulfill this criterion, we selected tracks that were *not* performed by women solo artists, released after 2009, not associated with LGBTQ+ themes, and either in English or performed by artists from the US, Canada, or UK. Sources included ranked lists from the Netherlands and the US, featuring songs that have remained at number one on their respective charts for the longest periods.

4. Results

We first present a quantitative analysis of the participants' song ranking, followed by a qualitative analysis of participants' perceptions and reasons behind their song categorization.

4.1. Quantitative Analysis

This section presents the ranking of songs across the four fairness categories. It also shows correlations between participants in terms of song ranking and whether their selections changed across different ranking rounds.

[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.

4.1.1. Song Rankings by Category

As mentioned in Section 3.3, we categorized the songs and, in order to investigate whether a particular category would rank consistently higheror lower overall compared to others in terms of average ranking. Table 1 shows the results from both ranking rounds. As expected, mainstream songs had the highest rankings, as shown by their low average ranking values (with 1.0 as the best possible ranking). Orientation, gender, and location followed. This pattern remained the same in the second round, with only minor differences between the first and second ranking.

Table 1
Average Rankings of Song Categories Across Two Rounds

Category	First Ranking	Second Ranking
Mainstream	9.34	9.43
Orientation	10.42	10.21
Gender	11.13	11.02
Location	11.50	11.55
Age	12.02	12.05

4.1.2. Kendal Tau Correlations

In order to explore whether any patterns were emerging among the participants of the study, we used Kendall's tau, a statistical measure that quantified how similarly the participants ranked the songs among themselves and between ranking rounds.

Participant Comparisons. To find if there were any patterns or similarities among the study participants and to see how closely their rankings matched, we created a 28x28 matrix (Figure 5), which displays the correlation between each pair of participants.

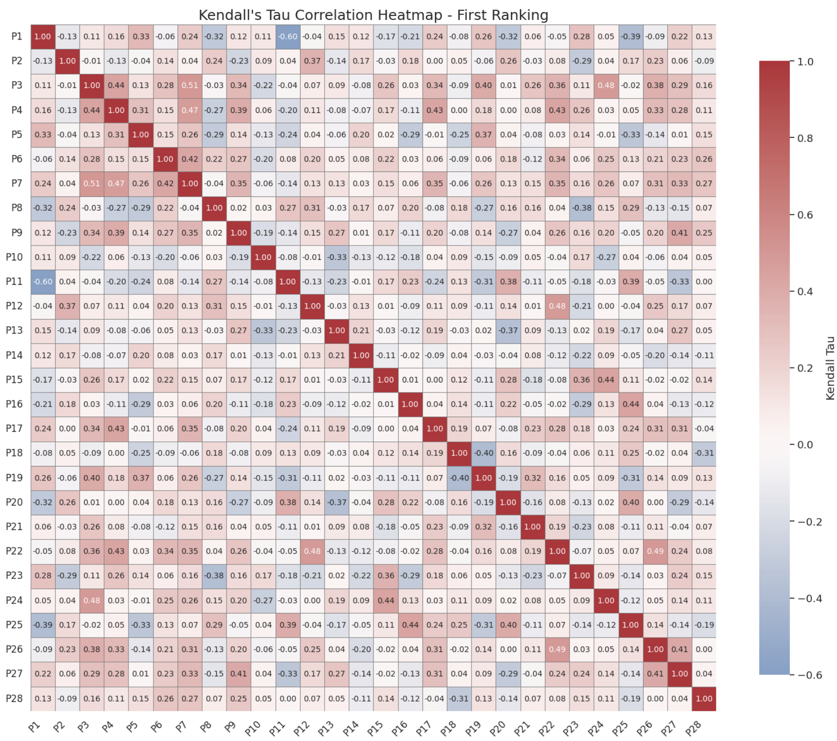

Figure 5: Kendall Tau Matrix

Elaborating on this, the results in Table 2 confirm that participants ranked songs in very different ways. Both the average ("τ" = 0.056) and median ("τ" = 0.053) Kendall's "τ" values are close to zero, meaning there was little agreement about the rankings. The highest τ value suggests that some pairs of participants agreed somewhat, but the lowest value shows that some ranked the songs almost in opposite order.

