

Participant and Musical Diversity in Music Psychology Research

Kelly Jakubowski*¹, Nashra Ahmad¹, James Armitage¹, Logan Barrett¹, Aliya Edwards¹, Elizabeth Galbo¹, Juan S. Gómez-Cañón^{1,2}, Thomas A. Graves¹, Akvilė Jadzgevičiūtė¹, Connor Kirts¹, Imre Lahdelma¹, Thomas M. Lennie^{1,3}, Aliyah Ramatally¹, Joshua L. Schlichting^{1,4}, Chara Steliou¹, Keerthana Vishwanath¹, & Tuomas Eerola¹

¹Music and Science Lab, Department of Music, Durham University, UK

²Pyrros AI / Time Machine Capital, UK

³Department of Psychology, American University in Bulgaria, Bulgaria

⁴Department of Psychology, Neuroscience & Behaviour, McMaster University, Canada

*Correspondence concerning this article should be addressed to Kelly Jakubowski, Department of Music, Durham University, Palace Green, Durham, DH1 3RL, UK

(kelly.jakubowski@durham.ac.uk).

Author Contributorship Statement: KJ conceived of the study, and all authors contributed to co-designing the study and coding the dataset. TE and KJ led the analyses, and all authors contributed to interpreting the results. KJ wrote the initial manuscript draft, TE generated the figures and analysis scripts, and all authors contributed to revising the manuscript.

Data Availability Statement: The data that support the findings of this study are openly available at (anonymised for peer review): <https://anonymous.4open.science/w/WEIRD-90F8/>

Acknowledgements: Thanks to all members of the Durham University Music & Science Lab for their feedback and support. We have no conflicts of interest to disclose.

Abstract

Research on music psychology has increased exponentially over the past half century, providing insights on a wide range of topics underpinning the perception, cognition, and production of music. This wealth of research means we are now in a place to develop broad, testable theories on the psychology of music, with the potential to impact our wider understanding of human biology, culture, and communication. However, the development of widely applicable and inclusive theories of human responses to music requires these theories to be informed by data that is representative of the global human population and its diverse range of music-making practices. The goal of the present paper is to survey the current state of the field of music psychology in terms of the participant samples and musical samples used. We reviewed and coded relevant details from all articles published in *Music Perception*, *Musicae Scientiae*, and *Psychology of Music* between 2010 to 2022. We found that music psychologists show a substantial tendency to collect data from young adults and university students in Western countries in response to Western music, replicating trends seen across psychology research as a whole. Even data collected in non-Western countries tends to come from a similar demographic to studies of Western participants (e.g., university students, young adults). Some positive trends toward increasing participant diversity have been evidenced over the past decade, although there is still much work to be done, and certain subtopics in the field appear to be more prone to these sampling biases than others. Recommendations for future diversification of research in our field are made, with the aim of increasing our confidence in music psychological theories and their relevance to humans and music in the broadest sense.

Keywords: music psychology, sampling, diversity, WEIRD, cross-cultural research, generalizability, stimuli

Participant and Musical Diversity in Music Psychology Research

Music psychology is a growing research field, with the number of published papers in this area increasing exponentially since the launch of its first dedicated journal, *Psychology of Music*, in 1973 (Anglada-Tort & Sanfilippo, 2019). Music psychology research has addressed a wide range of topics, from the perception of musical features to musical learning, memory, and development, from emotional and aesthetic responses to music to the psychological processes underpinning performing and composing music, and its rehabilitative potential (Hallam et al., 2014; Thompson, 2015). This work intersects with several other psychology subdisciplines, including cognitive, developmental, social, evolutionary, clinical, and performance psychology, whilst also drawing upon methods and theories from a range of other fields such as neuroscience, computer science, sociology, and musicology. Research in music psychology thereby has great potential to inform and impact broader theories of human evolution, perception, development, learning, emotion, creativity, and communication.

Given the now established and relatively mature state of this field, music psychologists are adequately placed to develop broad, testable theories about how humans perceive, respond to, and make sense of music. However, to develop widely applicable and inclusive theories on the psychology of music, which may subsequently shape broader theories about human cognition, it is essential that research in this field captures the full spectrum of human experiences with music. This includes the need for responses from human participants across the lifespan, from varying geographic regions and cultural groups, and with a range of previous experiences with music. It also means “music” should be understood in a broad sense to encapsulate the diversity of musical styles and genres that humans engage with across the globe.

A range of research in psychology in general has demonstrated that theories and models based on responses from a limited participant demographic may be biased or incomplete. That is, generalisations about human cognition that are made based on relatively demographically homogenous samples have in various cases failed to replicate in other participant groups. Perhaps the most highly cited example of such work is that of Joseph Henrich and colleagues on the “WEIRD” problem (Henrich et al., 2010). WEIRD stands for Western, Educated, Industrialized, Rich, and Democratic, and this acronym was developed in reference to the fact that most previous psychological research has been conducted with participants from WEIRD societies, such as the US and UK. Specifically, “96% of psychological samples come from countries with only 12% of the world’s population” (Henrich et al., p. 63). The original article by Henrich et al. (2010) goes on to provide numerous examples of how research findings on a diverse range of topics including visual perception, spatial cognition, economic decision making, cooperation, conformity, and moral reasoning differ between more versus less WEIRD populations, thereby accentuating the potential pitfalls of developing theories about human cognition based solely on data from WEIRD samples. Henrich et al. (2010) also criticise the overreliance of psychology research on undergraduate student samples, mainly in relation to the “E” in WEIRD. It is worth furthermore noting the significant age biases introduced by such an approach, especially given the age differences known to exist in relation to various cognitive, emotional, and social processes (e.g., Defoe et al., 2015; Reed et al., 2014; Sparrow et al., 2021; Spencer & Raz, 1995). A follow-up article published 10 years after Henrich et al. (2010) summarises many notable advances in the field since the introduction of the WEIRD concept, whilst also highlighting that the problem is far from solved (Apicella et al., 2020). For instance, a

review of all papers published in *Psychological Science* in 2014 and 2017 showed that around 95% of studies still relied primarily on WEIRD participants (Rad et al., 2018).

Subsequent research has advocated for broadening or going beyond the WEIRD framework to consider other key demographic and cultural differences that predict behavioural diversity. For instance, differences in race and ethnicity, religion, community size, and cultures with strong versus weak social norms are not explicitly captured within the WEIRD framework, despite evidence that these variables can also have substantial impacts on human values and behaviours (Barrett, 2022; Clancy & Davis, 2019; Gelfand et al., 2011; Muthukrishna et al., 2020; Roberts et al., 2020). In addition, there has been a prominent tendency within cross-cultural research to make comparisons between participants from different countries without capturing the wide cultural diversity that typically exists within a single country (Barrett, 2022; Hartmann et al., 2013; see Jacoby et al., 2024 for one recent music psychological study revealing rhythm perception differences between participant subgroups both across *and* within countries).

In music psychology, similar sampling biases seem to exist. For instance, in a review of studies published in three music psychology journals from 1973 to 2017, Anglada-Tort and Sanfilippo (2019) found that the vast majority of the corresponding authors on these papers were from North America, Western Europe, or Australia. A review focused on the representation of authors from different European countries in articles published in *Musicae Scientiae* from 1997 to 2017 revealed a greater prevalence of authors from western and northern than eastern and southern Europe (Sloboda & Ginsborg, 2018). One might thereby reasonably assume that most research participants in these studies were likewise from these countries, although this was not examined in either of these two reviews. A preliminary study in our research group indicated that, within 90 studies on the effects of background music listening on concurrent task

performance, 71% of the studies were from the US or UK, and 66% relied solely on university students, with a mean age across all the studies (weighted by sample size) of 22.6 years.¹ In addition, an analysis of articles published in *Music Perception* from 1983 to 2010 revealed that 75% of studies reviewed used musicians as at least one subgroup of their participants (Tirovolas & Levitin, 2011). This suggests a potential bias toward relying on participant samples with a higher level of domain-specific expertise than the general population, although the exact amount of musical training could not be definitively specified due to a lack of consistent reporting across studies; it also remains to be seen whether such a trend has persisted in the field since 2010.

