Understanding Individual and Collective Diversity of Cultural Consumption through Large-Scale Music Listening Events

Harin Lee[1,2,3,4,*], Romain Hennequin[1,†] and Manuel Moussallam[1,†]

[1]Deezer Research, France
[2]Max Planck Institute for Human Cognitive and Brain Sciences, Germany
[3]Max Planck Institute for Empirical Aesthetics, Germany
[4]Faculty of Life Sciences, Leipzig University, Germany

Abstract
Emerging research suggests that cultural richness and complexity intensify with population size. Yet the mechanism underlying this phenomenon remains unclear: Do populated areas exhibit more cultural diversity simply due to there being a larger spectrum of individuals with varied backgrounds, or does the urban environment itself stimulate individuals to explore a wider variety of cultural experiences, raising the population's baseline? To decipher this, we leverage a large-scale dataset of 69 million music listening events from the real world, examining the listening patterns of over 408 thousand unique individuals across 96 regions in France. Our study presents a dual perspective on diversity by (1) measuring one's diversity of musical consumption by evaluating the breadth of their music listening history, and (2) assessing the shared repertoire among individuals as a collective. We found that both individual and collective levels of musical consumption diversity increase with population size. This trend held true when segmenting the population by gender and age groups, while a gender-specific divergence in consumption appeared from a particular age. We further delineate potential confounding variables to consider in future research aimed at identifying causal pathways, presenting this model using a Directed Acyclic Graph (DAG). Together, our preliminary work represents a crucial step towards unravelling the complexity of cultural diversity and its ties to population dynamics.

Keywords
cultural diversity, population size, music consumption, cultural evolution

1. Introduction

With our world becoming increasingly interconnected, the incredible variety of human cultures emerges as a prominent aspect of our collective existence. Throughout history, from primal cave paintings to the triumph of modern architecture, every artefact, technology, and tradition paints a unique hue onto our collective cultural spectrum.

CHR 2023: Computational Humanities Research Conference, December 6 – 8, 2023, Paris, France
*Corresponding author.
†These authors contributed equally.
hlee@cbs.mpg.de (H. Lee); rhennequin@deezer.com (R. Hennequin); manuel.moussallam@deezer.com (M. Moussallam)
www.harinlee.info (H. Lee)
0000-0001-5579-4547 (H. Lee); 0000-0001-8158-5562 (R. Hennequin); 0000-0003-0886-5423 (M. Moussallam)

Contemporary research has begun to apply computational techniques to understand the constant flux of cultures, translating it into discernible patterns and delving into the intricate processes that shape our shared identity. A plethora of methods have been utilized in this endeavour, including computer simulations [6, 7], experimental studies [8, 24, 17, 11], and studying the patterns of archaeological records [5, 13, 18]. Together, these research approaches shed light on the underlying mechanisms for cultural richness and diversity, with population size and cultural exchanges (particularly via immigration) suggested as the main drivers. The larger and more diverse a population, the greater the chance for cultural interaction, sparking novel ideas and developments [14, 15]. Similarly, immigration can act as a vibrant channel for cross-cultural exchange, introducing fresh perspectives to a culture, which fosters its evolution and enhances its complexity [6, 27].

However, empirically testing these relationships remains a significant challenge since understanding the complex nature of cultural dynamics requires data that is high in temporal and spatial resolutions, capturing both micro and macro levels. Computational models provide valuable theoretical underpinning but often portray speculative scenarios that necessitate real-world data for validation [9]. Our study addresses this gap by assembling and analyzing an extraordinarily large dataset of music listening events, featuring 69 million streams from over 408 thousand unique individuals living in France. This finely-grained behavioural data, both temporally and spatially, provides a natural experiment to investigate the influence of population size and demographic composition on cultural diversity at both individual and collective scales. Music serves as an ideal test-bed for examining such phenomenon, with its universal prevalence [22] and regional distinctiveness [16, 21], and its reflection of the socio-cultural spirit of the times [10].

Just as an ecologist might analyze the biodiversity of species within a particular habitat, we study individuals' music-listening behaviour within specific geographical regions to understand the diversity of musical consumption. According to prior research that observed elevated levels of diversity with population size in various domains, including language [28] and music [30], we would anticipate finding more musical diversity in urban and metropolitan areas. However, the underlying mechanisms behind such increased levels of cultural diversity remain puzzling and subject to debate.

