Sex differences in music perception are negligible

Mila Bertolo*[1,2,3,4], Daniel Müllensiefen[5], Isabelle Peretz[3], Sarah C. Woolley[1,2,6], Jon T. Sakata†[1,2,6] & Samuel A. Mehr*†[2,4,7]

[1]Integrated Program in Neuroscience, McGill University, Quebec, QC, Canada H1A 2B4. [2]Center for Research on Brain, Language, and Music, Quebec, QC, Canada H3G 2A8. [3]International Laboratory for Brain, Music, and Sound Research (BRAMS), Department of Psychology, University of Montreal, Montreal, QC, Canada. [4]Yale Child Study Center, Yale University, New Haven, CT 06520, USA. [5]Department of Psychology, Goldsmiths, University of London. [6]Department of Biology, McGill University, Quebec, QC, Canada H3A 1B1. [7]School of Psychology, University of Auckland, Auckland 1010 New Zealand.

†These authors contributed equally.

*Corresponding authors. E-mail: mila.bertolo@mail.mcgill.ca, sam@auckland.ac.nz

Since Darwin[1], researchers have proposed that human musicality evolved in a reproductive context in which males produce music to signal their mate quality to females. Sexually selected traits involve tradeoffs in the costs of high-quality signal production and high-fidelity signal detection[2], leading to observable sexual dimorphisms across many species[3,4]. If musicality is a sexually selected trait in humans, males and females should then differ in their music perception ability, music production ability, or both. The evidence for this possibility is unclear, because previous reports of sex differences in human auditory perception are restricted in scope and inconsistent in direction[5–15]. Here, we report a test of music processing ability in 360,009 men and 194,291 women from 208 countries. In contrast to other non-musical human traits[16–19], and in contrast to music-related traits in non-human animals[20–23], we found no consistent advantage for either sex. The sex differences we did observe were negligible (Cohen's d range: 0.009-0.111) and Bayesian analyses indicated evidence in favor of the null hypothesis of no sex difference in general musical ability (Bayes Factor = 0.6). These results suggest that it is unlikely that music evolved in the context of sexual selection.

Main Text

Music is ubiquitous across human societies, and there are numerous similarities in the forms and functions of musical styles across societies and cultures[24–32]. The biological and evolutionary roots of musicality, accordingly, have been debated at length[5,7,33–38]. One common hypothesis, originally proposed by Darwin[1], but endorsed by many others[39–43] is that music evolved via sexual selection as a credible signal of mate quality: males produce music to signal mate quality to females, while females assess potential mates based on their musical ability. This idea echoes credible signals of mate quality found in many non-human species, such as in the tungara frog (*Physalaemus pustulosus*), where females select males based on the quality of advertisement calls[44]; in birds-of-paradise (*Paradisaeidae*), where males evolved highly complex behavioral, morphological, and acoustic "courtship phenotypes"[45]; and in red-winged black birds (*Agelaius phoeniceus*), where females but not males display acute differentiation between male calls and other species' imitations of them[3].

Sexually selected traits demonstrate a reliable pattern of sexual dimorphism because the costs involved with signal production and detection differ across the sexes. In species in which females incur the larger cost for reproduction (e.g., larger gametes, internal fertilization, pregnancy, and lactation), females tend to be more discriminating of social signals such as courtship vocalizations because the cost of reproducing with a lower-quality mate is higher for females[2,3,46–49]. Indeed, female advantages in the sensory detection and discrimination of courtship signals have been documented in bird[20,21], amphibian[50], and mammalian species[51]. If music evolved in the context of sexual selection, human males and females should therefore differ in their music perception ability, music production ability, or both.

The evidence for this prediction is mixed[5–9,52]. Some studies report higher musical abilities in men[10] and in individuals with lower digit ratios (a trait taken to indicate higher in-utero testosterone exposure)[11]. Others report a female advantage in melodic recognition[13] and auditory sensitivity[14]. Still others report no sex difference in pitch or beat perception abilities overall but possible differences in prevalence at the low extremes of these abilities[53]. Two studies of associated predictions of a sexual selection account have had mixed results: testosterone levels, as measured via saliva assays, were not predictive of men's musical aptitude[12], and in a large twin study, higher musical ability correlated with *lower* reproductive success[15].

Further, several papers reporting evidence of sexual selection for musicality have been retracted or have documented failures to replicate: the claim that women are more attracted to men with apparent musical ability[54] was retracted in 2020[55]; the claim that women are more attracted to higher quality dancing in men[56], retracted in 2013[57]; and the claim that women prefer more complex music around ovulation[58] failed to replicate[59]. Moreover, even in cases where a *bona fide* sex difference in found, such a difference could reflect sociocultural forces rather than biological forces[7]. This pattern has led to some skepticism regarding the sexual selection hypothesis for musicality evolution[5,7,52].

Here, we employ a large-scale, citizen-science approach to examine sex differences in music perception. 554,300 people participants from 208 countries (see *Figure 1*) were recruited via https://themusiclab.org; of these, n = 194,291 (34.5%) self-identified as "female", and n = 360,009 (64%) self-identified as "male". They completed three musical perception-based tests presented in a random order, measuring their i) mistuning perception[60], ii) melodic detection[61], and iii) beat alignment perception[62]. They also reported demographic information and details of their musical training (if any) and listening habits (see *SI Table 2*).

Are there sex differences in music perception?

We found statistically significant but negligibly sized sex differences in a measure of general musical ability derived from the three tests, along with similarly negligible differences on each of the tests ($p < 0.05$ for each; *Figure 1, Table 1*). Women, on average, scored higher than men on general musical ability, but barely so (mean ± SD; women: 0.002 ± 0.331; men: -0.001 ± 0.34). This difference is approximately 200 times smaller than the size of the average height difference between human males and females, a well-documented sexual dimorphism ($d = 1.63$[17]). Further, a Bayesian approach to this analysis provided evidence leaning in favor of the null hypothesis of no true sex difference in general musical ability (Bayes Factor = 0.6; *Table 1*).

The pattern of results on the individual perception tests was internally inconsistent. Women outperformed men in the mistuning perception test (women; 0.486 ± 0.857; men: 0.402 ± 0.9), but men outperformed women on the beat alignment (women: m = 0.126 ± 0.979; men: 0.234 ± 0.962) and melodic discrimination tests (women: 0.309 ± 1.046; men: 0.33 ± 1.064). Bayesian approaches to these differences supported their robustness (Bayes Factors >1000), despite their very small sizes and inconsistent directions. Cohen's d effect sizes ranged from 0.009-0.111, and all of these values would be considered as null or very small effects[16,63,64] (see *Table 1* for other measures of effect sizes).

This lack of clear sex differences cannot be explained by methodological quirks or poor testing; multiple other participant measures explained substantial variability in general musical ability scores. For example, participants' self-assessed music production skill level was strongly related to performance on the music perception tests: those reporting they "have no skill" performed far worse than those reporting being "experts" ($d = 1.66$; *Figure 2*). This relation also held when analyzing men and women separately; in women, the difference in general musical ability between "no skill" and "expert" was $d = 1.53$, and in men it was $d = 1.73$ (see *SI Figure 2*). Similarly, the degree of self-reported music lessons was highly predictive of performance of the tests; those participants who reported they had taken lessons scored higher than those who reported having never taken lessons ($d = 0.68$; *SI Figure 3*); and moderately sized effects of language experience on music perception ability in a subset of these data have been reported elsewhere[65].