Another issue more specific to the music psychology field is how “music” is operationalised in research. The definition of what constitutes “music” varies widely across different countries and contexts, and thereby needs to be flexible and culturally informed (Jacoby et al., 2020). Beyond definitional issues, biases related to the background of the researchers and participant sample used can lead to biases in the stimuli selected for use in music psychological studies. For instance, 74% of the music collated in a review article on music stimuli previously used in emotion research by Warrenburg (2020) fell into the genre categories of (primarily Western) popular music (e.g., US chart-topping songs), Western art music (primarily by canonical composers such as Mozart and Beethoven), and Western film music. An analysis of the genres of music stimuli used in emotion research over time showed a heavy reliance on Western art (classical) music in the 1920s-1990s, followed by an upsurge of more studies using popular music in the 2000s-2010s (Warrenburg, 2020; see also similar results in Diaz & Silveira, 2014 and Tirovolas & Levitin, 2011). In a review of studies published in *Music Perception* from

¹ <https://musicscience.net/2016/09/30/weird-music-psychology/>

1983 to 2010, a trend toward using more ecologically valid stimuli over time (i.e., sequential music, rather than “isolated sounds”; Tirovolas & Levitin, 2011) was also found.

However, large-scale studies focusing specifically on the demographic makeup and participant and musical sampling methods within music psychology as a whole are lacking. Given that previous research has found that perceptual, emotional, aesthetic, and imaginative responses to music may vary significantly across different countries and cultural groups (Jacoby et al., 2019; Jacoby & McDermott, 2017; Jakubowski et al., 2022; Laukka et al., 2013; McDermott et al., 2016), there is a need to probe whether the conclusions made in our field to date may be limited by the participants samples from which these are drawn. Similarly, theories about how music is perceived or responded to that are developed solely in relation to Western (art) music may not be applicable to groups of people whose music is organised in a different way (for example, in terms of its harmonic structure; Athanasopoulos et al., 2021; Lahdelma et al., 2021; Lahdelma & Eerola, 2023). Indeed, multiple recent position papers in our field have called for the inclusion of a greater diversity of musics, cultures, and collaborators in our research (Baker et al., 2020; Jacoby et al., 2020; Sauv e et al., 2023).

The aim of this article is to summarise the current state of the field of music psychology in relation to the participant samples and musical samples that are used. Specifically, we reviewed all articles published in *Music Perception*, *Musicae Scientiae*, and *Psychology of Music* between 2010 to 2022, and extracted details about both the participant samples and musical samples from each paper, as well as the keywords and bibliographical information. This allowed us to provide a comprehensive overview of the typical samples and music being employed in the field as a whole over the past 13 years, as well as within several subdomains in the field, with a view toward making recommendations for future research priorities.

Method

Data Selection

We reviewed all articles published in *Music Perception*, *Musicae Scientiae*, and *Psychology of Music* between 2010 to 2022. These three journals were chosen as they represent the most prominent and longest-running journals in music psychology; our approach also follows other bibliometric reviews in this field (Anglada-Tort & Sanfilippo, 2019).² We note that there exist many other articles that could be classified within the field of “music psychology” that have been published in general psychology journals, journals on music education, neuroscience journals, and so on. However, we chose this approach to allow us to provide a comprehensive review of three journals that fit squarely within this field rather than, for instance, searching all existing journals in the fields of music and psychology, to avoid potentially missing particular relevant articles on account of our selected search terms, and to avoid having to make a subjective judgement on each individual article identified as to whether it should be considered a “music psychology” article. Articles that were purely theoretical in nature or review articles that did not report any new data were excluded from consideration in our study.

Data Coding

For each included article, we coded details from each reported study into a spreadsheet. Most articles (85%) reported only one study, but some articles reported as many as 8 separate studies. For each study, the paper title, year published, journal, first author’s name, and keywords of the paper were noted. To provide an overview of the geographic location of the work, the

² Specifically, *Music Perception*, *Musicae Scientiae*, and *Psychology of Music* were selected by Anglada-Tort and Sanfilippo (2019) as being the most impactful journals (determined by the SJR ranking provided in SCImago, www.scimagojr.com) that focused on the specific topic of music psychology. A search of the keyword “music” on the SCImago site in July 2023 revealed *Psychology of Music* and *Music Perception* were ranked 1 and 2 respectively, and *Musicae Scientiae* was ranked number 5. The third and fourth ranked journals (*Journal of Music Therapy*, *Journal of Research in Music Education*) did not focus specifically on music psychology and were hence excluded from consideration.

country of the first author's affiliated institution, country(s) the data were collected in, and the primary country(s) of origin of the participant sample(s) were also noted. If a first author reported two affiliated institutions in different countries, we retained only the first affiliation listed (this operation affected less than 1% of our dataset, specifically, 12 studies from 7 articles). We also coded the primary ethnicity of the participant sample, total sample size, sample age mean, sample age standard deviation, number of female participants, and number of male participants.³ In addition, the mean years of education of the participant sample, any musicianship description of the participant sample, and any other description (e.g., "undergraduate students") were coded, as well as the recruitment method (i.e., "volunteers", "course credit", "paid", "other"). If any music was used (for instance, as stimuli or in corpus analysis), we coded the primary musical genre(s) used for each study, origin country(s) of the music, and compositional source of the music (i.e., whether it was precomposed or composed for research purposes).

In cases where the word "primary" is used in the above variable descriptions, such variables were coded on the basis of the majority of a participant sample or musical stimulus set falling into a particular category. For example, "primary country of origin of the participant sample" was coded as the country where the majority of participants came from (e.g., if 60% of participants were from the US and 40% were from Canada, this was coded as "US"). If samples were equally divided (e.g., 50% of participants were from the US and 50% were from Canada), both countries were coded. In the case of the "primary country of origin of the participant sample" variable, we also coded multiple countries if an explicit cross-cultural comparison was

³ Numbers of participants falling into other gender categories, or declining to provide gender information, could thus also be estimated by subtracting the number of male and female participants from the total sample size.

made by the authors, regardless of whether the split of participants was exactly equal (e.g., if a cross-cultural study was conducted comparing musical responses between 60% participants from the US and 40% from India, both “US” and “India” were coded).

All missing data were coded as “NA” values; for example, papers that reported only musical corpus studies (with no human participants) received “NA” entries for all participant-related columns, but the details of the music sample were still included.

Data Availability Statement

A full list of instructions provided to the coders who processed the data (all authors on this paper), the data from this coding process, and all analysis scripts are openly available here (anonymised for peer review): <https://anonymous.4open.science/w/WEIRD-90F8/>

Data Analysis

In the data cleaning stage, we collapsed some subcategories that emerged in the coding process. For the musicianship description variable, we included the following subcategories as “musicians” for our subsequent analyses: professional musicians, amateur musicians, music students, and music teachers. For the broader description of the participant sample (beyond musicianship), the following subcategories of participants emerged: adults, amusics, audience, caregivers, children, clinicians, infants, older adults, patients, undergraduate students, university students, and university+ (“university+” comprised samples of primarily university students supplemented by community members, academic staff, etc.). For our analyses we considered any studies that involved the undergraduate students, university students, and university+ subcategories as “university samples.”

For analyses related to the keywords of articles, we created a simplified list of all keywords used, reducing variant spellings and forms (e.g., *expressiveness*, *expressivity*, and

expression were all categorised under “expression”) and grouping similar concepts together (e.g., *health*, *well-being*, and *wellbeing* were all categorised as “health”). We also omitted the keyword “music”, given its high prevalence across the dataset. A full list of these operations can be found in the accompanying analysis scripts on GitHub.

We classified the country data (specifically, the first author’s affiliation country, data collection country⁴, and country of origin of participants) as WEIRD or non-WEIRD using a recent list of WEIRD countries by Kryś et al. (2024) which allocates countries in the EU or EFTA, Australia, Canada, New Zealand, UK, US, and Israel as Western countries. Alternative metrics were also initially explored, including classifying countries by membership in the Western European and Other States Group (WEOG), a measure of cultural distance from the US developed by Muthukrishna et al. (2020), and Mensah’s (2014) taxonomy of cultures, and resulted in minimal differences from the results reported here.