In one hypothetical scenario, increased cultural diversity in densely populated areas, like capital cities, might stem from the convergence of diverse individuals that form a cultural melting pot. In assessing the diversity of these regions, we may register increased diversity simply due to the higher likelihood of encountering individuals from varied backgrounds. Here, the observed regional diversity would then largely be attributed to demographic diversity, meanwhile the extent of an individual's breadth of consumption may be relatively isolated from regional influences; instead, constrained by the general capacity of musical variety one might engage with.

Alternatively, the catalysts for diversity might lie within principles of consumer behaviour and exposure. Inhabitants of populous regions, continuously surrounded by a variety of cultural offerings and privileged with abundant access to cultural events, might gradually incorporate this variety into their personal tastes. This continuous exposure to, and easy access to diverse cultural manifestations could subtly mould individual preferences, encouraging exploration and adoption of a broader range of cultural experiences. Over time, these influences might alter the population's baseline, leading to a higher level of diversity in dense metropolitan areas.

To this end, our study seeks to untangle the potential pathways towards cultural diversity, considering diversity at two levels: (1) an individual's diversity of musical consumption based on their listening profile, and (2) a collective's diversity based on the alignment of consumption between individuals. We present preliminary findings on how demographic factors and population size play a role on these two accounts. Importantly, we outline our future plans to go beyond mere associations and comprehend the causal mechanisms shaping cultural diversity. For this purpose, we have collected data on potential confounding variables and represent the model we aim to test with a Directed Acyclic Graph (DAG). This paper thus sets the stage for a deeper understanding of the interplay between demographics and consumption behaviour, and their influence on cultural diversity. To benefit the community and future research, we plan to openly make available the aggregated data at the regional level and socio-demographic data gathered from various sources, along with all the statistics computed for the analysis.

2. Method

2.1. Music Streams

Our research was conducted on Deezer, a globally available music streaming platform present in 187 countries, boasting a catalogue of over 90 million tracks. The user's listening behaviour is captured with comprehensive logs, including the date and time of song playback, duration, the listener's self-reported age and gender, preferred language, type of streaming device and internet connection, and geographical coordinates derived from the IP address.

Initially, we gathered all music streaming events that happened in France between April 3rd and 30th, 2023, a four-week period ensuring a balanced representation of weekdays. To reduce noise and potential biases in our data, we implemented various stringent filters. First, we limited our study to tracks played longer than 30 seconds. Second, we considered streams solely over local area networks (LAN) for robust geolocation. Finally, since our study relies on accurate user-to-region attribution, we developed an additional filter to exclude users with large travel radius. To do this, we first derived each user's listening locations using latitude and longitude data extracted from their IP addresses, excluding users with more than five distinct IP addresses. We then determined each user's travel radius over the four-week study period by calculating the haversine distance between all pairs of their geographical coordinates. We applied a cut-off at a median distance of 50km (see Figure 1B), which was a reasonable threshold to strike a balance between geographical specificity and retaining a substantial portion of the user sample. This process excluded 32% of the users and we translated the remaining users' geographical coordinates into NUTS3 level[1] regions using the 'gioscoR' R package, which formed the basis of our regional groupings. To test if these geographical filtering might bias the result (e.g. users who travel far to work from home or students who live away from home), we made another sample of users without the geolocation filters, which replicated the patterns reported

[1]Nomenclature of Territorial Units for Statistics or NUTS is a geocode standard for referencing the administrative divisions of countries for statistical purposes. https://ec.europa.eu/eurostat/web/nuts