Is there Greater Male Variability?

The "Greater Male Variability Hypothesis" argues that some dimorphisms manifest as greater variability - not greater averages - in males than in females (e.g., "more idiots, more geniuses"[36]). Significant ($p < 0.05$ for each) but small gender differences in variances were observed, but like the effects found in individual tests, they were of internally inconsistent directions: variances were significantly larger for men in the mistuning perception (men:women variance ratio: 1.101) and melodic discrimination tests (men:women variance ratio: 1.034), and larger for women in the beat alignment test (men:women variance ratio: 0.965; *Table 3*). The variance ratios all being close to 1, however, indicates highly similar variances between men and women[63]. Thus, we find no clear evidence in support of higher male variability.

How well do other factors drive music perception ability?

We asked whether performance on the music perception tests might have been driven by participant covariates beyond their sex, by repeating the main analyses while adjusting for participants' age, education, age of start

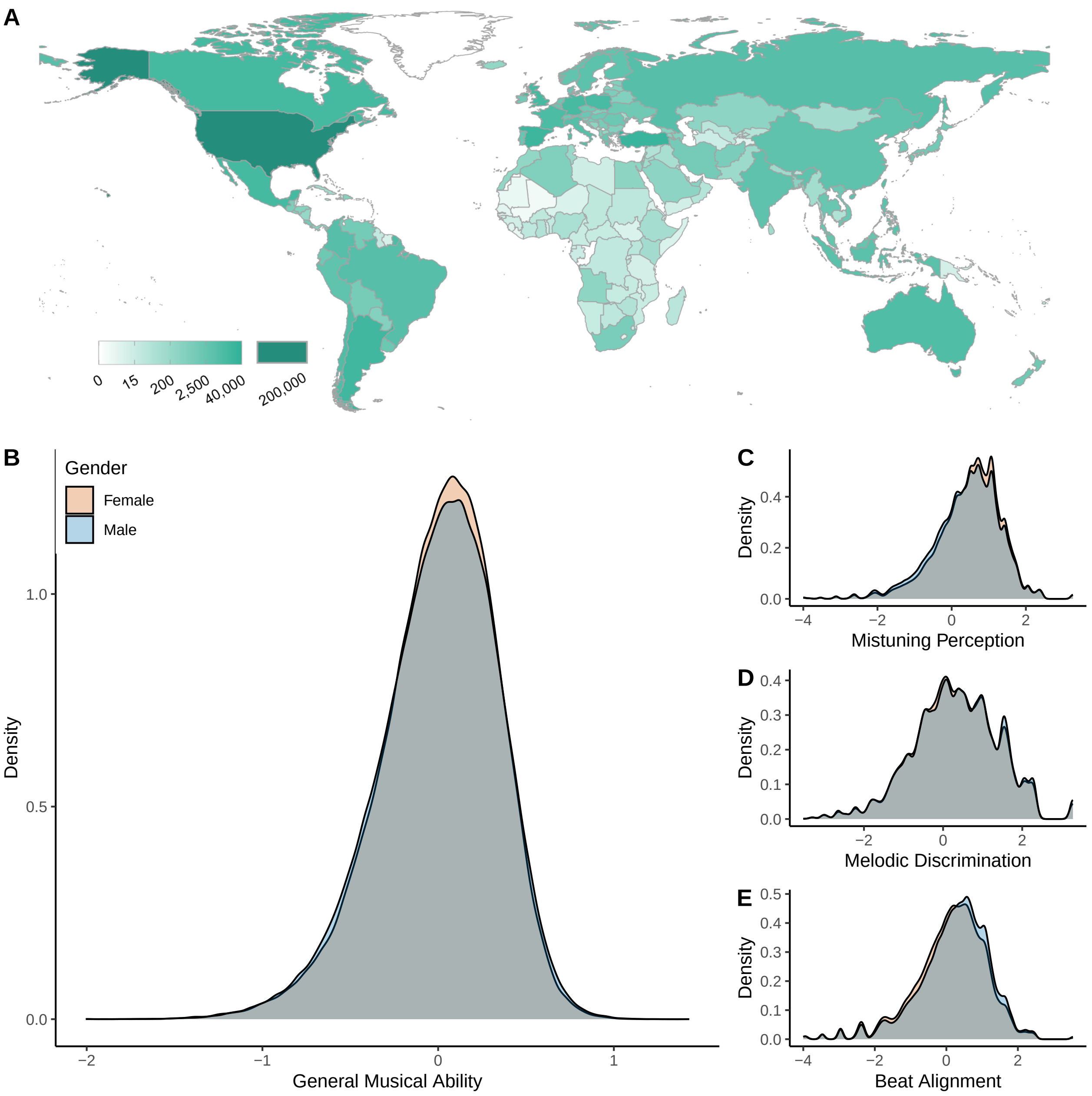


Fig. 1 | In a large global sample of participants, we find negligibly small sex differences in general musical ability. **A** Geographic spread of participants who completed music perception tests. The shading of each country indicates the number of participants who self-reported that country as their location. **B** In an aggregate measure of general musical ability (composed of measures **C, D, E**), there is a negligibly small difference between men and women's scores. The shaded areas represent kernel density estimations.