Three levels of analysis were utilised as appropriate. The first was at the level of individual articles (N=1,360). The second was at the study level, since some articles comprised multiple studies (N=1,622). For analyses of features of musical samples, we focused on the 1,047 of these studies that utilised music (either for corpus analysis or as stimuli for research participants). Of the 1,622 total studies, 1,532 (94%) involved human participants (the others were solely corpus studies). Some of these human studies involved participants from multiple countries (e.g., a study comparing participant groups in Finland and India); for such studies we expanded the dataset into a third level which we refer to as the sample level (N=1,589). Our main analyses of demographic differences (age, musicianship, etc.) across WEIRD versus non-WEIRD countries were focused at this sample level.

⁴ When online data collection was used, we attributed these based on the country of origin of participants.

Results

In total, 1,360 articles reporting 1,622 studies and 1,589 samples were included in our dataset. Over half of the articles were published in *Psychology of Music* (738), with 348 articles published in *Music Perception* and 274 in *Musicae Scientiae* (see Figure 1B for number of articles by year). In these 1,360 articles, the most frequently listed countries for the first author's affiliated institution were the US (22% of articles), UK (15%), Australia (11%), Germany (9%), Canada (6%), and Finland (5%); overall, 91% of first author affiliations were from WEIRD countries (see Figure 1A). Similarly, data for the studies reported in these articles were most frequently collected in the following countries: US (22% of studies), UK (11%), Australia (8%), Germany (7%), Canada (7%), and Finland (3%), while 9% of datasets were collected online. The non-WEIRD countries where data collection most frequently occurred were Japan (1.2%), China (1.0%), and Brazil (1.0%). Figure 1C shows the total number of participants tested across all studies by country of data collection. The primary country of origin of the participant sample was reported less frequently than the country where the data were collected (54% of the time versus 96% of the time, respectively), although we might presume that in cases where the former was not reported it is because the sample primarily originated from the country in which the data were collected. For studies where the origin country of participants was reported, the most frequently appearing 6 countries remain the same as the two lists above (although with Germany now listed third and Australia fourth in rank order).

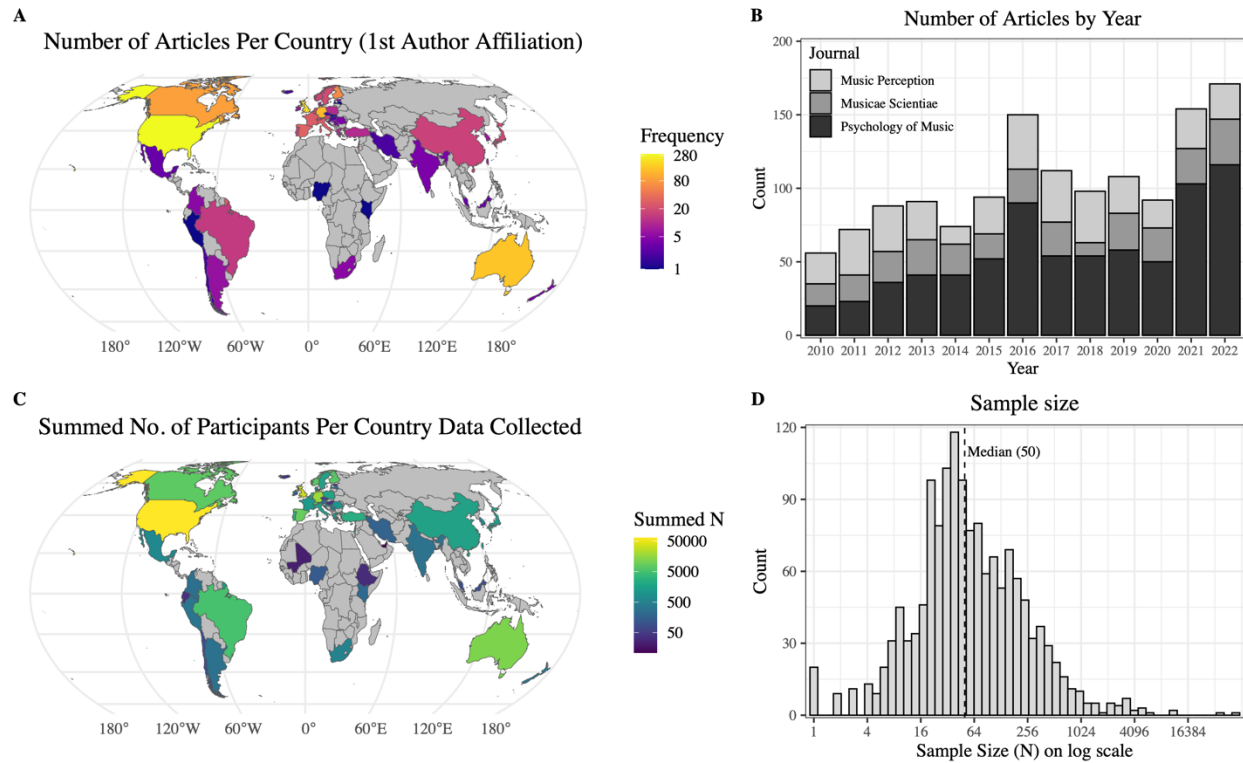


Figure 1. Summary of dataset regarding A) number of articles per country based on first author affiliation, B) number of articles per year by journal, C) total number of participants collected per country across all participant samples, and D) distribution of sample size (log scale) across all samples. Country data prepared with `countrynames` (Arel-Bundock et al., 2018) and `rnatuarearth` (Massicotte & South, 2023) packages.

Overall, the sample size of the studies ranged from 1 to 56,626 (a study of Twitter users) (*Median* = 50); 65% of studies had a sample size of fewer than 100 participants (see Figure 1D). The mean age of the participant samples across all studies was 26.5 years, and the mean age weighted by study sample size was 27.1 years; 63% of all participant samples had a mean age between 18-30 years. For studies that reported these details (76.2% of studies with 214,367 participants in total), 61.1% were female.

Table 1 shows the participant and musical sample variables across studies with data collected in WEIRD versus non-WEIRD countries. Participant samples collected in non-WEIRD countries were somewhat larger but not significantly so (Mann-Whitney test, $U = 103461$, $p = .486$). However, samples collected in non-WEIRD countries were significantly younger (sample size-weighted, $t_w(203.7) = 9.65$, $p < .001$) and showed less age variation ($t(151.2) = 8.26$, $p < .001$) than samples collected in WEIRD countries. In both WEIRD (59%, 95% CI [57%, 61%]) and non-WEIRD (54%, 95% CI [49%, 56%]) countries there was a tendency to collect more data from female than male participants, but this bias was more exaggerated in WEIRD countries ($t_w(139.2) = 4.21$, $p < .001$).

Table 1. Summary of key variables across WEIRD and non-WEIRD countries.

	Country Data Collected		p
	WEIRD	Non-WEIRD	
Sample Size (Md)	48 (42-52), n=1251	54.5 (38-68), n=160	.486
Age M (years)	27.6 (25.2-30.1), n=893	19.9 (17.1-22.1), n=119	<.001
Age SD	7.1 (6.2-8.1), n=673	3.7 (2.4-4.7), n=99	<.001
Gender balance (% female)	59 (57-61), n=923	54 (49-56), n=109	<.001
Solely musicians (%)	30 (28-33), n=381	30 (23-37), n=48	.920
Musicians & non-musicians (%)	53 (50-56), n=662	50 (42-58), n=80	.494
Solely non-musicians (%)	10 (9-12), n=131	14 (9-19), n=22	.223
University sample (%)	41 (38-44), n=510	43 (36-51), n=69	.597
Recruitment: volunteers (%)	30 (27-32), n=370	34 (27-42), n=54	.321
Experimenter-created music (%)	34 (31-38), n=270	34 (27-41), n=59	.925
Western music (%)	73 (70-76), n=572	59 (53-67), n=104	<.001

Notes. All participant related summaries are based on individual samples within studies utilising participants where primary country of data collection was given (N=1,411). The descriptors relating to musical samples (experimenter-created music and Western music) are derived from a subset of studies utilising music (N=1,047). Sample sizes are compared statistically with a Mann-Whitney test. Age and gender balance and t-tests have been weighted by the sample sizes. Proportions of musicians/non-musicians, university students, volunteers, experimenter-created music, and Western music have been analysed with χ^2 tests. Values in brackets refer to 95% confidence intervals obtained through bootstrapping.