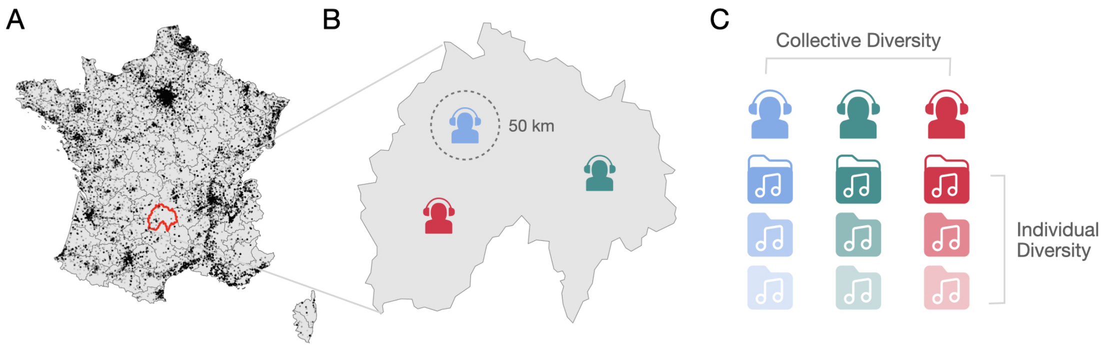

Figure 1: (A) Each dot on the map of France represents one of the total 408K users included in the study. (B) An illustration of our method for assigning a user to a specific NUTS3 region based on the criterion of 50km median travel distance. (C) Collective diversity is computed by sampling streams across users in a given region, while individual diversity is calculated based on the user's listening history and their dispersion within the general song embedding.

here.

Our final dataset encompassed over 69 million unique listening events from 408 thousand individual users (see Figure 1A for all users projected on the map). This data was augmented with statistics on regional population size, median age, and gender ratio gathered from Eurostat.[2] When compared across the regional population of France, our dataset's demographics exhibited strong correlations in terms of regional population size (rho = 0.97, p < .001) and median age (rho = 0.60, p < .001). However, when comparing to the entire French population as a whole, there were 25% more young users on Deezer (between ages 15-30) while 30% less among older (between ages 51 to 80), with the mid-age group being similar in their proportions (3% more on Deezer user base). Regarding self-reported gender, the Deezer users were skewed towards the male demographic (34% more males) compared to France's general population (6% more females), with no significant correlation found when comparing gender ratios across regions (rho = 0.06, p = 0.53). This suggests that, while our dataset effectively represents distribution across France's region in terms of user numbers and age, platform-based age and gender biases exist.

2.2. Measuring Diversity

In this study, we draw two perspectives on musical consumption diversity, at the individual (within-user) and collective (between-users) levels (see Figure 1C). Given that sample size can bias the estimates of diversity [3], all statistics we report in this study standardize the sample size at every level of analysis through random sampling with replacement. We apply this procedure across 1,000 users and 10,000 music streams when conducting individual and collective level analyses, respectively. We then aggregate across 100 bootstrap iterations of this sampling process to gain reliable estimates of the mean and variance.

[2]Eurostat is the statistical office of the European Union, publishing Europe-wide statistics and indicators that enable comparisons between countries and regions.

2.2.1. Individual Diversity

To assess an individual's diversity of musical consumption, we employed the Generalist-Specialist score (GS-score), a previously validated metric in user music exploration and discovery studies [1, 23]. The GS-score of a user (u_i) is computed by taking the cosine similarity between a given user's mean vector $\vec{\mu}_i$ (calculated from all the songs the user consumed within the last 28 days) and the vector representation of a randomly selected song ($\vec{s_j}$) they have listened to, weighted by the number of times they listened (w_j). These vectors are extracted from high-quality SVD song embeddings [19] computed based on the co-occurrence of millions of songs derived from user-generated playlists and listening behaviour [2]. Formally, the equation for GS-score can be written as:

$$GS\left(u_i\right) = \frac{1}{\sum w_j} \sum_j w_j \frac{\vec{s_j} \cdot \vec{\mu}_i}{\|\vec{s_j}\| \cdot \|\vec{\mu}_i\|} \tag{1}$$

The GS-score effectively captures the dispersion of a user's listening behaviour in the general song embedding space, whereby a *specialist* would have a more focused preference (e.g. only listens to Jazz), whereas a *generalist* would exhibit a broader range of listening habits (e.g. listens to Jazz, Classical, and Metal). To make this score consistent in the direction of our collective diversity measure, we inverted the score (1 - GS(u_i)) and normalized it to a range between 0 to 100, where 100 represents maximal diversity.