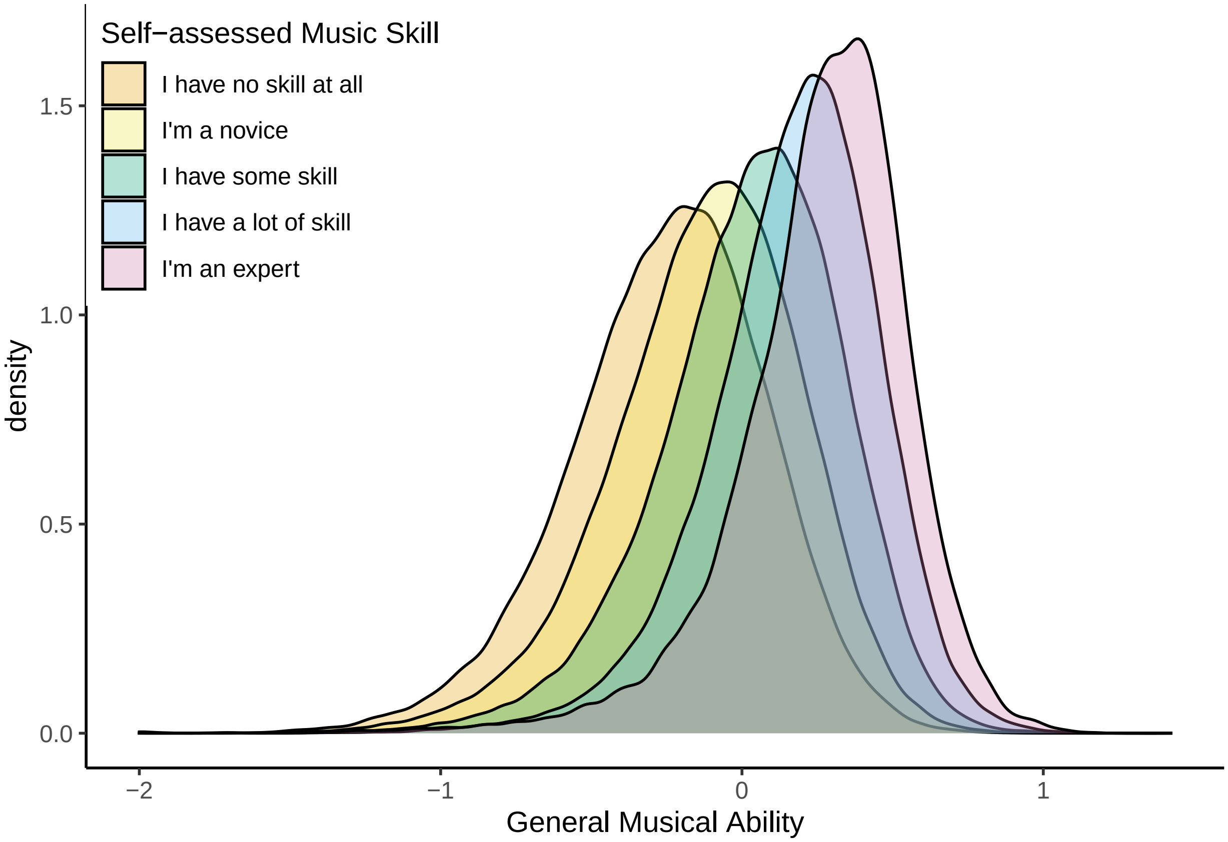


Fig. 2 | Distributions of general musical ability scores, by self-reported musical skill. As expected, general musical ability scores differ much more along other participant covaraites, such as self-estimated musical skill, than they differ by sex. The shaded areas represent kernel density estimations. The verbal prompt was *'Think of your skill at making music (using a musical instrument or singing). How would you rate your own skill?'* The participant covariate of self-assessed musical skill captured much more variance in participants' general musical ability scores, compared to their sex (partial η^2 =0.172 for self-assessed musical skill, relative to partial η^2 =0.007 for sex). Cohen's d values were all in the range of what would conventionally be described as medium effect sizes (Cohen's d between 'no skill' and 'novice' = 0.396; between 'novice' and 'some skill' =0.472; between 'some skill' and 'a lot of skill' =0.486; between 'a lot of skill' and 'expert' =0.339).

Task	Cohen's D	% Overlap	Probability Superior	η^2	R^2	Bayes Factor	Variance Ratio (m/f)
Mistuning Perception	0.096	0.954	0.527	0.00205	0.00205	7.41e+244	1.101
Melodic Discrimination	-0.020	0.934	0.494	0.00009	0.00009	4.49e+08	1.034
Beat Alignment	-0.111	0.950	0.469	0.00280	0.00280	Inf	0.965
General Music Ability	0.009	0.980	0.503	0.00002	0.00002	6.07e-01	1.057

Table 1 | Effect Sizes Effect sizes of gender differences in musical listening tasks. Note: Negative Cohen's D values indicate that men's scores are higher, and positive scores that women's scores are higher.

of music lessons (for participants who had completed some music lessons), and factors indicative of their musical training and music listening habits (for details about the factors, see *Analysis, & Table 3*). The inclusion of these covariates did not substantially alter effect sizes of sex differences (*SI Table 2*). Further analyses using structural equations modelling approaches also detected minuscule sex differences on the tests; these too found little influence of covariates on the magnitude of our estimated sex differences (see *Analysis & SI Figure 1*).

Discussion

Together, these findings provide no evidence for substantive sex differences in human music perception. Not only did the most general test of sex differences only find weak evidence for a sex difference (i.e. a statistically significant differences with a negligible effect size, and a low Bayes Factor of 0.6), but the reliable sex differences we *did* find were of a very small size and of inconsistent direction.

Those few reliable differences do not seem to support any claim of sexual selection, for two main reasons. First, human sex differences in traits argues to result from sexual selection tend to be large in size; some examples include height (d = 1.63)[17], vocal acoustic features (d = 2.7 - 5.7)[18], and rates of engagement in physical aggression (d = 1.11)[16]. These effects dwarf the sex differences we found on the three music tests.

Second, large sex differences have been observed in a variety of non-human animals that *do* use acoustic signals for courtship and sexual behavior. Female black-capped chickadees (*Poecile atricapillus*) outperform males on the discrimination of pitch ratios (partial $\eta 2 = 0.163$)[20], and are significantly faster than males at discriminating between socially dominant and subordinate male songs (Cohen's $d = 2.854$)[21]; and male Bengalese finches do not show the same individual recognition of other males (as seen in heart rate habituation) that females do[66]. Because in these species males and females often have categorically different behavioral responses to song, studies directly comparable to our own - where both males and females are presented with the same stimuli and assessed on the same behavioral assay - are sparse in the literature, and so there is little consensus as to whether these sex differences are strictly perceptual and/or motivational. While the exact mechanisms underlying these sex differences in non-humans animals' responses to these courtship signals is unclear, the magnitude of sex difference in behavioral responses equally dwarf the effects we found on the three music tests.

The lack of sex differences in music perception casts doubt on the mate quality hypothesis, but does not rule it out entirely. The ability to produce music requires complex sensory processing and sensorimotor integration; one can imagine unusual co-evolutionary dynamics at work, where music production is substantively dimorphic but music perception is not. We find this unlikely: in our data, self-reported music production skill was highly correlated with music perception ability in both males and females, suggesting the negligible sex difference in music perception extends to a comparable equivalence in music production abilities. While self-reports of music production ability are coarse measures relative to structured assessments of music production ability, we predict that future research using such tests will find comparably poor evidence for sexual dimorphism in musicality.

Methods

This research was approved by the Committee on the Use of Human Subjects and Harvard University's Institutional Review Board (protocol 2000033433). Data was collected from an online experiment hosted at https://www.themusiclab.org/quizzes/miq and advertised as a "Test Your Musical IQ" game. Recruitment was driven mainly by organic social media sharing which advertised the game to our global participant pool.

Participants

Our publicly accessible and globally disseminated online test reached 562,853 participants from 208 countries (at time of writing, where we considered data gathered between 22 Nov 2019 and 14 Dec 2020). Included in this dataset are only individuals who self-reported that they had not played the game before or did not have a hearing impairment. Of these, n = 194,291 (34.5%) self-identified as "female", n = 360,009 (64%) self-identified as "male" and n = 7,521 (1.3%) indicated "other" for their gender. Because sexual selection theory does not make explicit predictions with regards to how non-binary participants may differ from men and women, we only used data from participants who self-identified as either male or female. While individuals from over 200 countries participated in the test, the majority of participants were from North America (173,690, or 30.9%) and Europe (221,602, or 39.4%). Further information about participants' age and educational achievement is detailed in Table 2 and SI Table 1.

Gender	n	Mean Age (years)	Music Lessons	Age Music Lessons Started	Music Theory Training	Familiarity World Music
Female	194,291	23.6	72%	9.1	2.5	2.4
Male	360,009	25.3	63%	11.0	2.5	2.5

Table 2 | Participants Participant information, broken down by self-reported gender, age, and music lessons. Note for Music Theory Training: answers ranged from 'No music theory traning.' (scored 0) to 'A lot of music theory training' (scored 4). Note for Familiarity Traditional Music: answers ranged from 'I have never heard traditional music' (scored 0) to 'I am very familiar with traditional music' (scored 4).

