

User-Driven Fairness in Music Recommendations: Effects on Experience

Shah Noor Khan^{1,*}, Jesse Nieuwkoop¹ and Judith Masthoff¹

¹*Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands*

Abstract

This study investigates how user-driven customization of fairness and diversity affects satisfaction in music recommender systems. We developed a prototype allowing listeners to adjust four fairness dimensions: popularity, artist gender, nationality, and genre diversity. In 42 sessions with Dutch participants, interactive controls substantially improved perceived fairness, control, and added value. Genre diversity was most influential, while nationality was least engaged, with gender and popularity falling in between. Findings highlight that parametric design—not algorithmic complexity—drives improved user experience. We show that transparent, customizable fairness levers can make recommendation systems both fairer and more engaging.

Keywords

Music Recommender Systems, Fairness, User driven

1. Introduction

Popular platforms such as Spotify, YouTube, TikTok, and Instagram use recommendation algorithms optimized primarily for engagement, often ignoring users' own fairness and diversity preferences. They rarely offer tools for users to customize the underlying values guiding recommendations [1]. This lack of user control raises concerns about how recommender systems (RS) can better reflect individual expectations and preferences, particularly regarding fairness [2].

Fairness in RS is crucial, as these algorithms shape access to culture, information, and opportunities. While most prioritize accuracy, they often amplify popularity bias and demographic imbalances [3]. Fairness-aware models have been proposed to mitigate these issues, but are typically designed top-down, giving users limited say in defining or influencing what fairness means in practice [4, 5].

These issues are especially clear in Music Recommender Systems (MRS), where subjective tastes meet structural biases like overrepresentation of mainstream artists, underrepresentation of marginalized creators, and limited genre diversity [6, 7]. Most existing MRS address these issues through system-side adjustments or static personalization strategies, but few allow users to directly influence fairness or diversity outcomes [8]. Although prior studies have examined algorithmic bias and collected passive user feedback, they rarely empower listeners to actively shape their recommendation experience [9, 4].

This study investigates whether giving users explicit control over fairness dimensions in MRS enhances perceived fairness, control, and value. We built a prototype where participants select favorite tracks representing their taste, adjust sliders for four fairness dimensions, and finally receive recommendations. This design allows us to test whether parametric fairness controls can meaningfully improve users' listening experiences compared to traditional, system-driven recommendation approaches.

2. Related Work

Fairness and diversity in RS are crucial, especially in cultural domains like music [10, 2]. While research has focused on algorithmic fairness, user perceptions of fairness and their impact on satisfaction remain underexplored. This study explores users' perceptions of user-driven fairness customization in MRS.

MuRS 2025: 3rd Music Recommender Systems Workshop, September 22nd, 2025, Prague, Czech Republic

*Corresponding author.

✉ s.n.khan1@uu.nl (S. N. Khan); j.j.m.nieuwkoop@students.uu.nl (J. Nieuwkoop); j.f.m.masthoff@uu.nl (J. Masthoff)

🆔 0009-0006-4899-2987 (S. N. Khan); 0000-0001-8359-4657 (J. Nieuwkoop); 0000-0001-8099-0515 (J. Masthoff)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Fairness Challenges in Music Recommendation: MRS algorithms often exhibit structural biases, particularly popularity bias [11], which favors mainstream content and marginalizes lesser-known artists, reducing diversity for niche audiences [3]. Female artists are also consistently underrepresented in generated playlists, underscoring the need for demographic equity and systemic representation rather than one-size-fits-all corrections [7, 12, 13]. Most existing fairness interventions operate from a system-driven perspective with little user input, and algorithmic accuracy does not guarantee a fair user experience. This discrepancy between metric-based and perceived fairness [2] highlights the importance of involving users in collaboratively defining fairness [4].

Multi-Dimensional Fairness and Personalization: Fairness in recommendation is increasingly recognized as multi-dimensional, with growing evidence that allowing users to personalize fairness goals across dimensions can enhance satisfaction [5, 14]. Adaptive, participatory systems that respond to user-defined fairness criteria tend to feel more equitable. Re-ranking approaches that incorporate user preferences have been shown to improve the user experience, especially when individuals control how fairness is applied to their lists [15]. Preferences for diversity and popularity vary widely across cultures [16], while heavy reliance on historical genre labels can lead to filter bubbles and repetitive recommendations [6, 17]. Diversity-focused approaches mitigate these issues by incorporating cross-genre exposure and balancing relevance with novelty through multi-objective optimization [18, 19], enriching listening experiences by introducing fresh yet contextually fitting tracks.

