Vocal similarity predicts the relative attraction of musical chords

Daniel L. Bowling[a,1], Dale Purves[b,1], and Kamraan Z. Gill[c]

[a]Department of Cognitive Biology, University of Vienna, Vienna 1090, Austria; [b]Duke Institute for Brain Sciences, Duke University, Durham, NC 27708; and [c]Department of Pathology, CBLPath, Rye Brook, NY 10573

Contributed by D. Purves, November 21, 2017 (sent for review July 27, 2017; reviewed by Aniruddh D. Patel and Laurel J. Trainor)

Musical chords are combinations of two or more tones played together. While many different chords are used in music, some are heard as more attractive (consonant) than others. We have previously suggested that, for reasons of biological advantage, human tonal preferences can be understood in terms of the spectral similarity of tone combinations to harmonic human vocalizations. Using the chromatic scale, we tested this theory further by assessing the perceived consonance of all possible dyads, triads, and tetrads within a single octave. Our results show that the consonance of chords is predicted by their relative similarity to voiced speech sounds. These observations support the hypothesis that the relative attraction of musical tone combinations is due, at least in part, to the biological advantages that accrue from recognizing and responding to conspecific vocal stimuli.

music | consonance | evolution | vocalization

Music comprises periodically repeating sound signals (tones) that are combined sequentially as melodies or simultaneously as harmonies (1). Although the tones used to make music vary among different traditions, the frequency relationships between tones (musical intervals) are remarkably consistent across cultures and musical styles (2, 3). In particular, relationships defined by small integer ratios, such as 2:1 (the octave), 3:2 (perfect fifth), and 4:3 (perfect fourth) play important roles in major traditions from Europe, Africa, India, the Middle East, and East Asia (4). Moreover, based on ancient texts and surviving instruments, these relationships have been relatively stable over time (5–10).

The prevalence of intervals in music is closely related to their consonance, a term defined in musicological literature as the "affinity" between tones, as well as the clarity, stability, smoothness, fusion, and pleasantness that arise from their combination (11). In the current study, we define consonance as the subjective attractiveness of tone combinations. In general, the intervals that occur most frequently across cultures and historical eras correspond to those considered the most consonant by culturally diverse listeners (2, 12, 13). Various theories have sought to explain consonance, but their merits remain debated (14–23). Over the last decade, however, evidence has accumulated that vocal similarity can account for many features of tonal music (12, 19, 24–31). In this interpretation, the appeal of a particular tone combination is based on the relative resemblance of its spectrum to the spectra that characterize human vocalization. The rationale for this theory is that tonal sound stimuli in nature are effectively limited to animal sources, the most biologically important of which are typically conspecific vocalizations. A key feature of human (and many other animal) vocalizations that distinguishes them from inanimate environmental sounds is the harmonic series of acoustic vibrations produced by the quasiperiodic vibration of vocal membranes. Because these spectra—whether prelingual or as speech—harbor critical information about the physical size, age, gender, identity, and emotional state of the vocalizer, selective (and developmental) pressure on their perceptual appeal would have been intense. The implication is that the perceptual mechanisms we use to contend with tonal stimuli have been fundamentally shaped by the benefits of recognizing and responding to conspecific vocalization. Accordingly, we here ask whether the consonance of tone combinations in music can be rationalized on this basis: that is, whether our attraction to specific chords is predicted by their relative similarity to human vocalization.

Answering this question requires perceptual data that document the relative consonance of chords. Previous evaluations have focused on the two-tone combinations ("dyads") that define the chromatic scale, a set of 12 tones over an octave used in much music worldwide (Table S1). Studies of dyadic consonance have been repeated many times over the last century and, despite some variation in consonance ranking, listeners broadly agree on the dyads heard as the most and least attractive (12, 32). Surprisingly, comparable perceptual data are not available for more complex tone combinations, such as triads (three-tone chords) and tetrads (four-tone chords). Studies that have examined the consonance of some of these higher-order chords have typically focused on the small set commonly used in popular music [e.g., the major, minor, augmented, and diminished triads, and various seventh chords (33–36)]. We are aware of only two studies that have examined triads and tetrads more broadly; one did not include perceptual data (37), while the other did not specify which chords were tested (38). Thus, earlier studies have examined only a small fraction of the data available for evaluating theories of consonance biased in favor of chords prevalent in popular music.