2.2.2. Collective Diversity

To quantify the diversity found for a given population, existing research has commonly applied measures like the Gini coefficient, Simpson's index, or Shannon's entropy. However, these measures provide arbitrary scales, making comparisons challenging. As a solution to this, Hill's number (also commonly known as the effective number of species) has become a method increasingly popular to quantify the diversity of an assemblage, which allows for standardized comparisons [3, 7]. By treating each song as a 'specie' in a given population, we calculate collective diversity of consumption as *effective number of songs*(${}^qD$), expressed as:

$${}^qD = \left(\sum_{i=1}^{S} p_i^q\right)^{1/(1-q)} \tag{2}$$

Here, q defines the order of the Hill's number, where higher values of q emphasize the contribution of rare songs, while lower values of q focus on the abundance of popular songs. In our analysis, we set the order of q to be 1, which essentially becomes the exponential of the familiar Shannon index. The S represents the total number of unique songs, and p signifies the relative abundance of each song.

3. Result

3.1. Individual's Musical Consumption

We begin by examining the dynamics of an individual's musical consumption, and how these may be influenced by age, gender and geographical environment they live in.

We first assessed the evolution of one's musical consumption diversity with respect to self-reported age and gender. Our findings were consistent with established literature showing that one's musical consumption consolidates mostly during adolescence and less exploration happens with ageing [23, 20, 1, 31]. The results illustrated in Figure 2A clearly demonstrate that one's consumption diversity increases and converges around their mid-20s and then follows a gradually declining trend from the 30s. Interestingly, we observed that this peak is also the point of gender-based divergence, whereby male listeners display a slower rate of decline compared to females, with group differences remaining relatively consistent over time. This suggests that female listeners become more rapidly specialized in their consumption from mid-age. In fact, other studies have also found more musical diversity among male users [25], and some speculate that this is due to males considering mainstream music to be *unhip* [4]. This is in contrast, however, with a recent study [1] that found no gender differences in the GS-score of Spotify users. Yet this study did not evaluate the interaction of age and gender in their group-level comparisons, which might explain the discrepancy. Our data suggest that gender differences may only become apparent after a certain age, which could have potentially been overlooked in an aggregated group-level analysis. Alternatively, this divergence might be a phenomenon specific to Western culture, or a bias introduced from the embedding space where we derive the user's diversity metric (i.e. embedding space has more male representation given the gender imbalance in the user base). Thus, it necessitates further investigation to determine if the same pattern can also be replicated in other cultures and make comparisons with other measures for diversity that do not rely on embedding space.

Our next analysis focused on the impact of population size, established by the number of users in each NUTS3 region. We observed a positive and substantial correlation ($rho = 0.59$, $r = 0.49$, $ps < .001$) on individual's consumption diversity as a function of population size, indicating that individuals residing in densely populated regions tend to exhibit a more diverse consumption behaviour. To examine whether the effect is uniformly present across different gender and age groups, we subdivided the population into *younger* (age 18-30) and *older* (30+), further stratified by gender. Across all subgroups, we found an overall increasing trend in diversity with population size when fitting the data with the generalized additive model (GAM) (see Figure 2B). GAM is a statistical model that extends the concept of the generalized linear model by allowing for non-linear relationships between the dependent variable and the independent variables. GAMs can capture complex relationships by using smooth functions, such as splines, to model the non-linear components. Together, these results indicate that both demography and geographical environment play crucial roles in shaping one's breadth of musical consumption.

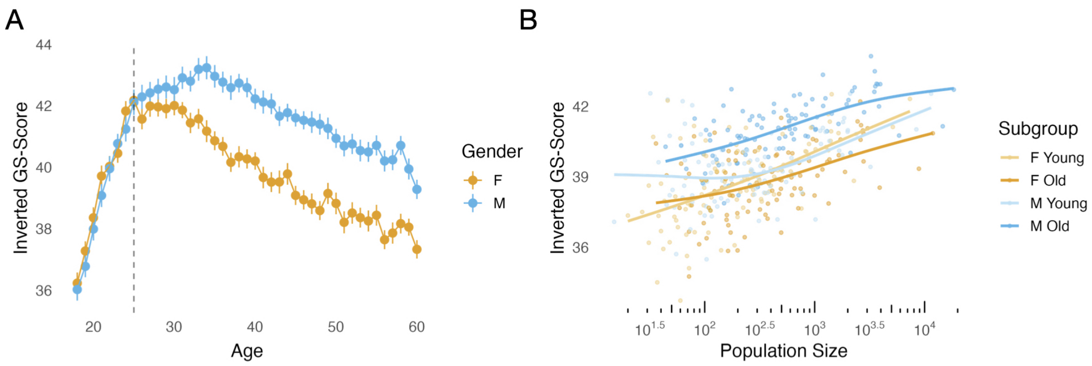

Figure 2: Individual's diversity of musical consumption measured with inverted GS-score. (A) Effect of age and gender with a dashed line indicating the point of divergence between male and female at 25 years of age. (B) Effect of population size, sub-grouped by age and gender. Each dot represents the mean across 100 bootstrap samples of 1,000 random users. Error bars denote one standard deviation from the mean, and fitted curves represent estimates produced via GAM.