In total, 30% of all samples were solely musicians (e.g., performers, composers, music teachers, music students), 10% were solely non-musicians, and 22% involved both musicians and non-musicians; the musicianship of participants was not reported for 37% of samples. Table 1 shows that the percentage of samples comprising these different musicianship categories did not significantly vary across WEIRD and non-WEIRD countries (for the purpose of this

comparison, the 37% of samples not reporting this information were classified as comprising both musicians and non-musicians). Overall, 21% of samples comprised solely university student samples, and a further 12% were solely undergraduates, while 9% of samples involved children or infants, and information on this aspect of the sample was not reported in 53% of cases. The proportion of samples using university populations did not significantly differ between WEIRD (41%) and non-WEIRD countries (43%; see Table 1). In terms of participant recruitment, 30% of samples were solely volunteers, 10% were solely paid participants, 9% comprised solely those recruited for course credit, and 9% used some combination of these strategies (or other rewards), while this factor was not reported for 41% of samples. The balance of volunteer versus non-volunteer participants did not significantly differ across samples collected in WEIRD (30%) versus non-WEIRD (34%) countries (see Table 1).

Only 9% of samples included ethnicity data (within this, 53% of samples were primarily white and 26% were primarily Asian) and years of education were reported for 6% of samples (within this, the mean number of years of education was 12.6). Due to the large amount of missing data, these two factors of ethnicity and years of education were not considered further in our analyses.

In total, 65% of all studies ($N = 1,047$) used some sort of music (the others were, for example, interview or questionnaire studies that did not explicitly present any musical stimuli to participants or analyse a corpus of music). From this set of studies, 56% used precomposed music, 33% used experimenter-created music, and 5% used semi-precomposed music (e.g., existing music that was reworked or edited for the purposes of the experiment); the other 6% of studies used some combination of these categories. The percentage of studies using experimenter-created music did not significantly differ between WEIRD and non-WEIRD

countries (see Table 1). Artificial sounds (e.g., sine waves, single tones) were used as stimuli in 21% of all studies. Solely Western music was used in 71% of studies, and the most frequently used Western genres were: classical (47% of studies using Western music), pop (15%), rock (9%), and jazz (9%). Western music stimuli were used significantly more often in WEIRD countries (73%, 95% CI [70%, 76%]) than non-WEIRD countries ($\chi^2 = 11.95, p < .001$, see Table 1), although even in non-WEIRD countries Western music was used in the majority of studies (59%, 95% CI [53%, 67%]).

We also examined trends over time in relation to the year of publication of each study. Figure 2 shows that the percentage of participant samples from WEIRD countries is decreasing over time, as is the percentage of studies with first authors from WEIRD countries, although both at a relatively slow (but statistically significant) rate, with the number of samples and first authors from WEIRD countries in the final year of data collection (2022) still both above 80%. Samples from WEIRD countries are gradually becoming more varied in age; the same seems to be the case for samples from non-WEIRD countries, although it is difficult to make strong conclusions due to the low number of observations per year. Studies conducted in both WEIRD and non-WEIRD countries are tending to use fewer university participant samples over time, although these trends are not statistically significant. The use of Western music samples and percentage of female participants show relatively little systematic change over time.

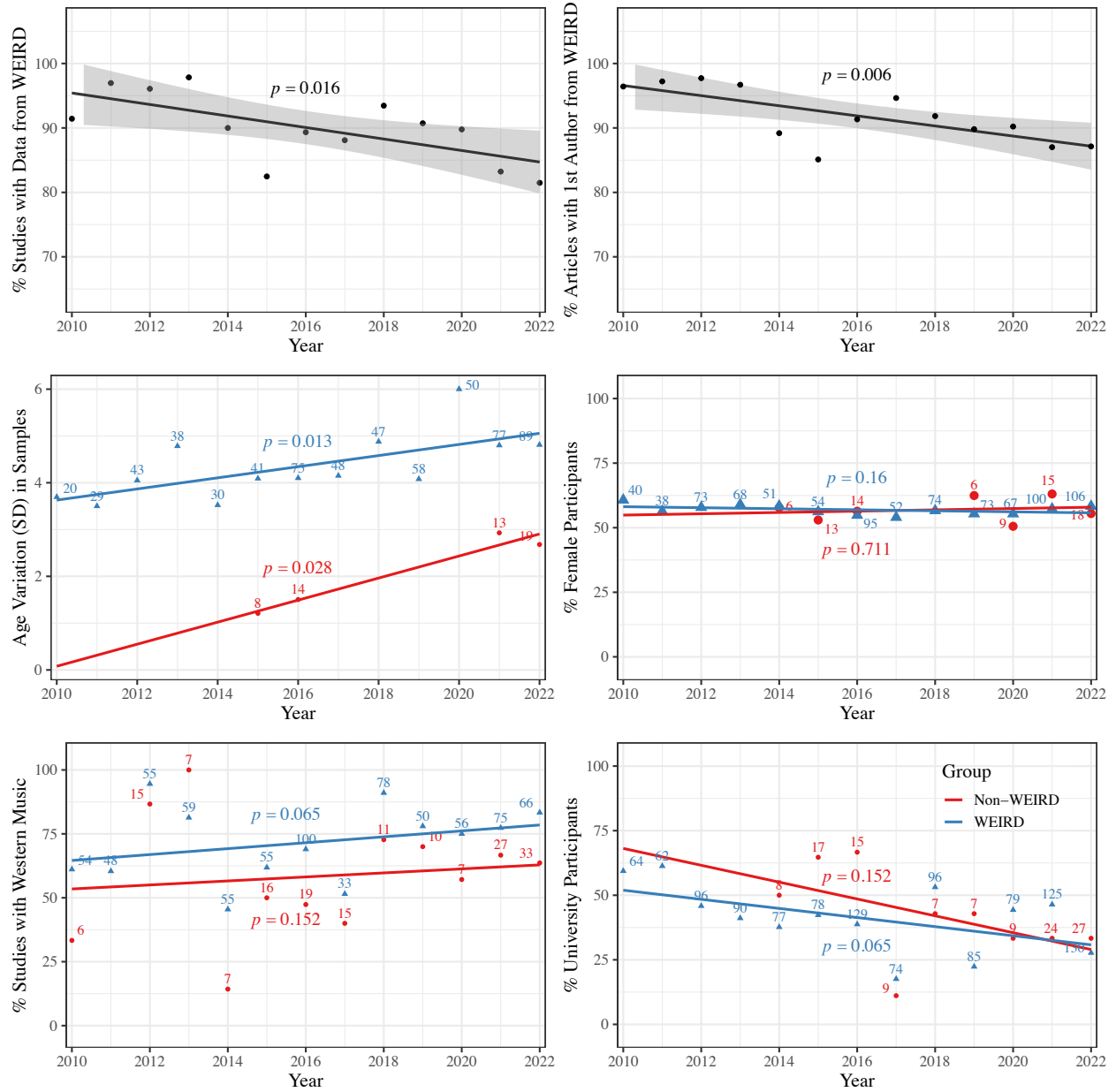


Figure 2. Trends across time for key metrics. The numbers next to the markers refer to the number of observations in the dataset (observations of 5 cases or fewer have been omitted). The p-values refer to statistical significance of linear trends.