Measures of music perceptual abilities

Participants were given three rounds of musical perception-based tests, each of which were based on existing and pre-validated music perception tests. These tests measures participants' abilities in the following domains:

i) Melodic discrimination[61]: In this test, participants listen to three different versions of the same melody. All three versions are played at different absolute pitch levels, with subsequent versions always transposed a semitone higher compared to the previous version. Two versions are identical in their interval structure but one of the three versions differs in one note. Differences can be in melodic interval, contour or tonality or any combination of the three factors. The participant's task is to identify the odd-one-out. The test is based on an explanatory item response theory model and is adaptive (for details see Harrison et al. (2017)[61]). This allows the procedure to select melody items dynamically such that item difficulty is matched to estimated participant ability for each trial. Final participant scores are computed according to the underlying item response theory model and have a theoretical range from -4 to 4, where higher scores represent greater melodic discrimination ability. All three music perception tests are adaptive in this way, with participant scores computed as described above.

ii) Mistuning perception[60]: In this test, participants are given two nearly identical short excerpts of vocal music. In one excerpt the vocal track was pitch-shifted against the instrumental back-track; the participant's task is to indicate in which of the two excerpts the singer sounds more out of tune.

iii) Beat alignment[62]: In this test, participants listen to the same clip of music twice, with an overlaid beep track. In one version the beep track is perfectly aligned with the musical beat of the music, and in the other version the beep track is slightly shifted in time. The participant's task is to identify in which of the two clips the beep track is best aligned with the beat of the music.

The short excerpts of music were all excerpts of Western popular music, i.e., music in a style that would be familiar to listeners acculturated to Western music, but not songs that they were likely to have heard before; all stimulus materials were sourced from stock libraries containing only songs that have not been commercially released. While our tests consisted only of musical excerpts in the Western musical tradition, because men and women from all geographic regions participated in the study, this test remains useful for quantifying sex differences around the world; any culture-driven differences in performance would apply equally to both men and women, and therefore are unlikely to confound the presently reported findings.

Procedure

Participants completed the experiment on computers, smartphones, or tablets. Participants were first asked to report their age, country, self-identified gender, native language, and whether they had a hearing impairment. They were also asked about the noisiness of the environment in which they were at the time of the experiment, whether they were wearing headphones, and if they had played the game before. Participants then completed the three music perception tests, in a randomized order. After the tests, they were asked a number of questions concerning their musical skill and experience (see *SI Table 3*).

Analysis

Data pre-processing

Of the >2 million people who had begun the online experiment at https://www.themusiclab.org/miq, we considered only those who had completed all three musical perception tests, only analyzed the data from participants who self-reportedly had not completed this game before, and excluded participants who reported an age younger than 7 or older than 100. Applying these filters resulted in a participant pool of n = 562,853. Questions that had written response labels were re-coded into binary or ordinal variables (see *SI Table 3*); these items formed conceptually meaningful groups: items pertaining to *i)* musical perception ability (ie. the three musical perception tests), *ii)* self-reported musical ability, *iii)* formal musical training, and *iv)* extent, intensity, and breadth of musical listening.

Factor analysis to extract composite scores

For each of the above groups of measures, information from their constituent sources was aggregated by factor analysis. For each of these factor analyses, the data sample was randomly split into two equal sized subsamples. An exploratory minimum residual factor analysis was run on one subsample, with the correlation matrix of the variables belonging to the same conceptual group. Questions differed in their response options and, by extension, the measurement level of corresponding variables. In these cases a mixed-type correlation matrix was computed employing polychoric, tetrachoric, biserial, point-biserial or Pearson correlations as appropriate for each pair of questions. The resulting correlation matrix was then subjected to a factor analysis. In all cases only a single factor was specified and only variables that loaded with a minimum of 0.3 were retained in the factor model. The second subsample was used for a confirmatory factor analysis and the factor structure was confirmed by inspecting measures of absolute model fit. On the condition that fit measures were satisfactory, the confirmatory factor model was then computed a second time, using the entire sample of participants, and factor scores were computed using only data from those participants with complete data on all the variables included in the factor model. We then used the Empirical Bayes Modal approach for computing factor scores from the confirmatory model. Table 3 details each factor's individual items, their loading on the factor, and how much variance among the indicators is explained by the single factor. For all factors, the confirmatory fit measures indicated a good model fit (General musical ability RMSEA $< .001$, SRMR, $< .001$; self-reported musical ability RMSEA $= 0.029$, SRMR $= 0.037$; self-reported musical training RMSEA $< .001$, SRMR, $< .001$; self-reported musical listening RMSEA $< .001$, SRMR, $< .001$).

Factor	Item	Loading onto Factor	Variance Explained
Music ability	Beat Alignment	0.442	0.290
	Melodic Discrimination	0.547	
	Mistuning Perception	0.612	
Self-reported musical ability	Tap In Time	0.821	0.522
	Out of Tune	0.689	
	Musical Skill	0.666	
	Music Listening Skill	0.706	
Musical Training	Enjoyed Lessons	0.517	0.235
	Compared to Peers in Lessons	0.528	
	Theory Training	0.397	
Musical Listening	Music Enjoyment	0.584	0.237
	Music Listening Time	0.443	
	Familiarity with World Music	0.416	

Table 3 | Factor Analysis Aggregation of related items into factors, and individual items' loadings onto their factor.

Estimating sex differences in music perception

To test whether there were any sex differences in music perception, after controlling for the above participant covariates, we tested for sex differences using a sequence of different analysis approaches.

First, a general additive model (GAM) was fit to each dependent variable, i.e., the three individual ability tests (melodic discrimination, mistuning perception, and beat alignment), and the aggregate factor score of general musical ability. This allowed us to calculate means and variances of these variables and test whether they differed between men and women. Men and women's distributions on each variable were plotted to visualize the extent of their overlap (*Figure 1*). On the basis of these data, we also calculated the probability superior (the probability that a randomly selected woman would have a higher score than a randomly selected man), the η^2 of how much variance sex explains in the models, the R^2 of the models, and the variance ratios (variance men / variance women) (*Table 1*). As a measure of effect size of a potential sex difference, we calculated Cohen's d as a parametric effect size, and percentage overlap of the two distributions as a non-parametric effect size (*Table 1*).

Secondly, in order to further test whether men and women differed on this latent variable "general musical ability" that we purport to measure using our three musical perception tests, we computed structural equation models (SEMs) using the factor scores on the three perceptual tests as indicators of a latent variable general musical ability (g) (see *SI Figure 1*). We constrained the intercepts and the loadings of the three tests on g to be equal across men and women, tested the difference in the latent means of g for significance, and computed Cohen's d effect sizes. This model estimated that the difference in g between men and women was 0.009, with a pooled variance of 0.552. This corresponds to an effect size of $d = 0.016$, which is comparable to the above estimates derived from the GAM.