User Perception and Fairness Awareness: Perceived fairness is shaped by how familiar and relevant recommendations feel; both excessive familiarity and excessive novelty reduce satisfaction, emphasizing the need for calibrated diversity that balances personal taste with exploration [19]. Subtle interface design interventions—such as making genre visibility adjustable or guiding user interactions—can improve perceived fairness and exploration without reducing comfort [20]. Interactive controls also boost perceived control and trust by making recommendation processes more transparent and interpretable [21].

Algorithmic Decisions and User Trust: Fairness constraints applied without user understanding or input can undermine trust, making transparency and user agency central, particularly in systems affecting demographic exposure [7]. In line with HCI principles of interpretability and participatory design [4], personalized fairness settings enhance user control and reduce dissatisfaction arising from rigid, opaque interventions [9]. While personalization and fairness-aware re-ranking approaches exist, few studies examine their effects on perceived recommendation quality, and the psychological impacts of user-controlled fairness remain underexplored [20, 15]. Offline studies using static fairness settings fail to capture real user experience. Collecting live feedback from users interacting with customizable fairness dimensions offers more actionable insights for system design [2].

This study introduces a system that lets users configure fairness preferences and examines how these choices affect satisfaction and perceived fairness. By emphasizing user agency, it advances a view of fairness as a dynamic user experience rather than a fixed system metric [4, 14, 9].

3. Methodology

We conducted a user study with 42 participants residing in the Netherlands, recruited through convenience sampling, to explore how giving users control over fairness and diversity settings in an MRS affects their satisfaction and listening experience. The study focused on user perception rather than system accuracy. Participants used a simple, locally run web app to try the music recommender. The Ethics and Privacy Quick Scan by the Utrecht University Research Institute of Information and Computing Sciences deemed the research as low-risk, requiring no further assessment.

3.1. Study Setup

The study was conducted in person. The procedure was as follows:

Step 0: Pre-survey. After providing informed consent, participants completed a short survey with their demographics (age, gender, nationality). Next, they were directed to the custom app.

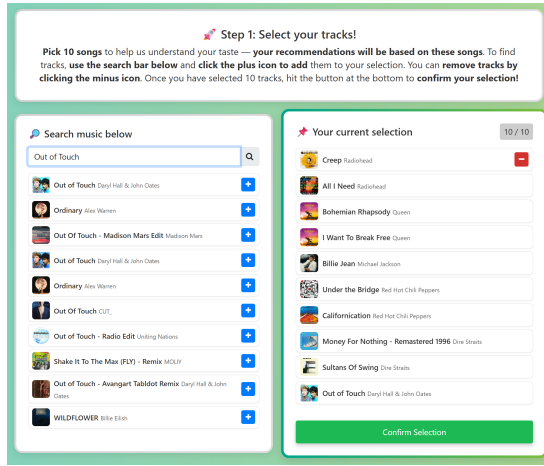


Figure 1: Seed track selection

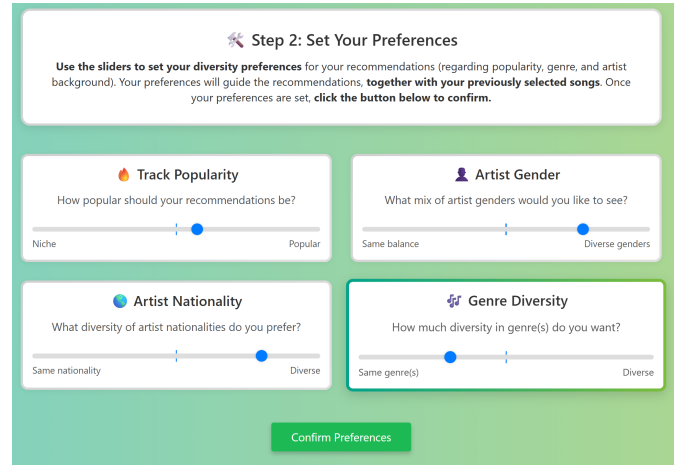


Figure 2: Fairness preference selection

Step 1: Seed track selection. Participants selected 10 seed tracks based on their preferences, as illustrated in Figure 1. The selection is facilitated through a search bar connected to the Spotify API, allowing users to find and add songs by title and artist. Once added, the songs appear in a dynamic list with a counter on the right that displays the number of selected tracks. Participants could listen to an audio preview of maximum 30 seconds per song, similar to a study by Fatahi et al. [22]. Each track can be individually removed, and the “Confirm Selection” button remains disabled until 10 tracks have been added. This number of tracks was chosen to prevent any single track from disproportionately influencing the recommendations. An open-ended search was used instead of a predefined list to promote user autonomy and reduce bias.