Among the reasons why previous investigations have focused on dyads is that psychophysical theories designed to predict consonance often fail when applied to more complex chords (39). This deficiency has led some investigators to argue that the perception of higher-order chords is dominated by cultural learning and is therefore not amenable to principled analysis (40). It is not clear, however, why a perceptual attribute as fundamental to music as

Significance

The foundations of human music have long puzzled philosophers, mathematicians, psychologists, and neuroscientists. Although virtually all cultures uses combinations of tones as a basis for musical expression, why humans favor some tone combinations over others has been debated for millennia. Here we show that our attraction to specific tone combinations played simultaneously (chords) is predicted by their spectral similarity to voiced speech sounds. This connection between auditory aesthetics and a primary characteristic of vocalization adds to other evidence that tonal preferences arise from the biological advantages of social communication mediated by speech and language.

Author contributions: D.L.B., D.P., and K.Z.G. designed research; D.L.B. performed research; D.L.B. and K.Z.G. analyzed data; and D.L.B., D.P., and K.Z.G. wrote the paper.

Reviewers: A.D.P., Tufts University; and L.J.T., McMaster University.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. Email: dan.bowling@univie.ac.at or purves@neuro.duke.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1713206115/-/DCSupplemental.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

consonance should be limited in this way, nor why the influence of cultural learning should be dependent on the number of tones in a chord. In contrast, we argue that a robust theory of consonance should be able to explain the relative attraction of any tone combination, regardless of the number of tones involved (25). Thus, in addition to retesting the 12 chromatic dyads, we determined average consonance ratings by listeners of all 66 possible chromatic triads and all 220 possible chromatic tetrads that can be formed over a single octave. We then measured the degree of vocal similarity by two metrics and compared the results with perceived consonance.

Methods

Subjects. Thirty subjects took part in the study. Fifteen (eight male) were students at the Yong Siew Toh Conservatory of Music at the National University of Singapore (age range = 18–27 y). These subjects had taken formal lessons in Western tonal music at least once per week for an average of 13 y (SD = 3.8). The other 15 subjects (eight male) were students recruited from the University of Vienna (age range = 19–28 y). These subjects had less than a year of weekly music lessons on average (SD = 1.1). Ethical approval was provided by the University of Vienna Ethics Committee.

Stimuli. The chords tested were all 12 dyads, 66 triads, and 220 tetrads that can be formed using the intervals specified by the chromatic scale over one octave. The fundamental frequencies (F0s) of the tones in each chord were adjusted such that the mean F0 of all tones was 263 Hz (middle C; tone F0s ranged from 174 to 349 Hz, ~F3–F4), and the intervals between them were tuned using just intonation ratios (Table S1). Individual tones were created using the "Bosendorfer Studio Model" piano in Logic Pro-9 (v9.1) (41). Stimuli were presented over headphones (ATH-M50; Audio Technica; DT 770 PRO; Beyerdynamic) with the volume adjusted to a comfortable level for each participant before starting the experiment and then held constant.

Procedure. Upon arrival, participants were given written instructions explaining the concept of consonance and what would be required of them. The instructions defined consonance as "the musical pleasantness or attractiveness of a sound," and that "if a sound is relatively unpleasant or unattractive, it is referred to as dissonant." After reading the instructions, participants provided written informed consent and were played six example dyads to expose them to the general range of consonance/dissonance evaluated in prior studies; the order of the chords was octave, perfect fifth, minor third, tritone, major seventh, and minor second, and they were told that the progression moved from chords "typically considered more consonant to chords typically considered more dissonant." On each trial of the experiment, the participants heard a single chord and provided a rating of consonance/dissonance using a four-point scale (1 = "quite consonant"; 2 = "moderately consonant"; 3 = "moderately dissonant"; 4 = "quite dissonant") (42). Participants could listen to a chord as many times as they wished before entering a rating. Dyads, triads, and tetrads were tested in separate blocks, with the order of stimuli within each block randomized across subjects. Each dyad was rated multiple times (allowing assessment of intrarater reliability), whereas each triad and tetrad was only rated once.