Controlling for participant covariates

To test whether this magnitude of difference in g may have been confounded by other participant variables, the above SEM was extended to include other participant variables that may have plausibly been influencing their test performance - specifically participants' age, education, age at which they started music lessons (if any), musical training, and musical listening. This SEM approach additionally allows us to model the correlational structure among the covariates (see *SI Figure 1*). When taking covariate effects into account, the effect size increased slightly: the difference in g after covariate adjustment was 0.05, with a pooled variance of 0.47. This corresponds to an effect size of $d = 0.106$. This may reflect the influence of slight group differences on the above covariates.

Testing whether differences were robust across covariate subgroups

Lastly, we tested whether the very small sex differences observed above were robust across different combinations of participants' covariates (namely: age, education, start age of music lessons, and factors musical training and musical listening). To do this, we used a tree model based on recursive partitioning. The tree model splits covariates into subgroups for which the coefficient for sex in a linear model changes significantly. Given our large sample size, we set a significance threshold of $p < .001$ and required a minimum subgroup to have at least 1% of the total sample size. Tree models were automatically pruned using the Bayesian Information Criterion. We computed the range of the effect sizes (Cohen's d) across subgroups for the tree model of each dependent variable.

Most of these tree models formed subgroups based on participants' amount of musical training and start age of music lessons. The range of effect sizes across these tree model subgroups ranged from Cohen's $d = 0.001$ to 0.264, this higher end representing effect sizes that are conventionally considered small effects (i.e. $d > 0.2$)[67]. Across these tree models, most models estimated a sex difference of Cohen's $d < 0.2$: even when controlling for other participant variables, sex differences on tests of general musical ability are very small.

These tree models reveal that subgroups constructed from these other variables reveal slightly larger effect sizes than those estimated in the full distribution. These can be explained by the covariate musical training. Musical training is positively related to musical ability, and females participants reported having higher musical training on average. Hence, when data are conditioned on musical training (i.e. males and females are made statistically equal in terms of training), the estimated sex difference effect size slightly increased, with a small male advantage in most subgroups. This appears to be a version of Simpson's paradox, where there are (almost) no differences in the full sample, but slightly stronger differences emerge for (almost) all subgroups analysed.

End notes

Data, code, and materials availability

A fully reproducible manuscript; data, analysis code, and visualizations; other materials; and code for the naïve listener experiment are available at https://github.com/themusiclab/sex-differences. Readers may participate in the naïve listener experiment by visiting https://themusiclab.org/quizzes/miq.

Acknowledgments

This research was supported by the Harvard Data Science Initiative (S.A.M.), the National Institutes of Health Director's Early Independence Award DP5OD024566 (S.A.M), Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: 05016 (J.T.S.), Fonds de Recherche du Québec Nature et Technologies, Grant/Award Number: PR-299652 (S.C.W.), an Anneliese Maier research prize awarded by the Humboldt Foundation (D.M.), and the Center for Research on Brain, Language, and Music (M.B.). We thank the participants, members of The Music Lab and The Peretz Lab for feedback, C. Hilton for technical assistance, and P. Harrison for raw data-preprocessing.

Author contributions

Conception S.A.M., D.M.; experimental design and implementation, S.A.M.; participant recruitment, data management, and data processing, S.A.M., D.M., M.B.; analysis and visualization, M.B., D.M.; writing, M.B., I.P., S.C.W., J.T.S., S.A.M.

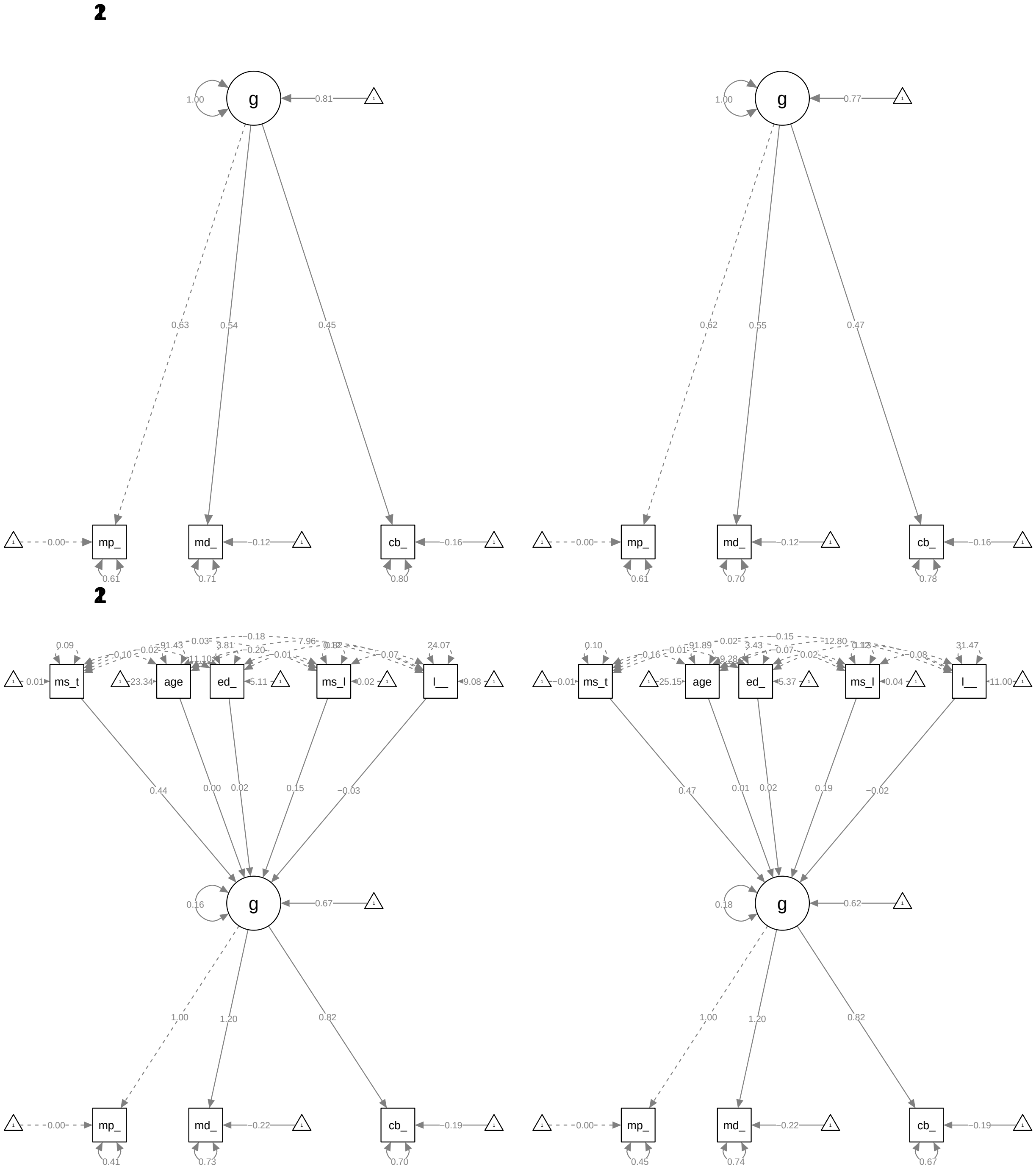
Supplemental Information

				Highest education completed			
Gender	n	Mean Age (years)	Age SD	High Schoool	Some Undergrad	Undergrad	Postgrad
Female	194,291	23.6	9.6	30,824	32,317	45,859	35,765
Male	360,009	25.3	9.5	64,320	66,161	88,386	72,823
Other	7,521	23.0	9.7	1,362	1,550	1,662	854
NA	1,032	25.7	9.0	166	152	206	189