3.2. Collective Musical Consumption

We now broaden our analysis to the diversity of musical consumption across individuals, effectively investigating how similar or distinct one's musical listening event might be from one another within the same group (i.e. collective). We have previously assessed an individual's consumption diversity trajectory as a function of age and gender. Here, we group individuals of the same gender and age, analyzing their co-consumption of music (i.e. the shared repertoire and frequency of songs within the group). High co-consumption indicates shared consumption, whereas low overlap points to unique consumption across individuals. The skewness or uniformity of this distribution serves as our diversity metric, determined using the effective number of songs (see Equation 2).

Age and gender effects on collective consumption diversity reflected similar trends to those observed in individual diversity (see Figure 3A). Much like individual diversity, between-user diversity increased with a notable surge at age 23 and peaking around the mid-30s (in other words, when gathering a group of mid 30 years old together in one room, they will have the least amount of shared repertoire of songs). We once again identified gender-based differences, with female listeners displaying more shared consumption within their age group than their male counterparts. Additionally, collective consumption diversity declined with age among female listeners, while it plateaued among males from the mid-30s onwards.

In alignment with the higher individual diversity we observed in more populated regions, we also observed an increasing trend for collective diversity with population size (*rho* = 0.96, *r* = 0.52, *ps* < .001), demonstrating a remarkably clear concave downwards curve across all four subgroups (see Figure 3B). This heightened diversity in populous regions becomes distinctly evident when examining the distribution of users in the user embedding, where users with similar listening habits are closer together in this embedding space. As depicted in Figure 3C, users from the capital area *Paris* are scattered throughout various parts of the space, implying

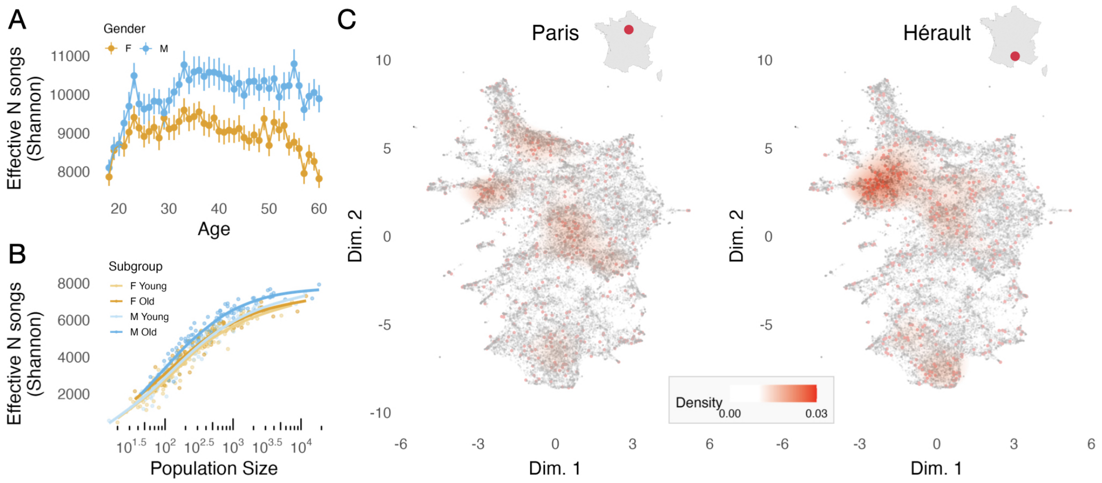

Figure 3: Comparison of collective diversity (A) grouped by age and gender, and (B) population size subdivided into gender and age groups. Diversity is estimated as the effective number of songs from 100 bootstraps of 10,000 randomly sampled music streams with replacement. Error bars denote one standard deviation from the mean, and fitted curves represent estimates produced via GAM. (C) A visualization of user embedding space in a two-dimensional UMAP projection. The user density is illustrated by the raster plot overlay. Notably, users from Paris exhibit a more dispersed distribution, indicating a broader range of users with more distinct listening habits. By contrast, users from smaller regions, such as Hérault cluster more closely, signifying a more concentrated music preference.

a higher individuality in their music consumption. In contrast, music listeners from a more rural region such as *Hérault* cluster more closely together, suggesting a more shared musical listening habit.