Table 2
Overall Kendall's Tau Statistics - First Ranking

Statistic	Average Tau	Minimum Tau	Maximum Tau	Median Tau
Value	0.056	-0.600	0.505	0.053

As no widely accepted closed-form approach exists for conducting power analyses directly for Kendall's τ, Pearson's correlation was used as an approximation. With 28 participants, a significance level of $\alpha = 0.05$, and such small τ values, there was not enough statistical power to reliably detect agreement.

Rankings Rounds Comparisons. We also further quantified the observed small differences between the initial and final rankings of participants. To do this, we used Kendall's tau to measure the variation in song rankings between the two rounds for each participant. For 24 participants, the Kendall's tau value was 1.0, indicating that they did not change their rankings between the two rounds at all. For the remaining 4 participants, the Kendall's tau value was slightly less than 1, indicating only some small changes in their rankings.

4.2. Qualitative Analysis

This section covers the qualitative part of the study, focusing on participants' perceptions and thoughts about how songs were categorized, as described in Section 3.2. The coding framework is described in Table 3.

Based on the between-participant Kendall tau matrix (Figure 5) , we created two groups of 5 participants: those with the highest average Kendall tau ("τ") values compared to all other participants (i.e. those with rankings that are most similar to the other participants) and those with the lowest average Kendall tau ("τ") values (indicating that their rankings had little to none agreement with other participants' rankings). This approach helped us to see whether participants' perceptions and interpretation of system choices may be related to (the lack of) shared understanding, as represented by their song ranking.

Table 3
Coding framework for the qualitative data

Code	**Definition**
Agreement	User agrees with the system's categorization.
Disagreement	User disagrees with the system's categorization.
Decision criteria	Cues users rely on to assign categories.
Category ambiguity	Uncertainty or critique of the category itself.
Task constraints	Effects of interface limitations.

4.2.1. Participants with high average Kendall Tau ("τ")

System Affinity. We found patterns in the qualitative data for participants with high average Kendall tau values (Table 4). These participants found their selections to be closer to the system's picks and agreed with the system for the categories. For example, "No brainer."P6 "What can I say, perfect score!" P7 These examples demonstrate concordance in their judgments, suggesting shared perceptions regarding the assignment of songs to categories.

Category Affinity. For categories like *female representation*, participants generally found the system's picks aligned with this aspect. For some songs, such as those by Lady Gaga, there was

Table 4
Participants with High Average Kendall Tau (“τ”) Values

Participant	Average Kendall Tau
P7	0.189
P3	0.128
P6	0.136
P4	0.128
P22	0.128

a consensus among participants (p4, p6, p22) that these fit both the *queer* icon and power women categories. For the *location* category, participants sometimes made choices based on song language and, at other times, on the artist's origin. For *age*, participants considered it relative to their own experiences, loosely defining songs from the 70s–90s as retro. “Parents would have listened to them.”P22

Overall, high average Kendall tau values show that participants who categorized songs similarly in one category tended to agree across multiple categories, reflecting shared cultural understandings of queerness, female empowerment, European identity, and musical retro appeal.

4.2.2. Participants with low average Kendall Tau (“τ”)

System Affinity. For participants with low average Kendall tau values (Table 5), we found patterns of disagreement with the system selections. They tended to diverge more from the system's selection logic. “That apparently only Canadian women can be powerful.”P11 Some participants indicated doubt and showed disconnection with the system's categorization.“I typically find it useless to categorize music based on the author's attribute.”P16

Category Affinity. In the gender category, participants found that the way songs were shown, with images and artist selection, made it easier for them to categorize. For the location category, disagreements arose over how 'European' should be interpreted, with some rejecting that it should be based on nationality or language. “That UK is also part of Europe.”P11 “Despacito is Spanish style!!!”P1 In the age category, participants disagreed with the system's association of certain music as 'retro,' viewing it as subjective. Divergence in the age category was less pronounced than for orientation.

In general, these low Kendall tau values highlight areas where the choices of the participants conflicted with the system, pointing to subjectivity,also showing how presentation affects interpretation.