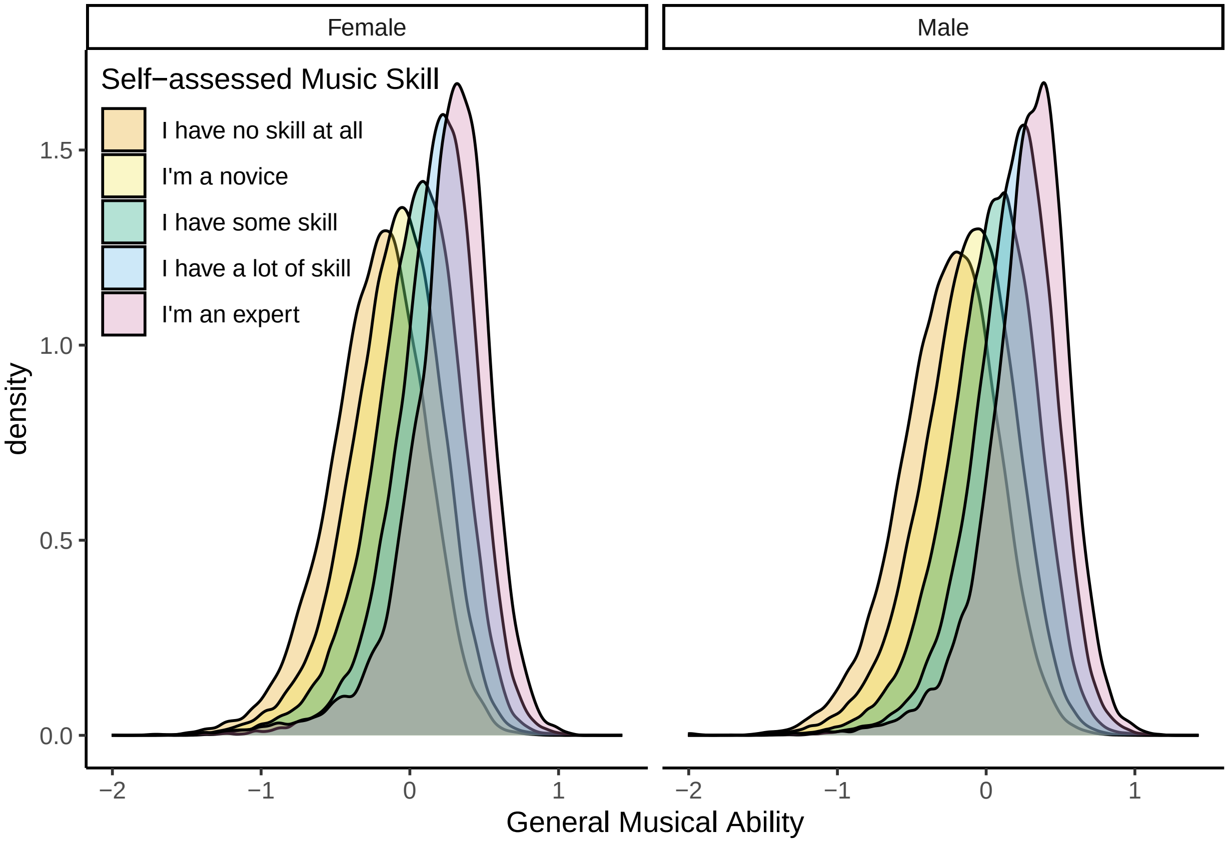
SI Table 1 | Participant information, broken down by self-reported gender, age, and education.

Task	η^2 of gender	R^2 full model
Mistuning Perception	0.001	0.074
Melodic Discrimination	0.001	0.076
Beat Alignment	0.004	0.037
General Music Ability	0.000	0.116

SI Table 2 | Full Model, gender and covariates The estimated gender effect remains of a comparable magnitude when other participant covariates - age, education, age at start of any music lessons, musical training, and musical listening - are added.



SI Figure 1 | Structural Equation Models Top row shows gender-only models, bottom row shows all-covariate models. Left panels in each row show the models for female participants; right panels show these models for males.

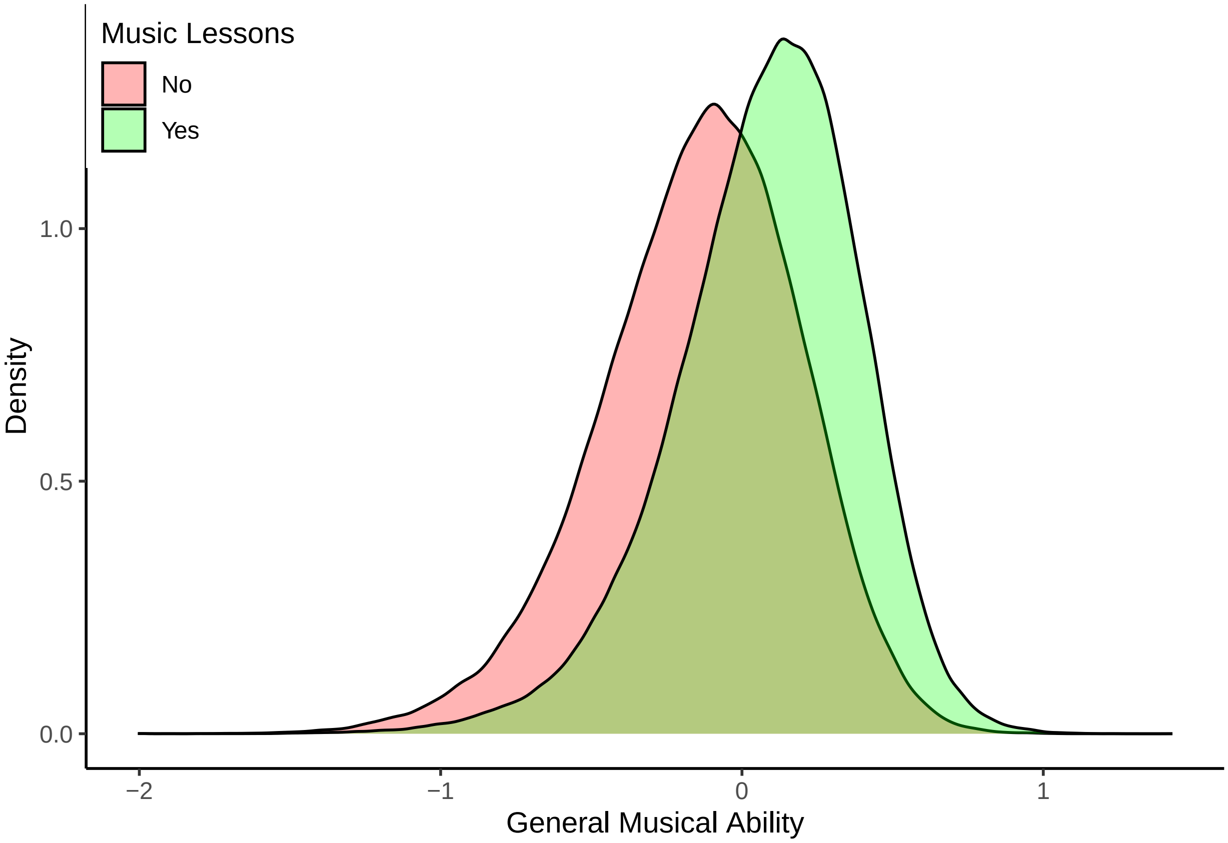


SI Figure 2 | Distribution of General Musical Ability scores by self-assessed musical skill, divided by sex.