Step 2: Setting fairness sliders. After confirming their seed track selection, participants provided their fairness preferences (see Figure 2), using four sliders, each representing a fairness dimension (see selection rationale in Section 3.2). The dimensions used were: track popularity (ranging from niche to popular), artist gender (from the same gender to diverse genders), artist nationality (from the same nationality to diverse), and genre diversity (from the same genre to diverse). Each slider was set to a neutral midpoint by default, allowing participants to express preferences in either direction. Sliders were chosen for their clarity and ease of use. Positioning this customization step after seed track selection reinforces that users actively shape their recommendation experience, setting the expectation that their input will influence the content.

Step 3: Recommendations. After submitting their fairness preferences, the participants were provided with a loading screen informing them that their recommendations were being generated. Once ready, participants were shown ten recommended tracks (Figure 3). Each recommendation was displayed as a card containing the song’s title, the artist’s name, and a preview button linked to the Spotify embedded audio player. This layout supports both evaluation and passive discovery.

Step 4: Post Survey. After listening, participants returned to the survey, where they assessed perceived added value, fairness, and sense of control of our system, and their current experience with MRS in general using 7-point Likert scales, as research emphasizes the importance of subjective RS user evaluations. For more details see Section 3.3. Additionally, they indicated for each fairness dimension whether it was important to them (ticking important ones). This survey completes the “feedback loop” central to the study’s design: users express their preferences, receive tailored recommendations, and evaluate whether the system respected their input. The study’s pacing and clarity were refined through pilot tests to ensure that participants clearly understood each step and could complete the study easily.

3.2. Selection of fairness dimensions

These dimensions were chosen because prior research has identified them as common sources of bias in music recommendation systems as mentioned in Section 2. **Popularity** was chosen as mainstream

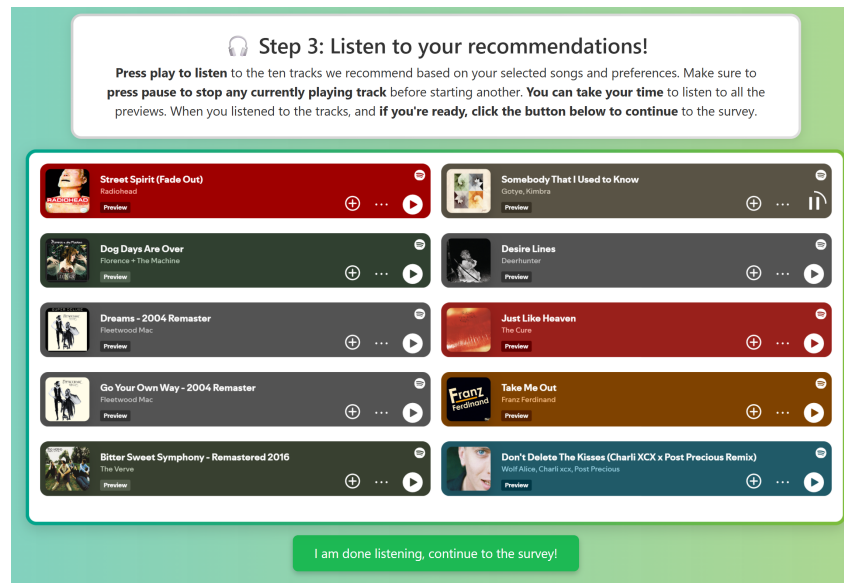


Figure 3: Recommendations screen

hits overshadow niche works [3], **Artist-gender diversity** as female and non-binary artists remain under-represented [7], **Nationality diversity** as local artists' visibility varies by region [16], and **Genre variety** as limited cross-genre exposure widens filter bubbles [6].