3.3. Underlying Mechanisms for Cultural Diversity

Our analysis reveals an elevated level of diversity in musical consumption in densely populated areas. This enhanced diversity is apparent not only at the individual level, but also within the broader spectrum of the region's population. Despite these findings, unravelling the underlying mechanisms that foster these varied consumption patterns and understanding their complex intertwined relationships is challenging. The strong link we observe between consumption diversity with increasing population size could be influenced by various factors, ranging from the confluence of diverse ethnic groups, and socioeconomic inequalities in accordance with the 'cultural omnivore theory' [26, 12], to the general expansion of varied consumption among individuals spurred by their increased exposure to a wide array of cultural products.

Determining causal relationships among these potential factors would ideally necessitate carefully controlled lab experiments. However, in the context of our study and many others examining social phenomena, it is often not feasible nor ethical to design experiments manipulating individuals' cultural environments. An alternative approach is to rely on observational data while thoroughly considering potential variables that could be confounding and carefully

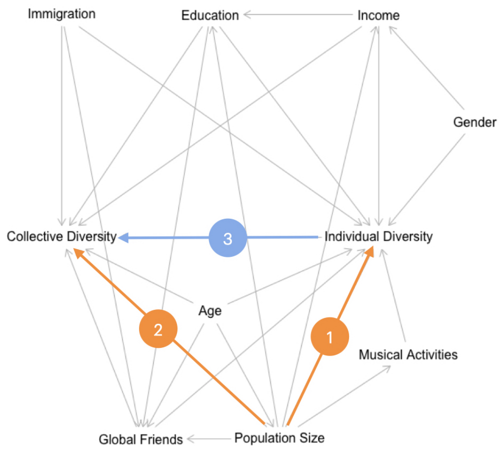

Figure 4: Proposed model represented with DAG to assess the causal pathways for individual and collective diversity of cultural consumption. The three main pathways of interest are highlighted with the first and second testing the influence of population size on individual and collective diversity, and the third testing the influence of individual's on the collective's diversity.

controlling for them with the available modern tools [29].

To this end, we have collected data to account for various potential factors mentioned in existing literature. These data were extracted and augmented from various external sources: (i) self-reported age and gender, (ii) population demographics (population size, immigration, education, income) obtained from Eurostat and the National Institute of Statistics and Economic Studies (INSEE) of France, and (iii) information on musical venues gathered using SongKick's API[3] to estimate the number of active musical events. We utilized these variables to construct a model for future testing, as shown in Figure 4. This model is represented by Directed Acyclic Graphs (DAGs), which offer a clear and efficient approach to identifying, presenting, and hypothesising the causal relationships between variables [29].

4. Discussion and Future Plans

By studying the behaviour of hundreds of thousands of music listeners in the real world, our study delves into the diversity of cultural consumption drawing two levels of perspectives, examining not only the breadth of an individual's consumption but also the alignment of consumption as a collective, stratified by age, gender, and geographical regions.

Our analysis reveals that both individual and collective levels of musical consumption diversity positively correlate with population size. This suggests that individuals residing in more densely populated regions not only have a broader spectrum of musical consumption but also exhibit more unique consumption profiles to one another within the population. This trend persists across different demographics, holding true when the population is segmented by gender

[3] https://www.songkick.com/developer

and age groups.

It is important to note, however, that our current findings only suggest correlations and do not establish causality. As such, in order to discern causal pathways that cultivate diversity at various levels, we outline potential confounding variables to consider for future research. These variables are incorporated into a model represented as a DAG. In subsequent research, we aim to test this model to determine the direct effects of various factors contributing to the observed diversity.

Our results align with prior research studying the trajectories of musical exploration across an individual's lifespan, revealing a general consolidation of consumption during adolescence. Yet, a notable observation from our study is the divergence in musical consumption between self-reported male and female listeners from a particular age onward. This disparity could potentially be a cultural phenomenon, or it may stem from other underlying factors that are hidden from our current observation.