Table 5
Participants with Low Average Kendall Tau (“τ”) Values

Participant	Average Kendall Tau
P10	-0.0604
P11	-0.0456
P18	-0.007
P1	-0.001
P16	0.0011

5. Discussion

The results of the song ranking in Section 4.1.1 show that *mainstream songs* were ranked higher by participants compared to other categories. One reason for this ranking is the effect of *popularity bias* [8], which is consistent with previous literature [22, 18]. Another likely reason for the preference toward mainstream songs is *familiarity* with the content itself, which aids in the decision-making process [36].

The Kendall's tau values for our sample set were low, indicating a surprisingly *low level of consensus* among participants in terms of their song rankings. This suggests that participants had diverse musical tastes and preferences. Such diversity could be attributed to several factors, including participants' personality traits [53], their individual goals during the study, and the way the available options [54] influenced their decision-making. Also, a recent representative survey revealed that patterns in genre preference do not necessarily translate to patterns for sub-genres or individual songs [55]. This emphasizes personalization as a critical user experience factor influencing interaction and outcomes.

Differences in familiarity with the music between the participants and the (intentional) lack of clear criteria during the study may have also contributed to this inconsistency, as each participant – as instructed – relied on their own preferences and interpretation of the ranking task. The lack of significant agreement indicates that no single song or type of music dominated in their selections – i.e. the songs were quite comparable in terms of popularity .

Furthermore, there was *no significant change in the rankings* for the majority of participants across both rounds, indicating that information about the fairness categories did not influence their decision-making strongly enough to compel them to change their initial preferences. A similar pattern has been observed in other studies [17], where participants also tended to maintain their original choices. One possible explanation for this behavior involves *loss aversion* and the *commitment principle* [56], which suggest that people tend to remain consistent with their earlier decisions, even when new information or better options become available, often to avoid a perceived loss.

Section 4.2 describes how participants evaluated their own song categorization *compared to the system's choices*. A pattern emerged: participants whose song rankings were moderately similar to other participants tended to view the system's picks as accurate and aligned. In contrast, those with least similarity often disagreed with the system.

When comparing the participants' remarks from both groups, it appears that participants who agreed more with the system were more familiar with the sub-group categories and held shared views on songs related to them. This suggests that alignment with system recommendations depends strongly on users' familiarity with and interpretation of the underlying categories.

6. Conclusion

This study explored how perceptions of fairness influence user decision-making. We observed that ranking behavior differed wildy between participants , confirming that song preferences are highly individual. Further, the results indicate that informing participants about the relation of a song with a particular fairness category (i.e. gender, age, orientation and nationality) had only little impact on their initial rating. These findings show that initial choices strongly influence participants, and – assumingly because of commitment bias and loss aversion – people rarely change their original selections. This suggests that if a systems aims to encourage fairness-related choices, it is more effective to present these options *before* they make a decision. The results also suggest that participants with similar selections tended to interpret the fairness categories in similar ways, while those with different tastes often had different interpretations.

With the above main findings, this study provides several insights to be taken into account in future fairness-related studies and insights to inform the design of persuasive strategies for fairness. First, interventions are likely to be more effective when users are made aware of them as early as possible, to avoid commitment bias. Second, the effectiveness of interventions may be further increased by actively furthering common, shared understanding.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] S. Frith, et al., Sound effects; youth, leisure, and the politics of rock'n'roll., Sound effects; youth, leisure, and the politics of rock'n'roll. (1981).

[2] A. C. North, D. J. Hargreaves, Music and adolescent identity, Music education research 1 (1999) 75–92.

[3] A. Laplante, Improving music recommender systems: What can we learn from research on music tastes?, in: ISMIR, 2014, pp. 451–456.

[4] M. Liu, X. Hu, M. Schedl, The relation of culture, socio-economics, and friendship to music preferences: A large-scale, cross-country study, PloS one 13 (2018) e0208186.

[5] M. Schedl, P. Knees, B. McFee, D. Bogdanov, Music recommendation systems: Techniques, use cases, and challenges, in: F. Ricci, L. Rokach, B. Shapira (Eds.), Recommender Systems Handbook, 3rd ed., 2022, pp. 927–971. URL: https://doi.org/10.1007/978-1-0716-2197-4_24. doi:10.1007/978-1-0716-2197-4_24.