3.3. Survey evaluation categories

The survey evaluation categories were chosen based on prior literature on subjective system evaluation. **Perceived Added value** is widely recognized as essential for assessing user experience, capturing not only relevance and enjoyment but also perceived personalization quality [23, 24]. **Perceived fairness** is central to our research, as prior work shows that algorithmic fairness metrics often diverge from user perceptions, making direct user assessment critical [4, 2]. **Sense of control** aligns with our core hypothesis: increasing user agency in shaping recommendations enhances both perceived fairness and satisfaction, leading to a better overall listening experience [21, 24]. Finally, **Current experience** with MRS (in terms of feeling treated fairly and in control) helps determine how our system performs relative to existing systems and highlights whether user-driven fairness customization offers a perceived improvement over mainstream platforms.

Survey items were adapted from validated constructs in Human-Computer Interaction (HCI) and recommender systems research to ensure reliability. Items measuring sense of control draw on established scales addressing user agency and transparency [24], while fairness perception items build on recent studies examining how users interpret and evaluate algorithmic fairness in personalized contexts [4]. The survey can be seen in the supplementary materials¹.

3.4. System Architecture

The MRS back end stores user data, processes seed tracks and fairness preferences, generates recommendations via OpenAI, enriches them with Spotify metadata, and delivers results to the front-end.

Data Storage and Recommendation Generation. User inputs (seed tracks and fairness preferences) are stored locally using JSON, for API compatibility, ease of debugging, and easy integration across system components. After submission, the back-end triggers an asynchronous recommendation process, while the front-end polls the server every few seconds. Results typically appear within 15–30 seconds.

¹<https://doi.org/10.5281/zenodo.17047978>.

API Integration and Validation Logic. Spotify's API (1) supports the seed track search bar and (2) validates GPT-generated recommendations and locates playable tracks. OpenAI's GPT API was selected as it supports structured JSON outputs, can be steered with detailed prompts, integrates well with Python, and allows incorporation of user-specified fairness dimensions.

Prompt Engineering with Fairness Dimensions. Prompts sent to GPT include seed tracks and normalized (0–100) values for popularity, gender diversity, nationality diversity, and genre diversity (e.g. 0 to 10 = Only very niche/obscure/underground tracks, so likely different from seed tracks). Detailed instructions guide GPT to respect these preferences while avoiding overfitting, hallucination, or invalid outputs. GPT is instructed to return exactly ten tracks in a fixed JSON format, including title, artist, and genre. Prompt design required extensive iteration. Early prompts led to extremes—recommending only frequent seed artists or generating entirely unfamiliar, overly diverse results. Adjustments were made to balance diversity with musical identity across genres and configurations. The prompts are available in the supplementary materials².

4. Results

In this section, first we report participant characteristics and the reliability of the measurement scales. Second, we analyze subjective evaluations, including perceived fairness, control, added value, and overall experience. Third, we examine behavioral preference measures derived from the fairness sliders and consider the importance ratings and the alignment between preferences and importance.

4.1. Demographics

Among the 42 participants, 25 (59.5%) identified as male and 17 (40.5%) as female, providing a modest male majority. Ages ranged from 19 to 25 years ($M = 20.8$, $SD = 1.6$), reflecting a predominantly early-twenties, university-age sample, making the findings most applicable to young adult listeners.

4.2. Reliability of Scales

To ensure the questionnaire reliably measured each category, we used **Cronbach's alpha** (α) with higher values (closer to 1) indicating stronger internal consistency. The Perceived Fairness scale showed excellent reliability ($\alpha = 0.95$), while Perceived Control was acceptable ($\alpha = 0.79$). The Added Value/Impact scale had marginal reliability ($\alpha = 0.69$), and the Current Experience scale showed low consistency ($\alpha = 0.57$), suggesting its items may capture varied aspects or be interpreted inconsistently.

4.3. Subjective user evaluations

Category Comparison & Correlation. Figure 4 shows the average scores across the four categories. Added Value scored highest (mid-5s to almost 6), suggesting users found the customization options beneficial. Perceived Control followed with (around 5.5), indicating participants felt confident adjusting the system. Perceived Fairness received moderate ratings (around 5), while Current Experience scored about 3.5, suggesting traditional MRS are perceived worse than our system in terms of fairness and control.