In the broader scheme of verifying and generalizing the findings reported in this paper, we intend to replicate our analysis in other countries — Brazil and Germany. Additionally, we aim to openly share the demographic data and indicators we have compiled, along with the aggregated music stream data and corresponding analysis code for reproducibility and transparency.

References

[1] A. Anderson, L. Maystre, I. Anderson, R. Mehrotra, and M. Lalmas. "Algorithmic Effects on the Diversity of Consumption on Spotify". In: *Proceedings of The Web Conference 2020*. Taipei Taiwan: Acm, 2020, pp. 2155–2165. DOI: 10.1145/3366423.3380281.

[2] L. Briand, G. Salha-Galvan, W. Bendada, M. Morlon, and V.-A. Tran. "A Semi-Personalized System for User Cold Start Recommendation on Music Streaming Apps". In: 2021, pp. 2601–2609. DOI: 10.1145/3447548.3467110.

[3] A. Chao, N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, and A. M. Ellison. "Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies". In: *Ecological Monographs* 84.1 (2014), pp. 45–67. DOI: 10.1890/13-0133.1.

[4] P. G. Christenson and J. B. Peterson. "Genre and Gender in the Structure of Music Preferences". In: *Communication Research* 15.3 (1988), pp. 282–301. DOI: 10.1177/009365088015003004.

[5] M. Collard, B. Buchanan, and M. J. O'Brien. "Population Size as an Explanation for Patterns in the Paleolithic Archaeological Record: More Caution Is Needed". In: *Current Anthropology* 54.S8 (2013), S388–s396. DOI: 10.1086/673881.

[6] N. Creanza, O. Kolodny, and M. W. Feldman. "Greater than the sum of its parts? Modelling population contact and interaction of cultural repertoires". In: *Journal of The Royal Society Interface* 14.130 (2017), p. 20170171. DOI: 10.1098/rsif.2017.0171.

[7] D. Deffner, A. Kandler, and L. Fogarty. "Effective population size for culturally evolving traits". In: *PLOS Computational Biology* 18.4 (2022), e1009430. DOI: 10.1371/journal.pcbi.1009430.

[8] M. Derex, M.-P. Beugin, B. Godelle, and M. Raymond. "Experimental evidence for the influence of group size on cultural complexity". In: *Nature* 503.7476 (2013), pp. 389–391. DOI: 10.1038/nature12774.

[9] M. Derex and A. Mesoudi. "Cumulative Cultural Evolution within Evolving Population Structures". In: *Trends in Cognitive Sciences* 24.8 (2020), pp. 654–667. DOI: 10.1016/j.tics.2020.04.005.

[10] C. N. DeWall, R. S. Pond Jr., W. K. Campbell, and J. M. Twenge. "Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular U.S. song lyrics". In: *Psychology of Aesthetics, Creativity, and the Arts* 5.3 (2011), pp. 200–207. DOI: 10.1037/a0023195.

[11] N. Fay, N. De Kleine, B. Walker, and C. A. Caldwell. "Increasing population size can inhibit cumulative cultural evolution". In: *Proceedings of the National Academy of Sciences* 116.14 (2019), pp. 6726–6731. DOI: 10.1073/pnas.1811413116.

[12] M. Flemmen, V. Jarness, and L. Rosenlund. "Social space and cultural class divisions: the forms of capital and contemporary lifestyle differentiation". In: *The British Journal of Sociology* 69.1 (2018), pp. 124–153. DOI: 10.1111/1468-4446.12295.

[13] J. Henrich. "Demography and Cultural Evolution: How Adaptive Cultural Processes Can Produce Maladaptive Losses–The Tasmanian Case". In: *American Antiquity* 69.2 (2004), pp. 197–214. DOI: 10.2307/4128416.

[14] J. Henrich. "The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter". In: *The Secret of Our Success*. Princeton University Press, 2015. DOI: 10.1515/9781400873296.

[15] J. Henrich and R. McElreath. "The evolution of cultural evolution". In: *Evolutionary Anthropology: Issues, News, and Reviews* 12.3 (2003), pp. 123–135. DOI: 10.1002/evan.10110.