Finally, we conducted an initial exploration into whether particular subtopics within the field show more sampling biases than others. Figure 3 shows the simplified keywords, following

the pre-processing steps described in the *Method*, in relation to several key participant and musical sample variables. For each variable we retained the top ranked 25 keywords in that subset of the data. Across all these keywords there was a consistent tendency to collect data more frequently in WEIRD than non-WEIRD countries; this bias tended to occur most often in relation to studies on musical performance and training (including learning and expertise), movement studies, studies using children or EEG, and (music) perception studies. Perhaps unsurprisingly, studies focused more on musician populations tended to be about anxiety, practice, performance, learning, and expertise; studies on music preferences and music listening, as well as studies of children, were typically focused on non-musicians. Studies on music perception and memory tended to utilise high percentages of both university students and female participants; these subtopics also fall fairly close to the median age of 24.1 years. Studies on health, singing, identity, and social factors (and unsurprisingly, children) tended to use fewer university samples. Health-related studies tended to focus on older participant samples; with the exception of the keyword “children” all other keywords fall within 5 years of the median age. Studies on harmony and movement showed the greatest tendency to use Western musical stimuli. It should also be noted that studies lying at the lower end of the range in terms of the percentage of Western music used are often simply studies that did not specify the genre of music used; for example, studies related to identity, anxiety, health, and personality often rely on self-selected music from participants, which can cover a wide range of genres.

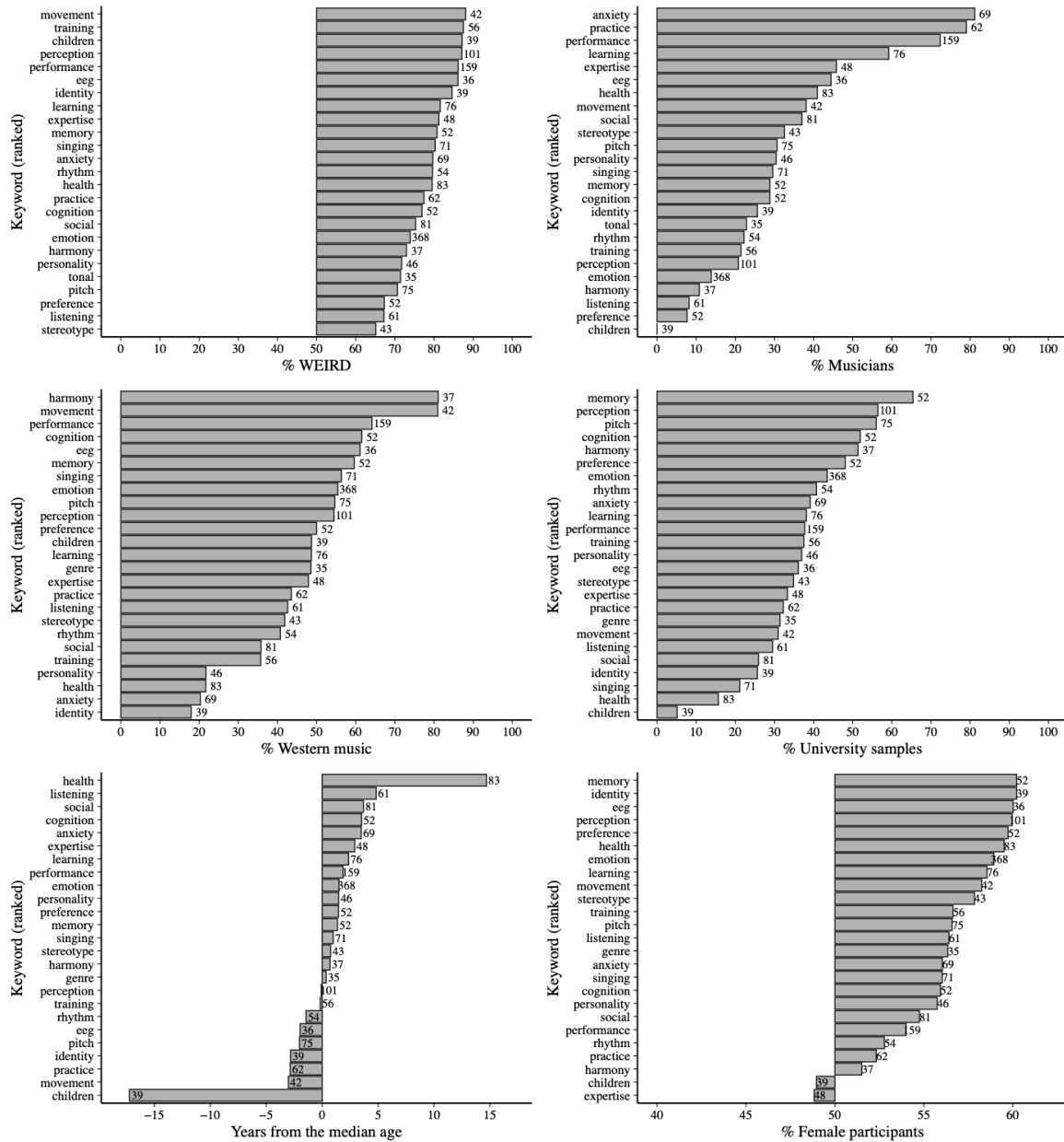


Figure 3. Summaries of keywords associated with key metrics (data collection in WEIRD countries, proportion of studies using musicians, Western music, university samples, age (years from the median age), and gender balance). Numbers next to the bars refer to the counts of articles with the keyword.

Discussion

The aim of this article was to survey the current state of the field of music psychology in terms of the diversity of its participant and musical samples. We have collected a range of measures of sampling practices from 1,360 articles published in 3 journals over 13 years, and considered how these practices vary across WEIRD and non-WEIRD countries, over time, and in relation to frequently occurring article keywords. Our results indicate that sampling biases are present throughout several dimensions of music psychology research and provide new insights into areas where future work is most needed.

The bias toward recruiting participants within WEIRD countries that has been shown in psychology research as a whole is also prevalent throughout music psychology research, across a diverse range of subtopics in this field (see Figures 1 and 3). The percentages of music psychology studies conducted in WEIRD countries are roughly similar to other areas of psychology research (Apicella et al., 2020; Rad et al., 2018). One positive point to note is that we have evidenced a gradual decrease over time in the number of studies conducted in WEIRD countries and the number of WEIRD first authors, although in both cases non-WEIRD groups are still consistently in the minority (see Figure 2). Although various recent calls have been made to increase the diversity of collaborators and participants in music psychology research (Baker et al., 2020; Jacoby et al., 2020; Sauv e et al., 2023) there is still much work to be done in practice, while the prevalence of first authors from WEIRD countries raises questions about the inherent power dynamics within research collaborations and the privileged position of native English speakers in academic publishing practices (Amano et al., 2023).

It is notable that several sampling trends that are prevalent within WEIRD populations are also replicated across non-WEIRD samples in this dataset. For instance, the proportion of

participants who were university students, volunteers, and musicians did not differ between samples from WEIRD versus non-WEIRD countries, and data collected in non-WEIRD countries was similarly skewed toward young adult samples (see Table 1). This suggests that even music psychological samples collected in non-WEIRD countries may be relatively more WEIRD than other subgroups from those countries. One recent large-scale, cross-cultural study that demonstrated the limitations of such an approach is the work of Jacoby et al. (2024), who showed that university students and/or online opportunity samples from non-WEIRD countries exhibit more similar rhythm priors to US participants than other non-WEIRD samples. Such sampling methods may thus underestimate the true cross-cultural variability of participant responses.

Beyond consideration of the countries in which data is collected, there are several other biases inherent in the current dataset. The first is a notable age skew, with around 2/3 of studies reporting a mean age between 18-30 years. As shown in Figure 3, very few subtopics showed much deviation from the median age of 24.1 years, although the age variation in samples does seem to be increasing gradually over time (see Figure 2). Over 40% of studies included university student samples; this proportion has also slightly (though not statistically significantly) decreased over the 13-year period of this dataset. This suggests that there has been some positive movement toward diversifying the age and educational background of music psychology participants, although there is still work to be done. One potential explanation for this diversification is the increase of online studies in recent years; although they come with their own limitations, online recruitment services (e.g., via Prolific or Qualtrics Online Sample service) allow for pre-screening and representative sampling in relation to a wide range of

demographic variables, including age, socioeconomic and geographic factors, and even hobbies such as music making (Eerola et al., 2021).