Factor	Item	Response Options	Numeric Coding
Extent, Intensity, and Breadth of Musical Listening	Have you ever experienced "chills" or "goosebumps" in response to music?	No, never	1
		Yes, but rarely	2
		Yes, sometimes	3
		Yes, often	4
		I'm not sure	NA
	How familiar are you with traditional music from around the world?	I've never heard traditional music	1
		I'm a little familiar with traditional music	2
		I'm somewhat familiar with traditional music	3
		I'm very familiar with traditional music	4
	On an average day, how much time do you spend listening to music and/or watching videos that include music?	No time at all	1
		1-5 minutes	2
		6-10 minutes	3
		11-15 minutes	4
		16-30 minutes	5
		31-60 minutes	6
		1-2 hours	7
		2-4 hours	8
		More than 4 hours	9
	Have you ever had any music theory training?	No music theory training	1
		A little music theory training	2
		Some music theory training	3
		A moderate amount of music theory training	4
		A lot of music theory training	5
	How much did you enjoy your music lessons?	I disliked them a lot	1
		I disliked them a little	2
		I liked them a little	3
		I liked them a lot	4
		They were a lot better than me	1
		They were a little better than me	2

Formal Musical Training	Think about your peers in the time when you first did music lessons. How did your musical skills compare to theirs?	I was a little better than them	3
		I was a lot better than them	4
		We had about the same level of musical skills	NA
Self-reported Musical Ability	Can you tap in time with a musical beat?	No	0
		Yes	1
		I'm not sure	NA
	How good do you think your music listening skills are? (things like remembering melodies, hearing out of tune notes, or hearing a beat that is out of sync with the music)	'I'm much worse than other people' —[sliding scale]— 'I'm much better than other people'	1
			2
			3
			4
			5
			6
	Think of your skill at making music (using a musical instrument or singing). How would you rate your own skill?	I have no skill at all	1
		I'm a novice	2
		I have some skill	3
		I have a lot of skill	4
		I'm an expert	5
	When you are singing, can you tell if you are out-of-tune or off-key?	No	0
		Yes	1
		I'm not sure	NA

SI Table 3 | Question Items Details of item verbal prompts, response options, and numeric coding of responses.



SI Figure 3 | Participants who had taken music lessons had higher musical ability than paticipants who had not. There was a difference in general musical ability scores of d=0.684 between the those who self report having had or not had music lessons in their life.

References

1. Darwin, C. *The descent of man.* (Watts & Co., 1871).
2. Wiley, R. H. Signal Detection and Animal Communication. in *Advances in the Study of Behavior* vol. 36 217–247 (Academic Press, 2006).
3. Searcy, W. A. & Brenowitz, E. A. Sexual differences in species recognition of avian song. *Nature* **332**, 152–154 (1988).
4. Irestedt, M., Jønsson, K. A., Fjeldså, J., Christidis, L. & Ericson, P. G. An unexpectedly long history of sexual selection in birds-of-paradise. *BMC Evolutionary Biology* **9**, 235 (2009).
5. Ravignani, A. Darwin, sexual selection, and the origins of music. *Trends in Ecology & Evolution* **33**, 716–719 (2018).
6. Fitch, W. T. On the Biology and Evolution of Music. *Music Perception* **24**, 85–88 (2006).
7. Marcus, G. F. Musicality: Instinct or acquired skill? *Topics in Cognitive Science* **4**, 498–512 (2012).
8. Bowling, D. L., Hoeschele, M. & Dunn, J. C. Progress without exclusion in the search for an evolutionary basis of music. *Behavioral and Brain Sciences* **44**, e97 (2021).
9. Zentner, M. Social bonding and credible signaling hypotheses largely disregard the gap between animal vocalizations and human music. *The Behavioral and Brain Sciences* **44**, e120 (2021).
10. Wolf, A. & Kopiez, R. Development and Validation of the Musical Ear Training Assessment (META). *Journal of Research in Music Education* **66**, 53–70 (2018).
11. Sluming, V. A. & Manning, J. T. Second to fourth digit ratio in elite musicians: Evidence for musical ability as an honest signal of male fitness. *Evolution and Human behavior* **21**, 1–9 (2000).
12. Borniger, J. C., Chaudhry, A. & Muehlenbein, M. P. Relationships among Musical Aptitude, Digit Ratio and Testosterone in Men and Women. *PLOS ONE* **8**, e57637 (2013).
13. Miles, S. A., Miranda, R. A. & Ullman, M. T. Sex Differences in Music: A Female Advantage at Recognizing Familiar Melodies. *Frontiers in Psychology* **7**, (2016).
14. Wisniewski, A. B. *et al.* Otoacoustic Emissions, Auditory Evoked Potentials and Self-Reported Gender in People Affected by Disorders of Sex Development (DSD). *Hormones and behavior* **66**, 467–474 (2014).
15. Mosing, M. A. *et al.* Did sexual selection shape human music? Testing predictions from the sexual selection hypothesis of music evolution using a large genetically informative sample of over 10,000 twins. *Evolution and Human Behavior* **36**, 359–366 (2015).
16. Archer, J. The reality and evolutionary significance of human psychological sex differences. *Biological Reviews* **94**, 1381–1415 (2019).
17. Lippa, R. A. Sex Differences in Sex Drive, Sociosexuality, and Height across 53 Nations: Testing Evolutionary and Social Structural Theories. *Archives of Sexual Behavior* **38**, 631–651 (2009).
18. Puts, D. A., Apicella, C. L. & Cárdenas, R. A. Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences* **279**, 601–609 (2011).
19. Archer, J. Sex Differences in Aggression in Real-World Settings: A Meta-Analytic Review. *Review of General Psychology* **8**, 291–322 (2004).
20. Hoeschele, M., Guillette, L. M. & Sturdy, C. B. Biological relevance of acoustic signal affects discrimination performance in a songbird. *Animal Cognition* **15**, 677–688 (2012).
21. Hahn, A. H. *et al.* Discrimination of male black-capped chickadee songs: Relationship between acoustic preference and performance accuracy. *Animal Behaviour* **126**, 107–121 (2017).
22. Rouse, A. A., Patel, A. D., Wainapel, S. & Kao, M. H. Within-species differences in vocal production learning in a songbird are associated with differences in flexible rhythm pattern perception. 2022.07.13.499954 (2023) doi:10.1101/2022.07.13.499954.