Correlations (Table 1) reveal a strong positive relationship between Perceived Fairness and Perceived Control ($r = .84$), suggesting fairness and agency are closely linked in shaping the user experience. Fairness also correlated strongly with Added Value ($r = .72$), and Control with Added Value ($r = .69$), indicating that both fairness and control enhance perception of the system's usefulness. In contrast, Current Experience was not significantly correlated with any other construct, suggesting that satisfaction with MRS in general was independent of perceptions of our system's fairness, control, and added value.

²<https://doi.org/10.5281/zenodo.17047978>.

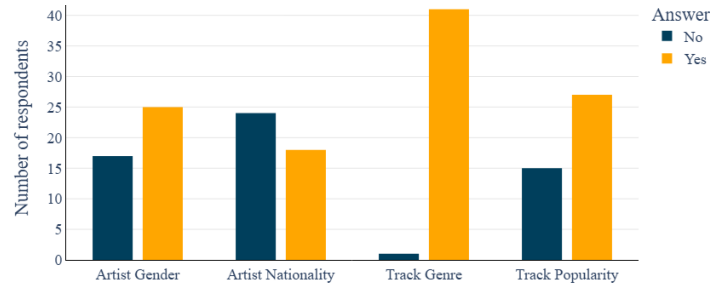


Figure 6: Importance of fairness dimensions.

4.4. Behavioral Preference Measures

Correlations Between Fairness Dimensions. Figure 5 summarizes mean slider positions (0–100) across participants: Popularity ≈ 45 , Genre ≈ 22 , Gender ≈ 43 , and Nationality ≈ 49 . All averages fall below the neutral midpoint (50), with Genre notably lowest and Nationality nearest to neutrality.

Correlations between slider values (Table 2) show the strongest link between Gender and Nationality ($r = .39$), indicating that participants favoring gender diversity also diversified by nationality. Moderate correlations emerged for Genre-Gender ($r = .34$), Popularity-Gender ($r = .33$), and Genre-Popularity ($r = .31$). Popularity-Nationality was not significant, suggesting mainstream versus niche choices were independent of nationality preferences. These patterns reveal “global fairness” tendencies in some participants, while others adjusted only one dimension, supporting the value of modular controls.

Importance of Dimensions. Figure 6 shows participants’ post-survey responses on whether they considered each dimension important, based on “Yes” (important) vs “No” (not important) answers. Nearly all judged track genre as important, with only one dissenting response. Track popularity followed, valued by roughly two-thirds of participants. Views on artist gender were more divided, though a slight majority marked it as important. Artist nationality ranked lowest, with more deeming it unimportant than important. These preferences contextualize slider behaviors by clarifying which fairness levers users consciously prioritized.

Preference Behavior vs. Ratings. Figure 7 compares the slider deviations from the midpoint (50) between participants who judged a dimension “important” versus “not important”. Greater deviation reflects stronger engagement, regardless of direction. Across all dimensions, “important” participants consistently moved sliders farther from 50, while others stayed near default. Patterns varied: for Popularity, importance was associated with widely ranging adjustments, whereas non-importance yielded near-zero deviations. For Genre, nearly all “important” participants moved the slider substantially, but the low overall average (22; Figure 5) indicates only a limited desire for genre exploration. Gender showed the strongest polarization: with extreme shifts for importance and minimal change for non-importance. Nationality followed a similar but less pronounced pattern.

Mann–Whitney U tests confirmed that absolute slider deviations were significantly greater among participants who rated a dimension as important for Popularity ($U = 83.0, p = .002$), Gender ($U =$

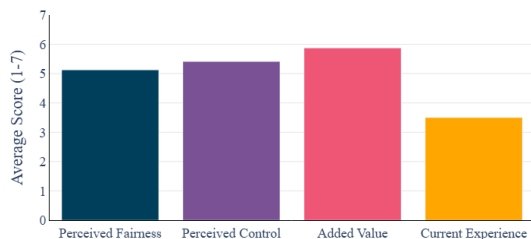


Figure 4

Average survey scores per category.

Scale	Fairness	Control	Value	Experience
Fairness		0.84***	0.72***	0.13
Control	0.84***		0.69***	0.03
Value	0.72***	0.69***		-0.01
Experience	0.13	0.03	-0.01	

Table 1

Pearson correlations between categories, *** means $p < .001$.