[6] M. Schedl, H. Zamani, C.-W. Chen, Y. Deldjoo, M. Elahi, Current challenges and visions in music recommender systems research, International Journal of Multimedia Information Retrieval 7 (2018) 95–116. URL: https://doi.org/10.1007/s13735-018-0154-2. doi:10.1007/s13735-018-0154-2.

[7] M. Ekstrand, M. Tian, I. Azpiazu, J. Ekstrand, O. Anuyah, D. McNeill, M. Pera, All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness, in: S. Friedler, C. Wilson (Eds.), Proceedings of the 1st Conference on Fairness, Accountability and Transparency, volume 81, PMLR, 2018, pp. 172–186. URL: https://proceedings.mlr.press/v81/ekstrand18b.html.

[8] D. Kowald, P. Muellner, E. Zangerle, C. Bauer, M. Schedl, E. Lex, Support the underground: characteristics of beyond-mainstream music listeners, EPJ Data Science 10 (2021) 14. URL: https://doi.org/10.1140/epjds/s13688-021-00268-9.

[9] R. C. Borges, M. G. d. Queiroz, Automatic music recommendation based on acoustic content and implicit listening feedback, Revista Música Hodie 18 (2018) 31–43.

[10] D. Jannach, I. Kamehkhosh, G. Bonnin, 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, 2016, pp. 281–285.

[11] B. Ferwerda, E. Ingesson, M. Berndl, M. Schedl, 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, 2023, p. 357–361. URL: https://doi.org/10.1145/3576840.3578287. doi:10.1145/3576840.3578287.

[12] R. Ungruh, K. Dinnissen, A. Volk, M. S. Pera, H. Hauptmann, 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, 2024. URL: https://doi.org/10.1145/3640457.3688102.

[13] K. Dinnissen, C. Bauer, 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, 2023. URL: https://doi.org/10.1145/3565472.3592960. doi:10.1145/3565472.3592960.

[14] A. Ferraro, X. Serra, C. Bauer, What is fair? Exploring the artists' perspective on the fairness of music streaming platforms, in: C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, K. Inkpen (Eds.), Human-Computer Interaction – INTERACT 2021: 18th IFIP TC 13 International Conference, volume 12933 of *INTERACT '21*, 2021, pp. 562–584. doi:10.1007/978-3-030-85616-8_33.

[15] G. Alves, D. Jannach, R. Ferrari De Souza, M. G. Manzato, User perception of fairness-calibrated recommendations, in: Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, UMAP '24, 2024, p. 78–88. URL: https://doi.org/10.1145/3627043.3659558. doi:10.1145/3627043.3659558.