Statistics. For each chord, the mean consonance rating was calculated across subjects. One-way ANOVAs were used to test for significant differences between mean consonance ratings for each chord type, and Tukey's range tests were used to determine which specific pairs of chords (out of all possible pairs) were significantly different while maintaining a family-wise α level at 0.05 (43). Measures of intrarater and interrater reliability were calculated using intraclass correlations (ICCs) (44), the values of which were interpreted according to the guidelines provided in Koo and Li (45).

Assessing Vocal Similarity. For each significant difference in the consonance ratings of two chords, we determined whether or not the chord with greater vocal similarity was judged to be more consonant. We tested vocal similarity in two ways. We first examined how closely the pattern of harmonics in each chord mimicked the single harmonic series characterizing human vocalizations (Fig. S1). The initial step in this analysis was to determine the single series that contained all harmonics in the chord. The fundamental frequency (F0) of this series was calculated as the greatest common divisor (GCD) of the F0s of the tones in the chord. For example, the GCD of a major triad comprising tones with F0s of 400, 500, and 600 Hz is 100 Hz. Thus, the single harmonic series with an F0 of 100 Hz contains all harmonics present in the triad. The next step was to calculate the percentage of harmonics in the single harmonic series that were also present in the chord. We refer to this percentage as the chord's harmonic similarity score; its use as an index of vocal similarity is justified by the fact that the voice is the primary source of harmonic stimulation in a natural auditory environment (see above). The highest frequency considered in the calculation of the harmonic similarity score was the least-common multiple (LCM) of the F0s of the chord (after which the pattern of harmonics repeats). The vocal similarity hypothesis predicts that a chord with a higher harmonic similarity score more closely mimics the pattern of harmonics heard in vocalizations and will thus be more attractive than a chord with a lower harmonic similarity score.

Although harmonic similarity provides a way to compare the spectra of chords with the harmonic patterns found in the voice, it does not address another critical feature of vocalization: the absolute frequency intervals between harmonics. Thus, in a second approach, we compared the absolute frequency intervals between the tones in each chord to the absolute frequency intervals that occur between harmonics in human vocalizations. The intervals between harmonics in the voice are determined by the F0 of phonation, which is restricted by the physical properties of the larynx. Although the human larynx operates across a wide range of F0s (46, 47), studies of vocal range in speech and singing indicate a lower limit of ~50 Hz (48). Because each harmonic is an integer multiple of the F0, the minimum absolute frequency interval between successive harmonics typically encountered in human vocalizations is ~50 Hz. Accordingly, chords containing intervals smaller than 50 Hz are treated here as having lower vocal similarity and are predicted to be heard as less consonant. This analysis was limited to the intervals between the F0s of chord tones because these represent the most powerful harmonics (at least for the piano tones we used). This second metric was only applicable when the minimum interval between the F0s of one or both chords in a given comparison was <50 Hz. If both chord comprised intervals <50 Hz, the chord with the greater minimum interval was predicted to be more consonant. See *SI Methods* for further details.