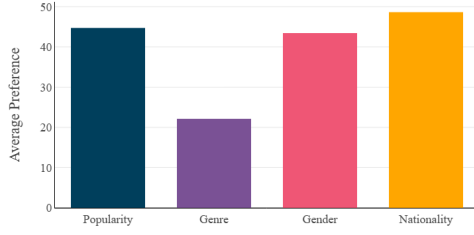


Figure 5
Average slider positions for fairness dimensions (0=homogeneity, 100=diversity)

	Popularity	Genre	Gender	Nationality
Popularity		0.31***	0.33***	0.15
Genre	0.31***		0.34***	0.27***
Gender	0.33***	0.34***		0.39***
Nationality	0.15	0.27***	0.39***	

Table 2
Pearson correlations between fairness dimensions,
*** means $p < .001$

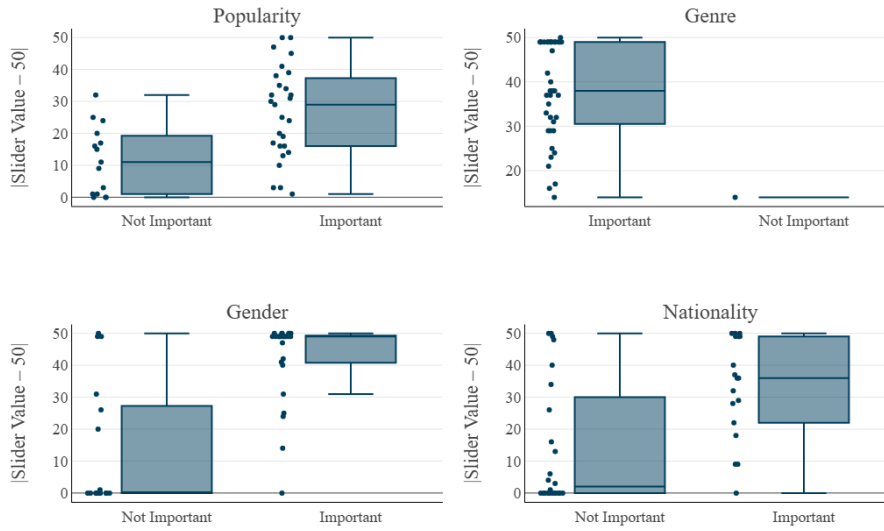


Figure 7: Absolute slider deviations from midpoint (50) by self-reported importance.

66.5, $p < .001$), and Nationality ($U = 98.5$, $p = .003$). For Genre, the difference was not statistically significant ($U = 0.5$, $p = .099$) despite a large effect size, due to the very low number of “unimportant” participants. Rank-biserial correlations (r) indicate moderate-to-large effects for all dimensions.

5. Discussion

Our findings indicate that transparent, parametric controls substantially enhance user perceptions of fairness and control, while also revealing that fairness priorities vary widely across individuals.

Main Findings. Participants reported high levels of perceived fairness ($M \approx 5.0$, $\alpha = .95$) and control ($M > 5.5$, $\alpha = .79$), with the two constructs strongly correlated ($r = .84$). This suggests that in interactive recommender settings, fairness is closely tied to agency: users who felt empowered to shape recommendations were also those most likely to perceive the system as fair. Both categories were also positively associated with the sense that the system added value to the listening experience, with fairness correlating at $r = .72$ and control at $r = .69$. These effects occurred despite the underlying recommendation system being relatively simple, reinforcing that the parametric design—not algorithmic complexity—drove perceptions of value. The current experience category, which reflects satisfaction with traditional MRS was not significantly correlated with any of these categories, and its lower average indicates participants preferred our system in terms of fairness and control.

Not all fairness dimensions were weighted equally. Genre diversity emerged as the most universally important: nearly all participants considered it relevant and adjusted its slider. Nationality diversity, by contrast, was less important, with many participants leaving this setting unchanged; however, those

who did prioritize it moved the slider significantly more ($U = 98.5$, $p = .003$, $r = -0.54$). Gender and popularity fell between these extremes. For both, participants who rated these dimensions as important made deliberate adjustments, with significant differences for gender ($U = 66.5$, $p < .001$, $r = -0.69$) and popularity ($U = 83.0$, $p = .002$, $r = -0.59$). These patterns suggest that while some fairness concerns are widely shared (e.g., genre), others reflect more individual preferences.