[16] S. Fatahi, M. Mousavifar, J. Vassileva, 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, 2023, pp. 66–77.
[17] K. Dinnissen, S. N. Khan, H. Hauptmann, E. Herder, J. Masthoff, 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, 2025, p. 48–57. URL: https://doi.org/10.1145/3720553.3746671. doi:10.1145/3720553.3746671.
[18] Ò. Celma, Music Recommendation, 2010, pp. 43–85. URL: https://doi.org/10.1007/978-3-642-13287-2_3. doi:10.1007/978-3-642-13287-2_3.
[19] A. Van den Oord, S. Dieleman, B. Schrauwen, Deep content-based music recommendation, Advances in neural information processing systems 26 (2013).
[20] S. Goel, A. Broder, E. Gabrilovich, B. Pang, Anatomy of the long tail: ordinary people with extraordinary tastes, in: Proceedings of the third ACM international conference on Web search and data mining, 2010, pp. 201–210.
[21] D. Kowald, M. Schedl, E. Lex, 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, 2020, pp. 35–42. URL: https://doi.org/10.1007/978-3-030-45442-5_5.
[22] H. Abdollahpouri, M. Mansoury, R. Burke, B. Mobasher, The unfairness of popularity bias in recommendation, arXiv preprint arXiv:1907.13286 (2019).
[23] B. Ferwerda, M. P. Graus, A. Vall, M. Tkalcic, M. Schedl, How item discovery enabled by diversity leads to increased recommendation list attractiveness, in: Proceedings of the Symposium on Applied Computing, 2017, pp. 1693–1696.
[24] M. C. Willemsen, B. P. Knijnenburg, M. P. Graus, L. C. Velter-Bremmers, K. Fu, Using latent features diversification to reduce choice difficulty in recommendation lists, in: Joint Workshop on Human Decision Making in Recommender Systems (Decisions@ RecSys 2011) and User-Centric Evaluation of Recommender Systems and Their Interfaces-2 (UCERSTI 2001), CEUR-WS. org, 2011, pp. 14–20. URL: https://research.tue.nl/en/publications/using-latent-features-diversification-to-reduce-choice-difficulty.
[25] A. Ferraro, X. Serra, C. Bauer, Break the loop: Gender imbalance in music recommenders, in: Proceedings of the 2021 Conference on Human Information Interaction and Retrieval, CHIIR '21, 2021, pp. 249–254. doi:10.1145/3406522.3446033.
[26] A. C. Epps-Darling, H. Cramer, R. T. Bouyer, Artist gender representation in music streaming, in: Proceedings of the 21st International Society for Music Information Retrieval Conference, ISMIR, 2020, pp. 248–254. URL: https://doi.org/10.5281/zenodo.4245416. doi:10.5281/zenodo.4245416.
[27] D. Shakespeare, L. Porcaro, E. Gómez, C. Castillo, 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), volume 2697 of *ComplexRec-ImpactRS 2020*, CEUR Workshop Proceedings, 2020, pp. 1–9. URL: https://arxiv.org/pdf/2009.01715.
[28] K. Dinnissen, C. Bauer, Fairness in music recommender systems: a stakeholder-centered mini review, Frontiers in Big Data 5 (2022) 1–9. URL: https://doi.org/10.3389/fdata.2022.913608. doi:10.3389/fdata.2022.913608.
[29] N. Sonboli, J. J. Smith, F. Cabral Berenfus, R. Burke, C. Fiesler, Fairness and transparency in recommendation: The users' perspective, in: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '21, 2021, p. 274–279. URL: https://doi.org/10.1145/3450613.3456835. doi:10.1145/3450613.3456835.
[30] N. Tintarev, J. Masthoff, Explaining recommendations: Design and evaluation, in: Recommender Systems Handbook, 2015, pp. 353–382. URL: https://link.springer.com/content/pdf/10.1007/978-1-4899-7637-6.pdf.
[31] R. Sinha, K. Swearingen, The role of transparency in recommender systems, in: CHI '02 Extended Abstracts on Human Factors in Computing Systems, CHI EA '02, ACM, New York, NY, USA, 2002,