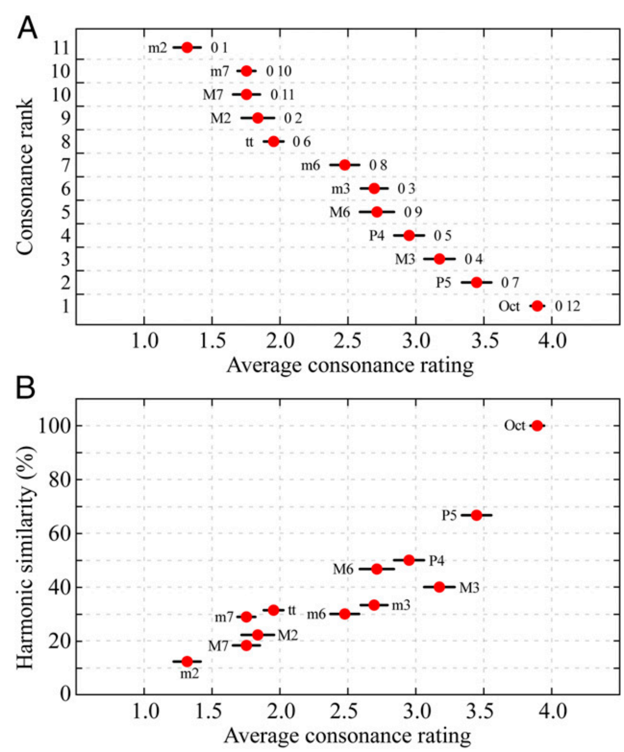

Fig. 1. Dyad ratings. (*A*) Mean consonance ratings calculated across all 30 subjects for the 12 chromatic dyads, sorted from lowest to highest and ranked (equal ranks assigned to chords with the same mean). Each dyad is labeled with an abbreviation of its common name (full names in Table S1) and its component tones, as specified by a list of numbers corresponding to semitone intervals above the lowest tone (labeled "0"). (*B*) The mean consonance ratings in *A* plotted against harmonic similarity score (*Methods*). Error bars represent ±1 SEM.

Results

Analysis of Subject Groups. The overall patterns of ratings for all chord types were similar for musically trained and untrained subjects. Spearman correlations between means calculated separately for each group were $r = 0.93$ for dyads, 0.92 for triads, and 0.88 for tetrads (Ps < 0.0001). The chords considered most and least consonant were also similar in both groups. The average absolute difference between group means for the same chord was less than half a scale point for dyads (mean = 0.4, SD = 0.26), triads (mean = 0.28, SD = 0.2), and tetrads (mean = 0.36, SD = 0.23). Given the degree of similarity in consonance ratings between the two subject groups, the analyses that follow are based on data from all 30 subjects combined. See *Supporting Information* for additional comparisons of musicians vs. nonmusicians.

Dyads. The mean consonance ratings for all 12 dyads are shown in Fig. 1. Intrarater reliability analyses showed that 29 of the 30 subjects exhibited "moderate" or "good" consistency across multiple ratings of the same chord (ICCs ranging from 0.54 to 0.89) (Table S2). The one exceptional subject showed extreme variation across repeated ratings, with an ICC falling nearly three SDs below that of the subject with the next lowest value (0.06 vs. 0.54). However, exclusion of this subject only had minimal effects on the overall results; all data are thus retained in subsequent analyses. The analyses of interrater reliability showed that as a group, subjects exhibited "moderate" consistency in their ratings of the same dyads (single measures ICC = 0.7) (Table S3). However, the reliability of the average consonance ratings (calculated across 30 subjects) was determined to be "excellent" (average measures ICC = 0.99). The average consonance ratings are thus highly reliable, justifying their use in subsequent analyses (44). ANOVA analysis indicated that there were significant differences between average ratings of individual chords [$F_{(11,\ 348)} = 63.08$, $P < 0.0001$]. Pairwise comparisons indicated that 76% of all possible dyad pairings (50 of 66) were perceived as significantly different in consonance. The harmonic similarity analysis correctly predicted the chord perceived as more consonant in 96% (48 of 50) of these cases. The frequency intervals analysis was applicable in 44% (22 of 50) of the cases and correctly predicted the chord perceived as more consonant in 86% (19 of 22) of them. At least one metric of vocal similarity correctly predicted perceived consonance in 96% of the pairwise comparisons between dyads determined to be significantly different at the group level (48 of 50). See *Supporting Information* for discussion of the two significant consonance differences (4%) incorrectly predicted by these metrics.

Triads. The mean consonance ratings for all 66 triads are shown in Fig. 2. The analyses of interrater reliability showed that as a group, subjects exhibited "moderate" consistency in their ratings of the same triads (single measures ICC = 0.59) (Table S2). However, the reliability of the average consonance ratings (calculated across all 30 subjects) was again determined to be "excellent" (average measures ICC = 0.98), indicating high reliability and justifying their further use (44). ANOVA analysis indicated that there were significant differences between average ratings of individual chords [$F_{(65,\ 1,914)} = 37.97$, $P < 0.0001$]. Pairwise comparisons showed that 50% of all possible triad pairings (1,065 of 2,145) were perceived as significantly different in consonance. The harmonic similarity analysis correctly predicted the chord perceived as more consonant in 86% (925 of 1,065) of these cases. The frequency intervals analysis was applicable in 93% (995 of 1,065) of the cases and correctly predicted the chord perceived as more consonant in 90% (894 of 995) of them. At least one metric of vocal similarity correctly predicted perceived consonance in 97% of the pairwise comparisons between triads determined to be significantly different at the group level (1,035 of 1,065). See *Supporting Information* for discussion of the 30 significant consonance differences (3%) incorrectly predicted by these metrics.