In addition, certain subtopics within the field appear to be studied in less diverse populations than others. For example, studies on memory, (music) perception, and EEG tend to focus on relatively WEIRD, young adult populations, and studies on movement tend to use highly WEIRD participants and WEIRD music. This tendency is likely related to the paradigms and equipment used in such research, which have been traditionally less conducive to usages outside a “controlled lab” setting. Since many such lab settings, in particular those with costly facilities such as motion capture systems, tend to be based in WEIRD countries, this can perpetuate the bias to recruit WEIRD participants into studies on these topics. However, technological advances continue to augment our ability to reach broader populations. Examples include the rise of online implementations of a range of behavioural tasks previously constrained to lab-based paradigms (e.g., sensorimotor synchronisation, eye tracking; Anglada-Tort et al., 2022; Saxena et al., 2023) and the development of more portable/flexible technologies for collecting and analysing movement data, EEG data, etc. in fieldwork and naturalistic settings (e.g., Clayton et al., 2024; Jakubowski et al., 2017; Pearson et al., 2024; Pearson & Pouw, 2022; Zamm et al., 2019).⁵

Around 30% of participant samples comprised solely musicians. This is perhaps to be expected for certain subtopics in the field, such as studies of expert performance routines and music performance anxiety, which may be chiefly of interest to people who make music. However, many other subtopics (e.g., studies on pitch, personality, perception and cognition in general) tended to include perhaps a disproportionate number of studies focused on musicians

⁵ See also ongoing work on motion capture data from Mali:
https://www.uio.no/ritmo/english/projects/djembedance/djembedance_revised-project-description.pdf.

(see Figure 3), if one considers the relative proportion of musicians to non-musicians in the general population. This leaves questions as to whether certain theoretical conceptions about music cognition may be biased toward evidence from expert samples, although unpacking specific theories in this regard is beyond the scope of this article. In addition, the definition of what constitutes a “musician” varied widely across the papers we surveyed; this issue was highlighted in Tirovolas and Levitin’s (2011) review and seems to have subsequently persisted in the field. Furthermore, what constitutes a “musician” is likely to vary widely across cultures. For instance, definitions specifying a certain number of years of “formal training” may not be globally relevant, while in some cultures there is no clear cut differentiation between “musicians” and “non-musicians”, hence the definitions adopted across WEIRD and non-WEIRD countries may be necessarily varied.

Finally, our results indicate there is still a heavy focus on Western music throughout our field. This was most prominent in studies of WEIRD participants, and thus may partially reflect the desire to utilise musical stimuli of cultural relevance to one’s research participants. But most studies within non-WEIRD countries also tended to employ at least some Western music, suggesting there is still a strong tendency to use Western music as a baseline comparator from which cultural differences are sought, which propagates potentially problematic contrasts between “the West and the rest” (Hall, 2018; Sauv e et al., 2023). Results from Figure 2 also indicate that this trend is not decreasing over time. In some ways this is surprising, given that *music* from non-WEIRD countries is arguably more easily accessible than participants from non-WEIRD countries. For instance, online streaming services and platforms such as YouTube host a diverse array of music from across the globe. Corpora that are specifically curated for research purposes are harder to come by, however, recent years have seen the compilation of research

corpora of music from countries as wide ranging as India, Uruguay, Mali, Cuba, Tunisia, and Georgia (Clayton et al., 2021; Rosenzweig et al., 2020; Srinivasamurthy et al., 2021). The Global Jukebox project provides a database of music and extracted features from 5,776 songs from across the globe (Wood et al., 2022) and a recent global comparison of speech and song from over 50 languages is accompanied by openly shared data and audio recordings (Ozaki et al., 2024). Of course, simply having access to a diversity of music does not in itself enable meaningful research, and it is likely that music psychologists will need to continue to build collaborative links with researchers in a wider range of countries and disciplines (e.g., ethnomusicology, anthropology) to be able to fully exploit the possibilities afforded by these collections (Savage et al., 2023).

Limitations and Future Directions

There are several limitations to the current approach that should be noted. Due to substantial variation in reporting methods, it was not possible to break down our analyses into specific subpopulations within each country. It should be noted that within countries there are likely to exist many regional differences in demographic and socioeconomic factors, such that some regions/subpopulations of a non-WEIRD country may be better classified as WEIRD and vice versa. Furthermore, our classification of WEIRD and non-WEIRD countries in line with the method proposed by Kryś et al. (2024) represents only one of a myriad of possible dimensions upon which these countries could be compared. Alternatives include postcolonial (or decolonial) distinctions between the Global North and South (Dussel, 2013; Grosfoguel, 2022; Mignolo, 2011; Quijano, 2000), as it is clear from our results that countries classified as being from the Global South are also highly underrepresented in music psychology research. In addition, broad cultural groupings of WEIRD versus non-WEIRD are likely to shift and change with time due to

economic development and educational changes, industrialisation/deindustrialisation, climate migration from threatened zones, and so on.

Several dimensions which have proved useful in characterising cultural variation in previous research were not readily available for analysis across our dataset. For example, we found that ethnicity and education data were rarely or inconsistently reported (interestingly, years of music education seemed to be reported relatively more often than general educational data). Other factors such as the religious background of participants and the degree to which they are sampled from urban versus rural populations are very rarely reported, despite previous evidence showing that these factors can have a substantial impact on behavioural diversity across different subpopulations (Barrett, 2022; Muthukrishna et al., 2020). At the same time, it would be naïve for us to advocate that all researchers should or could collect data on these and all other potentially relevant demographic factors if they are not being considered as a factor of interest in the research design, given potential ethical concerns. For example, the data minimisation principle of the European Union’s General Data Protection Regulation states that “personal data shall be... adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (‘data minimisation’)” (Article 5(1)(c)). As such, the recommendation we make here is that some of these factors that have been relatively under-investigated (e.g., urban-rural comparisons) could be put at the forefront of future comparative music psychological studies of a wider range of populations from across this continuum.

In this project we did not consider any articles published outside three specialist music psychology journals. We are aware that there exist many cross-cultural music psychological studies that are published in non-specialist journals, although arguably there are also a plethora of articles on WEIRD populations published in such journals as well, and attempts to identify *all*

articles across any journal that may be classified as “music psychology” are likely to be biased themselves (e.g., in relation to the selection of keywords used to identify these, etc.).

Furthermore, our focus on journals published in English means we have explicitly not included journals in which authors from countries where English is not the native language may be inclined to publish music psychology research (e.g., Germany: *Jahrbuch Musikpsychologie*; Japan: *Journal of Music Perception and Cognition*).

It should also be noted that the authors of this article are all affiliated to a university in a WEIRD country, and we thus acknowledge the biases that come with our own position within the field. We do, however, come from a diverse range of countries and disciplinary backgrounds. As such, the results of this work have benefited from reflection from researchers from music psychology, ethnomusicology, and music computing from both the Global North and Global South. We believe this variety of perspectives has strengthened and diversified the range of insights provided throughout this article.

Conclusion

In conclusion, our results show that music psychology research exhibits an overwhelming tendency to collect data from young adults and university students in Western countries in response to Western music. Some positive trends toward increasing participant diversity have been evidenced over the past decade, although there is still much work to be done, and certain subtopics in the field appear to be more prone to these sampling biases than others. The dataset and all code used to produce these summaries have been made openly available with the hope that they serve as a starting point for investigating possible biases in other dimensions of music psychological samples. Continued efforts to expand the diversity of participant and musical samples require dedicated, strategic effort, but come with the fundamental benefit of increasing

our confidence in music psychological theories and their relevance to humans and music in the broadest sense. Beyond our immediate field, inclusive theories of music psychology have the power to inform and catalyse change in related areas, such as Music Information Retrieval, music-related artificial intelligence, and music education (Bull et al., 2023; Campbell et al., 2005; Huang et al., 2023; Sturm et al., 2019).

Ethical Approval

No new data from human participants were collected in this project, and all analyses are based on secondary data available in previously published journal articles. As such, no institutional ethical approval was required.