p. 830–831. URL: https://doi.org/10.1145/506443.506619. doi:10.1145/506443.506619.
[32] F. Gedikli, D. Jannach, M. Ge, How should i explain? a comparison of different explanation types for recommender systems, International Journal of Human-Computer Studies 72 (2014) 367–382. URL: https://doi.org/10.1016/j.ijhcs.2013.12.007.
[33] G. Zhao, H. Fu, R. Song, T. Sakai, Z. Chen, X. Xie, X. Qian, Personalized reason generation for explainable song recommendation, ACM Trans. Intell. Syst. Technol. 10 (2019). URL: https://doi.org/10.1145/3337967. doi:10.1145/3337967.
[34] K. Dinnissen, C. Bauer, How control and transparency for users could improve artist fairness in music recommender systems, in: Ismir 2023 Hybrid Conference, 2023.
[35] J. Leonhardt, A. Anand, M. Khosla, User fairness in recommender systems, in: Companion Proceedings of the The Web Conference 2018, WWW '18, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2018, p. 101–102. URL: https://doi.org/10.1145/3184558.3186949. doi:10.1145/3184558.3186949.
[36] D. Jannach, L. Lerche, M. Zanker, Recommending Based on Implicit Feedback, Springer International Publishing, Cham, 2018, pp. 510–569. URL: https://doi.org/10.1007/978-3-319-90092-6_14. doi:10.1007/978-3-319-90092-6_14.
[37] Y. Liang, M. C. Willemsen, Promoting music exploration through personalized nudging in a genre exploration recommender, International Journal of Human–Computer Interaction 39 (2023) 1495–1518. URL: https://doi.org/10.1080/10447318.2022.2108060. doi:10.1080/10447318.2022.2108060.
[38] R. Burke, N. Sonboli, A. Ordonez-Gauger, Balanced neighborhoods for multi-sided fairness in recommendation, in: Conference on fairness, accountability and transparency, PMLR, 2018, pp. 202–214.
[39] M. Kearns, S. Neel, A. Roth, Z. S. Wu, Preventing fairness gerrymandering: Auditing and learning for subgroup fairness, in: International conference on machine learning, PMLR, 2018, pp. 2564–2572.
[40] F. Fabbri, F. Bonchi, L. Boratto, C. Castillo, The effect of homophily on disparate visibility of minorities in people recommender systems, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 14, 2020, pp. 165–175. URL: https://doi.org/10.1609/icwsm.v14i1.7288.
[41] R. Burke, Multisided fairness for recommendation, FATML'17 (2017). URL: https://arxiv.org/abs/1707.00093. doi:10.48550/arXiv.1707.00093. arXiv:1707.00093.
[42] N. Sonboli, F. Eskandanian, R. Burke, W. Liu, B. Mobasher, Opportunistic multi-aspect fairness through personalized re-ranking, 2020. URL: https://arxiv.org/abs/2005.12974. arXiv:2005.12974.
[43] G. Farnadi, P. Kouki, S. K. Thompson, S. Srinivasan, L. Getoor, A fairness-aware hybrid recommender system, 2018. URL: https://arxiv.org/abs/1809.09030. arXiv:1809.09030.
[44] R. Burke, N. Mattei, V. Grozin, A. Voida, N. Sonboli, Multi-agent social choice for dynamic fairness-aware recommendation, in: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, UMAP '22 Adjunct, Association for Computing Machinery, New York, NY, USA, 2022, p. 234–244. URL: https://doi.org/10.1145/3511047.3538032. doi:10.1145/3511047.3538032.
[45] Y. M. Baek, Relationship between cultural distance and cross-cultural music video consumption on youtube, Social Science Computer Review 33 (2014) 730–748. URL: https://doi.org/10.1177/0894439314562184. doi:10.1177/0894439314562184.
[46] H. Abdollahpouri, Popularity bias in ranking and recommendation, in: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, AIES '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 529–530. URL: https://doi.org/10.1145/3306618.3314309. doi:10.1145/3306618.3314309.
[47] M. Mansoury, H. Abdollahpouri, J. Smith, A. Dehpanah, M. Pechenizkiy, B. Mobasher, Investigating potential factors associated with gender discrimination in collaborative recommender systems, in: The thirty-third international flairs conference, 2020.
[48] O. Cornelis, M. Lesaffre, D. Moelants, M. Leman, Access to ethnic music: Advances and perspectives

in content-based music information retrieval, Signal Processing 90 (2010) 1008–1031. URL: https://doi.org/10.1016/j.sigpro.2009.06.020.
[49] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, et al., The youtube video recommendation system, in: Proceedings of the fourth ACM conference on Recommender systems, 2010, pp. 293–296.
[50] C. Davies, B. Page, C. Driesener, Z. Anesbury, S. Yang, J. Bruwer, The power of nostalgia: Age and preference for popular music, Marketing Letters 33 (2022) 681–692.
[51] F. Dhaenens, J. Burgess, ‘press play for pride’: The cultural logics of lgbtq-themed playlists on spotify, new media & society 21 (2019) 1192–1211.
[52] A. Simons, M. Weinmann, M. Tietz, J. vom Brocke, 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). URL: http://hdl.handle.net/10125/41687.
[53] L. Chen, W. Wu, L. He, How personality influences users’ needs for recommendation diversity?, in: CHI ’13 Extended Abstracts on Human Factors in Computing Systems, CHI EA ’13, ACM, New York, NY, USA, 2013, p. 829–834. URL: https://doi.org/10.1145/2468356.2468505. doi:10.1145/2468356.2468505.
[54] P. Cremonesi, A. Donatacci, F. Garzotto, R. Turrin, et al., Decision-making in recommender systems: The role of user’s goals and bounded resources., in: Decisions@ recsys, 2012, pp. 1–7. URL: 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.
[55] A. Siebrasse, M. Wald-Fuhrmann, 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.
[56] D. Kahneman, O. Sibony, C. R. Sunstein, Noise: a flaw in human judgment, Hachette UK, 2021.