Correlations between fairness dimensions revealed modest associations, such as gender with genre ($r = .34$) and popularity ($r = .33$), and a somewhat stronger relationship between gender and nationality ($r = .39$). This points to a general fairness orientation among some users but also indicates sufficient independence between dimensions to justify treating them as conceptually distinct.

Overall, these findings indicate that while fairness preferences are diverse and personal, enabling users to express them explicitly improves their experience and perception of the recommender.

Connection with Existing Literature. The present findings resonate strongly with prior research on user-centered fairness and transparency in RS. Much of the literature has focused on algorithmic strategies to improve fairness, but recent work increasingly emphasizes user agency as a determinant of fairness perceptions [5, 8]. The strong correlation between perceived fairness and perceived control ($r = .84$) supports claims by Dinnessen and Bauer [4] and Burke [5] that fairness is not simply a statistical outcome but a lived, interpretable experience tied to user influence. This aligns with broader principles in fairness-aware recommendation, where procedural justice—the ability to influence the process—often matters as much as distributive outcomes [9, 25].

Unlike many prior studies that rely on passive feedback or post hoc surveys, this work used interactive sliders, enabling users to modulate dimensions such as popularity, gender, nationality, and genre in real time. This parametric design approach responds to recent calls for more transparent fairness mechanisms [14, 21], particularly in music domains where taste and identity are tightly coupled [6, 7]. Whereas traditional re-ranking methods operate invisibly, the interface foregrounded user choices, aligning with Liang and Willemsen’s [20] concept of “fairness nudges”. The finding that fairness and control correlated strongly with added value ($r = .72$, $r = .69$) emphasizes that users derive satisfaction not only from outcomes but also from their ability to shape them.

The dimension-specific patterns observed here also connect with earlier findings. Genre diversity’s near-universal importance echoes evidence that stylistic variety, when aligned with personal taste, enhances satisfaction and engagement [17, 26]. Nationality, conversely, received the least engagement, consistent with studies showing mixed user sensitivity to cultural or geographic diversity in recommendations [16]. The diminished significance of nationality can be partially attributed to the phenomenon of glocalisation [27]. Gender and popularity, while less broadly endorsed, elicited substantial adjustments from those who prioritized them, reinforcing the significance of individual fairness preferences in interactive RS [28, 3]. These results complement population-level work on bias and diversity in MRS [10, 7, 6] by offering a more granular perspective on the dimensions users themselves are most inclined to control.

Finally, these findings raise questions about the long-term effects of user-controlled fairness. This study focused on a single interaction; future work could examine whether such control mechanisms sustain trust, satisfaction, and diversity over repeated sessions [1, 8]. Expanding the controllable dimensions (e.g., lyrical content, mood, or tempo) or tailoring sliders to specific user segments (e.g., minority or non-mainstream listeners [29]) could further enhance perceived relevance and fairness. Collectively, these results contribute to the growing body of research that views fairness not as a fixed property of recommendations, but as a dynamic, user-driven process.

Limitations. The Fairness ($\alpha = .95$) and Control ($\alpha = .79$) scales showed strong reliability, but Added Value ($\alpha = .69$) and Current Experience ($\alpha = .57$) were less reliable. The study used a homogeneous sample of young Dutch adults, limiting generalizability. The design lacked a control condition without fairness sliders, making it unclear how much customization drove the observed benefits. The study tested only four fairness dimensions. Finally, only a single, brief interaction was studied. Longitudinal research is needed to examine whether benefits of fairness slider persist or change over time.

6. Conclusion

This study explored how user-driven customization of fairness and diversity shapes satisfaction in MRS. Allowing listeners to adjust four fairness dimensions improved their experience, showing that parametric design - rather than algorithmic complexity or default settings - drives perceived value.

Not all fairness dimensions held equal weight. Genre diversity was most frequently prioritized, while nationality diversity drew the least engagement, though valued by a subset of users. Gender and popularity fell between, with notable slider adjustments among those rating them as important. Correlations among fairness dimensions (e.g., gender and nationality) suggest some users adopt a broad fairness orientation, while others focus on specific axes.

These findings show that modular, user-centric fairness controls bridge the gap between algorithmic metrics and lived experience. By emphasizing user agency, such systems enhance perceived fairness, control, and added value. This parametric approach supports viewing fairness as a dynamic negotiation rather than a fixed rule [14, 4]. Platforms should consider offering similar controls, tailoring available dimensions to individual preferences. Future research should refine measures, diversify samples, add control conditions, expand fairness axes, and assess effects longitudinally.