References

- Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., González-Trujillo, J. D., Montaña-Centellas, F., Paudel, K., White, R. L., & Verissimo, D. (2023). The manifold costs of being a non-native English speaker in science. *PLOS Biology*, *21*(7), e3002184. <https://doi.org/10.1371/journal.pbio.3002184>
- Anglada-Tort, M., Harrison, P. M. C., & Jacoby, N. (2022). REPP: A robust cross-platform solution for online sensorimotor synchronization experiments. *Behavior Research Methods*, *54*(5), 2271–2285. <https://doi.org/10.3758/s13428-021-01722-2>
- Anglada-Tort, M., & Sanfilippo, K. R. M. (2019). Visualizing Music Psychology: A Bibliometric Analysis of Psychology of Music , Music Perception, and Musicae Scientiae from 1973 to 2017. *Music & Science*, *2*, 205920431881178. <https://doi.org/10.1177/2059204318811786>
- Apicella, C., Norenzayan, A., & Henrich, J. (2020). Beyond WEIRD: A review of the last decade and a look ahead to the global laboratory of the future. *Evolution and Human Behavior*, *41*(5), 319–329. <https://doi.org/10.1016/j.evolhumbehav.2020.07.015>
- Arel-Bundock, V., Enevoldsen, N., & Yetman, C. (2018). countrycode: An R package to convert country names and country codes. *Journal of Open Source Software*, *3*(28), 848. <https://doi.org/10.21105/joss.00848>
- Athanasopoulos, G., Eerola, T., Lahdelma, I., & Kaliakatsos-Papakostas, M. (2021). Harmonic organisation conveys both universal and culture-specific cues for emotional expression in music. *PLOS ONE*, *16*(1), e0244964. <https://doi.org/10.1371/journal.pone.0244964>
- Baker, D. J., Belfi, A., Creel, S., Grahn, J., Hannon, E., Loui, P., Margulis, E. H., Schachner, A., Schutz, M., Shanahan, D., & Vuhan, D. T. (2020). Embracing Anti-Racist Practices in the

Music Perception and Cognition Community. *Music Perception*, 38(2), 103–105.

<https://doi.org/10.1525/mp.2020.38.2.103>

Barrett, H. C. (2022). Psychology Within and Without the State. *Annual Review of Psychology*, 73(1), 461–487. <https://doi.org/10.1146/annurev-psych-020821-110248>

Bull, A., Scharff, C., & Nooshin, L. (Eds.). (2023). *Voices for Change in the Classical Music Profession*. Oxford University Press New York.

<https://doi.org/10.1093/oso/9780197601211.001.0001>

Campbell, P. S., Drummond, J., Dunbar-Hall, P., Howard, K., Schippers, H., & Wiggins, T. (Eds.). (2005). *Cultural Diversity in Music Education: Directions and Challenges for the 21st Century*. Australian Academic Press.

Clancy, K. B. H., & Davis, J. L. (2019). Soy lent Is People, and WEIRD Is White: Biological Anthropology, Whiteness, and the Limits of the WEIRD. *Annual Review of Anthropology*, 48(1), 169–186. <https://doi.org/10.1146/annurev-anthro-102218-011133>

Clayton, M., Li, J., Clarke, A., & Weinzierl, M. (2024). Hindustani raga and singer classification using 2D and 3D pose estimation from video recordings. *Journal of New Music Research*, 1–16. <https://doi.org/10.1080/09298215.2024.2331788>

Clayton, M., Tarsitani, S., Jankowsky, R., Jure, L., Leante, L., Polak, R., Poole, A., Rocamora, M., Alborn, P., Camurri, A., Eerola, T., Jacoby, N., & Jakubowski, K. (2021). The Interpersonal Entrainment in Music Performance Data Collection. *Empirical Musicology Review*, 16(1), 65–84. <https://doi.org/10.18061/emr.v16i1.7555>

Defoe, I. N., Dubas, J. S., Figner, B., & van Aken, M. A. G. (2015). A meta-analysis on age differences in risky decision making: Adolescents versus children and adults. *Psychological Bulletin*, 141(1), 48–84. <https://doi.org/10.1037/a0038088>

- Diaz, F. M., & Silveira, J. M. (2014). Music and Affective Phenomena. *Journal of Research in Music Education*, 62(1), 66–77. <https://doi.org/10.1177/0022429413519269>
- Dussel, E. (2013). *Ethics of Liberation In the Age of Globalization and Exclusion*. Duke University Press.
- Eerola, T., Armitage, J., Lavan, N., & Knight, S. (2021). Online Data Collection in Auditory Perception and Cognition Research: Recruitment, Testing, Data Quality and Ethical Considerations. *Auditory Perception & Cognition*, 4(3–4), 251–280. <https://doi.org/10.1080/25742442.2021.2007718>
- Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'Amato, A., Subirats Ferrer, M., Fischlmayr, I. C., ... Yamaguchi, S. (2011). Differences Between Tight and Loose Cultures: A 33-Nation Study. *Science*, 332(6033), 1100–1104. <https://doi.org/10.1126/science.1197754>
- Grosfoguel, R. (2022). *De la sociología de la descolonización al nuevo antiimperialismo decolonial*. Akal.
- Hall, S. (2018). The West and the Rest: Discourse and Power (1992). In D. Morley (Ed.), *Essential Essays: Identity and Diaspora* (Vol. 2, pp. 141–184). Duke University Press. <https://doi.org/10.1215/9781478002710-009>
- Hallam, S., Cross, I., & Thaut, M. (Eds.). (2014). *The Oxford Handbook of Music Psychology, Second Edition*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198722946.001.0001>
- Hartmann, W. E., Kim, E. S., Kim, J. H. J., Nguyen, T. U., Wendt, D. C., Nagata, D. K., & Gone, J. P. (2013). In Search of Cultural Diversity, Revisited: Recent Publication Trends in

Cross-Cultural and Ethnic Minority Psychology. *Review of General Psychology*, 17(3), 243–254. <https://doi.org/10.1037/a0032260>

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>

Huang, R. S., Holzzapfel, A., Sturm, B. L. T., & Kaila, A.-K. (2023). Beyond Diverse Datasets: Responsible MIR, Interdisciplinarity, and the Fractured Worlds of Music. *Transactions of the International Society for Music Information Retrieval*, 6(1), 43–59. <https://doi.org/10.5334/tismir.141>

Jacoby, N., Margulis, E. H., Clayton, M., Hannon, E., Honing, H., Iversen, J., Klein, T. R., Mehr, S. A., Pearson, L., Peretz, I., Perlman, M., Polak, R., Ravnigani, A., Savage, P. E., Steingo, G., Stevens, C. J., Trainor, L., Trehub, S., Veal, M., & Wald-Fuhrmann, M. (2020). Cross-Cultural Work in Music Cognition. *Music Perception*, 37(3), 185–195. <https://doi.org/10.1525/mp.2020.37.3.185>

Jacoby, N., & McDermott, J. H. (2017). Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction. *Current Biology*, 27(3), 359–370. <https://doi.org/10.1016/j.cub.2016.12.031>

Jacoby, N., Polak, R., Grahn, J. A., Cameron, D. J., Lee, K. M., Godoy, R., Undurraga, E. A., Huanca, T., Thalwitzer, T., Doumbia, N., Goldberg, D., Margulis, E. H., Wong, P. C. M., Jure, L., Rocamora, M., Fujii, S., Savage, P. E., Ajimi, J., Konno, R., ... McDermott, J. H. (2024). Commonality and variation in mental representations of music revealed by a cross-cultural comparison of rhythm priors in 15 countries. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-023-01800-9>

- Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and Non-universal Features of Musical Pitch Perception Revealed by Singing. *Current Biology*, 29(19), 3229-3243.e12. <https://doi.org/10.1016/j.cub.2019.08.020>
- Jakubowski, K., Eerola, T., Alborn, P., Volpe, G., Camurri, A., & Clayton, M. (2017). Extracting Coarse Body Movements from Video in Music Performance: A Comparison of Automated Computer Vision Techniques with Motion Capture Data. *Frontiers in Digital Humanities*, 4. <https://doi.org/10.3389/fdigh.2017.00009>
- Jakubowski, K., Polak, R., Rocamora, M., Jure, L., & Jacoby, N. (2022). Aesthetics of musical timing: Culture and expertise affect preferences for isochrony but not synchrony. *Cognition*, 227, 105205. <https://doi.org/10.1016/j.cognition.2022.105205>
- Krys, K., de Almeida, I., Wasiel, A., & Vignoles, V. L. (2024). WEIRD–Confucian comparisons: Ongoing cultural biases in psychology’s evidence base and some recommendations for improving global representation. *American Psychologist*. <https://doi.org/10.1037/amp0001298>
- Lahdelma, I., Athanasopoulos, G., & Eerola, T. (2021). Sweetness is in the ear of the beholder: chord preference across United Kingdom and Pakistani listeners. *Annals of the New York Academy of Sciences*, 1502(1), 72–84. <https://doi.org/10.1111/nyas.14655>
- Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Beller, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion*, 13(3), 434–449. <https://doi.org/10.1037/a0031388>
- Massicotte, P., & South, A. (2023). *rnaturalearth: World Map Data from Natural Earth*.

McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature*, *535*(7613), 547–550. <https://doi.org/10.1038/nature18635>

Mensah, Y. M. (2014). An Analysis of the Effect of Culture and Religion on Perceived Corruption in a Global Context. *Journal of Business Ethics*, *121*(2), 255–282. <https://doi.org/10.1007/s10551-013-1696-0>

Mignolo, W. (2011). *Mignolo, Walter. The darker side of western modernity: Global futures, decolonial options*. Duke University Press.

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) Psychology: Measuring and Mapping Scales of Cultural and Psychological Distance. *Psychological Science*, *31*(6), 678–701. <https://doi.org/10.1177/0956797620916782>

Ozaki, Y., Tierney, A., Pfordresher, P. Q., McBride, J. M., Benetos, E., Proutskova, P., Chiba, G., Liu, F., Jacoby, N., Purdy, S. C., Opondo, P., Fitch, W. T., Hegde, S., Rocamora, M., Thorne, R., Nweke, F., Sadaphal, D. P., Sadaphal, P. M., Hadavi, S., ... Savage, P. E. (2024). Globally, songs and instrumental melodies are slower and higher and use more stable pitches than speech: A Registered Report. *Science Advances*, *10*(20). <https://doi.org/10.1126/sciadv.adm9797>

Pearson, L., Nuttall, T., & Pouw, W. (2024). *Landscapes of coarticulation: The co-structuring of gesture-vocal dynamics in Karnatak vocal performance*.

Pearson, L., & Pouw, W. (2022). Gesture–vocal coupling in Karnatak music performance: A neuro–bodily distributed aesthetic entanglement. *Annals of the New York Academy of Sciences*, *1515*(1), 219–236. <https://doi.org/10.1111/nyas.14806>

- Quijano, A. (2000). Coloniality of Power and Eurocentrism in Latin America. *International Sociology, 15*(2), 215–232. <https://doi.org/10.1177/0268580900015002005>
- Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences, 115*(45), 11401–11405. <https://doi.org/10.1073/pnas.1721165115>
- Reed, A. E., Chan, L., & Mikels, J. A. (2014). Meta-analysis of the age-related positivity effect: Age differences in preferences for positive over negative information. *Psychology and Aging, 29*(1), 1–15. <https://doi.org/10.1037/a0035194>
- Roberts, S. O., Bareket-Shavit, C., Dollins, F. A., Goldie, P. D., & Mortenson, E. (2020). Racial Inequality in Psychological Research: Trends of the Past and Recommendations for the Future. *Perspectives on Psychological Science, 15*(6), 1295–1309. <https://doi.org/10.1177/1745691620927709>
- Rosenzweig, S., Scherbaum, F., Shugliashvili, D., Arifi-Müller, V., & Müller, M. (2020). Erkomaishvili Dataset: A Curated Corpus of Traditional Georgian Vocal Music for Computational Musicology. *Transactions of the International Society for Music Information Retrieval, 3*(1), 31–41. <https://doi.org/10.5334/tismir.44>
- Sauvé, S. A., Phillips, E., Schiefelbein, W., Daikoku, H., Hegde, S., & Moore, S. (2023). Anti-Colonial Strategies in Cross-cultural Music Science Research. *Music Perception: An Interdisciplinary Journal, 40*(4), 277–292. <https://doi.org/10.1525/mp.2023.40.4.277>
- Savage, P. E., Jacoby, N., Margulis, E. H., Daikoku, H., Anglada-Tort, M., Castelo-Branco, S. E.-S., Nweke, F. E., Fujii, S., Hegde, S., Chuan-Peng, H., Jabbour, J., Lew-Williams, C., Mangalagu, D., McNamara, R., Müllensiefen, D., Opondo, P., Patel, A. D., & Schippers,

- H. (2023). Building Sustainable Global Collaborative Networks: Recommendations from Music Studies and the Social Sciences. In E. H. Margulis, P. Loui, & D. Loughridge (Eds.), *The Science-Music Borderlands* (pp. 347–366). The MIT Press.
<https://doi.org/10.7551/mitpress/14186.003.0032>
- Saxena, S., Fink, L. K., & Lange, E. B. (2023). Deep learning models for webcam eye tracking in online experiments. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-023-02190-6>
- Sloboda, J. A., & Ginsborg, J. (2018). 25 years of ESCOM: Achievements and challenges. *Musicae Scientiae*, 22(2), 147–160. <https://doi.org/10.1177/1029864918764574>
- Sparrow, E. P., Swirsky, L. T., Kudus, F., & Spaniol, J. (2021). Aging and altruism: A meta-analysis. *Psychology and Aging*, 36(1), 49–56. <https://doi.org/10.1037/pag0000447>
- Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging*, 10(4), 527–539.
<https://doi.org/10.1037/0882-7974.10.4.527>
- Srinivasamurthy, A., Gulati, S., Caro Repetto, R., & Serra, X. (2021). Saraga: Open Datasets for Research on Indian Art Music. *Empirical Musicology Review*, 16(1), 85–98.
<https://doi.org/10.18061/emr.v16i1.7641>
- Sturm, B. L. T., Iglesias, M., Ben-Tal, O., Miron, M., & Gómez, E. (2019). Artificial Intelligence and Music: Open Questions of Copyright Law and Engineering Praxis. *Arts*, 8(3), 115.
<https://doi.org/10.3390/arts8030115>
- Thompson, W. F. (2015). *Music, Thought, and Feeling: Understanding the Psychology of Music* (2nd ed.). Oxford University Press.

- Tirovolas, A. K., & Levitin, D. J. (2011). Music Perception and Cognition Research from 1983 to 2010: A Categorical and Bibliometric Analysis of Empirical Articles in Music Perception. *Music Perception*, 29(1), 23–36. <https://doi.org/10.1525/mp.2011.29.1.23>
- Warrenburg, L. A. (2020). Choosing the Right Tune: A Review of Music Stimuli Used in Emotion Research. *Music Perception*, 37(3), 240–258. <https://doi.org/10.1525/mp.2020.37.3.240>
- Wood, A. L. C., Kirby, K. R., Ember, C. R., Silbert, S., Passmore, S., Daikoku, H., McBride, J., Paulay, F., Flory, M. J., Szinger, J., D’Arcangelo, G., Bradley, K. K., Guarino, M., Atayeva, M., Rifkin, J., Baron, V., El Hajli, M., Szinger, M., & Savage, P. E. (2022). The Global Jukebox: A public database of performing arts and culture. *PLOS ONE*, 17(11), e0275469. <https://doi.org/10.1371/journal.pone.0275469>
- Zamm, A., Palmer, C., Bauer, A.-K. R., Bleichner, M. G., Demos, A. P., & Debener, S. (2019). Synchronizing MIDI and wireless EEG measurements during natural piano performance. *Brain Research*, 1716, 27–38. <https://doi.org/10.1016/j.brainres.2017.07.001>