23. Christie, P. J., Mennill, D. J. & Ratcliffe, L. M. Pitch shifts and song structure indicate male quality in the dawn chorus of black-capped chickadees. *Behavioral Ecology and Sociobiology* **55**, 341–348 (2004).
24. Mehr, S. A. *et al.* Universality and diversity in human song. *Science* **366**, 957–970 (2019).
25. Yurdum, L. *et al.* Mutual intelligibility in musical communication. *PsyArXiv preprint* (2022) doi:10.31234/osf.io/4kdx6.
26. Lomax, A. Universals in song. *The world of music* **19**, 117–129 (1977).
27. Singh, M. & Mehr, S. A. Universality, domain-specificity and development of psychological responses to music. *Nature Reviews Psychology* 1–14 (2023) doi:10.1038/s44159-023-00182-z.
28. Jacoby, N. *et al.* Universality and cross-cultural variation in mental representations of music revealed by global comparison of rhythm priors. (2021) doi:10.31234/osf.io/b879v.
29. Anglada-Tort, M., Harrison, P. M. C., Lee, H. & Jacoby, N. Large-scale iterated singing experiments reveal oral transmission mechanisms underlying music evolution. *Current Biology* **0**, (2023).
30. Sievers, B., Polansky, L., Casey, M. & Wheatley, T. Music and movement share a dynamic structure that supports universal expressions of emotion. *Proceedings of the National Academy of Sciences* **110**, 70–75 (2013).
31. Fritz, T. *et al.* Universal recognition of three basic emotions in music. *Current Biology* **19**, 573–576 (2009).
32. Trehub, S. E., Unyk, A. M. & Trainor, L. J. Adults identify infant-directed music across cultures. *Infant Behavior and Development* **16**, 193–211 (1993).
33. Mithen, S. *The singing neanderthals: The origins of music, language, mind and body.* (Weidenfeld Nicolson, 2005).
34. Fitch, W. T. The biology and evolution of music: A comparative perspective. *Cognition* **100**, 173–215 (2006).
35. Honing, H. On the biological basis of musicality. *Annals of the New York Academy of Sciences* **1423**, 51–56 (2018).
36. Pinker, S. *How the mind works.* (Norton, 1997).
37. Orians, G. H. *Snakes, sunrises, and Shakespeare: How evolution shapes our loves and fears.* (The University of Chicago Press, 2014).
38. Honing, H. *The Origins of Musicality.* (MIT Press, 2018).
39. Werner, G. M. & Todd, P. M. Too many love songs: Sexual selection and the evolution of communication. (1997).
40. Van Den Broek, E. M. F. & Todd, P. M. Evolution of rhythm as an indicator of mate quality. *Musicae Scientiae* **13**, 369–386 (2009).
41. Miller, G. Evolution of human music through sexual selection. in *The origins of music* (eds. Wallin, N. L., Merker, B. & Brown, S.) 329–360 (MIT Press, 2000).
42. Madison, G., Holmquist, J. & Vestin, M. Musical improvisation skill in a prospective partner is associated with mate value and preferences, consistent with sexual selection and parental investment theory: Implications for the origin of music. *Evolution and Human Behavior* **39**, 120–129 (2018).
43. Novaes, F. C. & Natividade, J. C. The sexual selection of creativity: A nomological approach. *Frontiers in Psychology* **13**, (2023).
44. Ryan, M. J. *et al.* Nineteen Years of Consistently Positive and Strong Female Mate Preferences despite Individual Variation. *The American Naturalist* **194**, 125–134 (2019).
45. Ligon, R. A. *et al.* Evolution of correlated complexity in the radically different courtship signals of birds-of-paradise. *PLOS Biology* **16**, e2006962 (2018).

46. Green, D. M. & Swets, J. A. *Signal detection theory and psychophysics.* xi, 455 (John Wiley, 1966).

47. Trivers, R. L. Parental investment and sexual selection. in *Sexual selection and the descent of man* (ed. Campbell, B. G.) 136–179 (Aldine, 1972).

48. Dabelsteen, T. & Pedersen, S. B. Song-based species discrimination and behaviour assessment by female blackbirds, Turdus merula. *Animal Behaviour* **45**, 759–771 (1993).

49. Otter, K. & Ratcliffe, L. Female initiated divorce in a monogamous songbird: Abandoning mates for males of higher quality. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **263**, 351–355 (1997).

50. Bernal, X. E., Stanley Rand, A. & Ryan, M. J. Sexual Differences in the Behavioral Response of Túngara Frogs, Physalaemus pustulosus, to Cues Associated with Increased Predation Risk. *Ethology* **113**, 755–763 (2007).

51. Krizman, J., Rotondo, E. K., Nicol, T., Kraus, N. & Bieszczad, K. M. Sex differences in auditory processing vary across estrous cycle. *Scientific Reports* **11**, 22898 (2021).

52. Mehr, S. A., Krasnow, M. M., Bryant, G. A. & Hagen, E. H. Origins of music in credible signaling. *Behavioral and Brain Sciences* **44**, e60 (2021).

53. Peretz, I. & Vuvan, D. T. Prevalence of congenital amusia. *European Journal of Human Genetics* **25**, 625–630 (2017).

54. Guéguen, N., Meineri, S. & Fischer-Lokou, J. Men's music ability and attractiveness to women in a real-life courtship context. *Psychology of Music* **42**, 545–549 (2014).

55. Guéguen, N., Meineri, S. & Fischer-Lokou, J. "Men's music ability and attractiveness to women in a real-life courtship context": Expression of concern. *Psychology of Music* **49**, 687–687 (2021).

56. Brown, W. M. *et al.* Dance reveals symmetry especially in young men. *Nature* **438**, 1148–1150 (2005).

57. Brown, W. M. *et al.* Retraction Note: Dance reveals symmetry especially in young men. *Nature* **504**, 470–470 (2013).

58. Charlton, B. D., Filippi, P. & Fitch, W. T. Do women prefer more complex music around ovulation? *PLoS ONE* **7**, e35626 (2012).

59. Charlton, B. D. Menstrual cycle phase alters women's sexual preferences for composers of more complex music. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20140403 (2014).

60. Larrouy-Maestri, P., Harrison, P. M. C. & Müllensiefen, D. The mistuning perception test: A new measurement instrument. *Behavior Research Methods* **51**, 663–675 (2019).

61. Harrison, P. M. C., Collins, T. & Müllensiefen, D. Applying modern psychometric techniques to melodic discrimination testing: Item response theory, computerised adaptive testing, and automatic item generation. *Scientific Reports* **7**, 1–18 (2017).

62. Harrison, P. M. C. & Müllensiefen, D. Development and validation of the Computerised Adaptive Beat Alignment Test (CA-BAT). *Scientific Reports* **8**, 1–19 (2018).

63. Hyde, J. S. Gender Similarities and Differences. *Annual Review of Psychology* **65**, 373–398 (2014).

64. Zell, E., Krizan, Z. & Teeter, S. Global Gender Differences Can Be Operationalized and Tested. *The American psychologist* **70**, 664–665 (2015).

65. Liu, J., Hilton, C. B., Bergelson, E. & Mehr, S. A. Language experience predicts music processing in 1/2 million speakers of 54 languages. (2023) doi:10.1101/2021.10.18.464888.

66. Ikebuchi, M., Futamatsu, M. & Okanoya, K. Sex differences in song perception in Bengalese finches measured by the cardiac response. *Animal Behaviour* **65**, 123–130 (2003).

67. Cohen, J. *Statistical power analysis for the behavioral sciences.* (L. Erlbaum Associates, 1988).