[16] N. Jacoby, E. A. Undurraga, M. J. McPherson, J. Valdés, T. Ossandón, and J. H. McDermott. "Universal and Non-universal Features of Musical Pitch Perception Revealed by Singing". In: *Current Biology* 29.19 (2019), 3229–3243.e12. DOI: 10.1016/j.cub.2019.08.020.

[17] M. Kempe and A. Mesoudi. "An experimental demonstration of the effect of group size on cultural accumulation". In: *Evolution and Human Behavior* 35.4 (2014), pp. 285–290. DOI: 10.1016/j.evolhumbehav.2014.02.009.

[18] M. A. Kline and R. Boyd. "Population size predicts technological complexity in Oceania". In: *Proceedings of the Royal Society B: Biological Sciences* 277.1693 (2010), pp. 2559–2564. DOI: 10.1098/rspb.2010.0452.

[19] Y. Koren, R. Bell, and C. Volinsky. "Matrix Factorization Techniques for Recommender Systems". In: *Computer* 42.8 (2009), pp. 30–37. DOI: 10.1109/mc.2009.263.

[20] C. L. Krumhansl and J. A. Zupnick. "Cascading Reminiscence Bumps in Popular Music". In: *Psychological Science* 24.10 (2013), pp. 2057–2068. DOI: 10.1177/0956797613486486.

[21] J. McDermott and M. Hauser. "The origins of music: Innateness, uniqueness, and evolution". In: *Music Perception: An Interdisciplinary Journal* 23.1 (2005), pp. 29–59. DOI: 10.1525/mp.2005.23.1.29.

[22] S. A. Mehr, M. Singh, D. Knox, D. M. Ketter, D. Pickens-Jones, S. Atwood, C. Lucas, N. Jacoby, A. A. Egner, E. J. Hopkins, R. M. Howard, J. K. Hartshorne, M. V. Jennings, J. Simson, C. M. Bainbridge, S. Pinker, T. J. O'Donnell, M. M. Krasnow, and L. Glowacki. "Universality and diversity in human song". In: *Science* 366.6468 (2019). DOI: 10.1126/science.aax0868.

[23] L. Mok, S. F. Way, L. Maystre, and A. Anderson. "The Dynamics of Exploration on Spotify". In: *Proceedings of the International AAAI Conference on Web and Social Media* 16 (2022), pp. 663–674. DOI: 10.1609/icwsm.v16i1.19324.

[24] M. Muthukrishna, B. W. Shulman, V. Vasilescu, and J. Henrich. "Sociality influences cultural complexity". In: *Proceedings of the Royal Society B: Biological Sciences* 281.1774 (2014), p. 20132511. DOI: 10.1098/rspb.2013.2511.

[25] M. Park, I. Weber, M. Naaman, and S. Vieweg. "Understanding Musical Diversity via Online Social Media". In: *Proceedings of the International AAAI Conference on Web and Social Media* 9.1 (2021), pp. 308–317. DOI: 10.1609/icwsm.v9i1.14620.

[26] R. A. Peterson and R. M. Kern. "Changing Highbrow Taste: From Snob to Omnivore". In: *American Sociological Review* 61.5 (1996), pp. 900–907. DOI: 10.2307/2096460.

[27] L. S. Premo. "Local extinctions, connectedness, and cultural evolution in structured populations". In: *Advances in Complex Systems* 15.01n02 (2012), p. 1150002. DOI: 10.1142/s0219525911003268.

[28] F. Reali, N. Chater, and M. H. Christiansen. "Simpler grammar, larger vocabulary: How population size affects language". In: *Proceedings of the Royal Society B: Biological Sciences* 285.1871 (2018), p. 20172586. DOI: 10.1098/rspb.2017.2586.

[29] J. M. Rohrer. "Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data". In: *Advances in Methods and Practices in Psychological Science* 1.1 (2018), pp. 27–42. DOI: 10.1177/2515245917745629.

[30] S. E. Street, T. Eerola, and J. R. Kendal. "The role of population size in folk tune complexity". In: *Humanities and Social Sciences Communications* 9.1 (2022), pp. 1–12. DOI: 10.1057/s41599-022-01139-y.

[31] A. Warde, D. Wright, and M. Gayo-Cal. "Understanding Cultural Omnivorousness: Or, the Myth of the Cultural Omnivore". In: *Cultural Sociology* 1.2 (2007), pp. 143–164. DOI: 10.1177/1749975507078185.