In sum, transparent, user-driven fairness controls offer a promising path to make MRS both equitable and engaging. By combining live customization with fairness-aware algorithms, MRS can meet fairness goals while accommodating diverse, subjective listening preferences.

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] W. Wang, F. Feng, L. Nie, T.-S. Chua, User-controllable recommendation against filter bubbles, in: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22*, ACM, 2022, p. 1251–1261. URL: <http://dx.doi.org/10.1145/3477495.3532075>. doi:10.1145/3477495.3532075.
- [2] 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>.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] D. Kowald, M. Schedl, E. Lex, The unfairness of popularity bias in music recommendation: A reproducibility study, in: *Advances in Information Retrieval*, Springer International Publishing, Cham, 2020, pp. 35–42. URL: https://doi.org/10.1007/978-3-030-45442-5_5. doi:10.1007/978-3-030-45442-5_5.

- [7] 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*, Association for Computing Machinery, New York, NY, USA, 2021, p. 249–254. URL: <https://doi.org/10.1145/3406522.3446033>. doi:10.1145/3406522.3446033.
- [8] D. Jannach, S. Naveed, M. Jugovac, User control in recommender systems: Overview and interaction challenges, in: *E-Commerce and Web Technologies*, Springer International Publishing, Cham, 2017, pp. 21–33.
- [9] 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.
- [10] 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.
- [11] O. Celma, *Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space*, Springer, 2010. doi:10.1007/978-3-642-13287-2.
- [12] 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.
- [13] D. Shakespeare, L. Porcaro, E. Gómez, C. Castillo, Exploring artist gender bias in music recommendation, 2020. URL: <https://arxiv.org/abs/2009.01715>. arXiv:2009.01715.
- [14] R. Burke, N. Mattei, V. Grozin, A. Volda, 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.
- [15] N. Sonboli, F. Eskandarian, 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.
- [16] 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.
- [17] M. Kunaver, T. Požrl, Diversity in recommender systems – a survey, *Knowledge-Based Systems* 123 (2017) 154–162. URL: <https://www.sciencedirect.com/science/article/pii/S0950705117300680>. doi:<https://doi.org/10.1016/j.knosys.2017.02.009>.
- [18] 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, SAC '17*, Association for Computing Machinery, New York, NY, USA, 2017, p. 1693–1696. URL: <https://doi.org/10.1145/3019612.3019899>. doi:10.1145/3019612.3019899.
- [19] 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.
- [20] 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.
- [21] M. Jesse, D. Jannach, Digital nudging with recommender systems: Survey and future directions, *Computers in Human Behavior Reports* 3 (2021) 100052. doi:10.1016/j.chbr.2020.100052.
- [22] 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, UMAP '23*, ACM, New York, NY, USA, 2023, p. 66–77. URL: <https://doi.org/10.1145/3565472.3592958>. doi:10.1145/3565472.3592958.

- [23] P. Pu, L. Chen, R. Hu, A user-centric evaluation framework for recommender systems, in: Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11, Association for Computing Machinery, New York, NY, USA, 2011, p. 157–164. URL: <https://doi.org/10.1145/2043932.2043962>. doi:10.1145/2043932.2043962.
- [24] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, C. Newell, Explaining the user experience of recommender systems, *User Modeling and User-Adapted Interaction* 22 (2012) 441–504. URL: <https://doi.org/10.1007/s11257-011-9118-4>. doi:10.1007/s11257-011-9118-4.
- [25] A. Singh, T. Joachims, Policy learning for fairness in ranking, in: Advances in Neural Information Processing Systems 31 (NeurIPS 2018), 2019. URL: <https://arxiv.org/abs/1902.04056>. arXiv:1902.04056.
- [26] L. Porcaro, E. Gómez, C. Castillo, Assessing the impact of music recommendation diversity on listeners: A longitudinal study, 2022. URL: <https://arxiv.org/abs/2212.00592>. arXiv:2212.00592.
- [27] W. Page, C. Dalla Riva, 'Glocalisation' of Music Streaming within and across Europe, Technical Report, EIQ Paper, 2023.
- [28] 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.
- [29] S. Khan, E. Herder, D. Kaya, Experiences of non-mainstream and minority users with music recommendation systems, 2024. doi:10.18420/muc2024-mci-ws11-141.