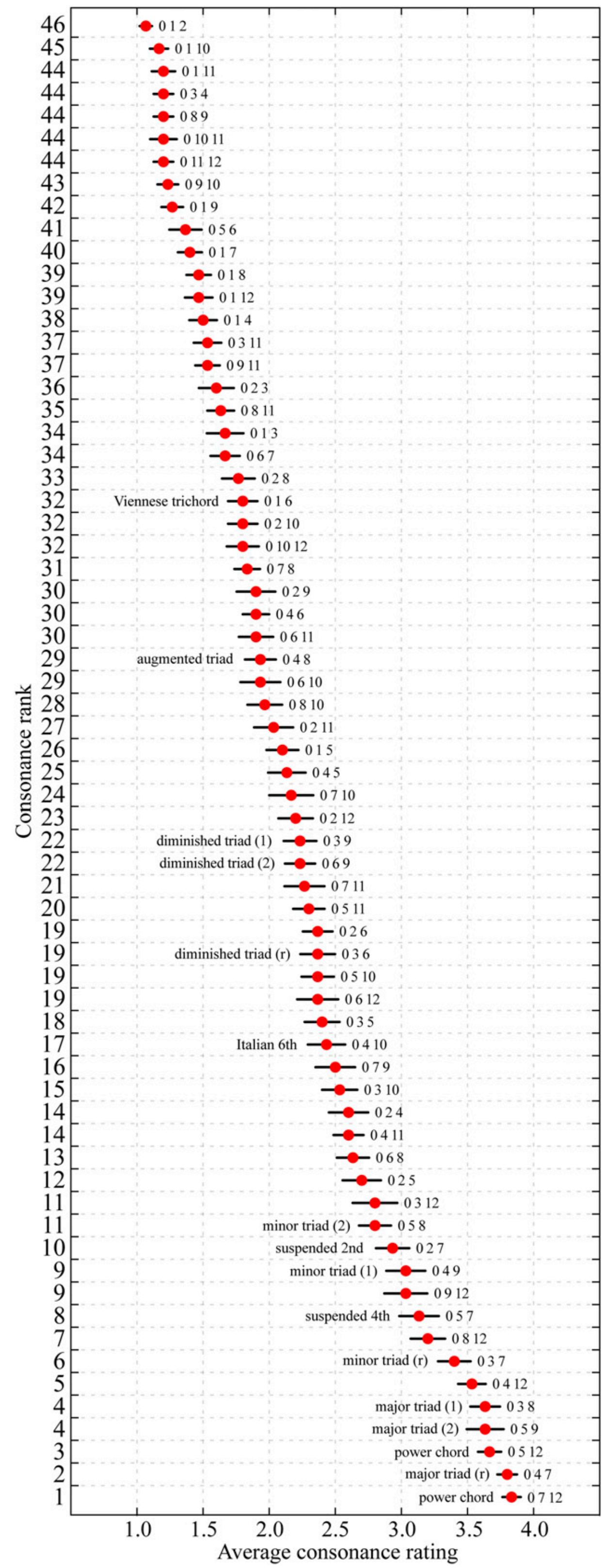

Fig. 2. Triad ratings. Mean consonance ratings calculated across all 30 subjects for the 66 chromatic triads, sorted from lowest to highest and ranked. Triads with common names are labeled accordingly (inversions are labeled for the major, minor and diminished triads: r = root, 1 = first inversion, 2 = second inversion). The format is otherwise the same as Fig. 1*A*.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Tetrads. The mean consonance ratings for a subset of the 220 tetrads tested are shown in Fig. 3. The analyses of interrater reliability showed that as a group, subjects exhibited "poor" consistency in their ratings of the same tetrad (single measures ICC = 0.46) (Table S2). This was primarily due to the large increase in chords associated with moderate levels of consonance/dissonance relative to dyads and triads; consistency for chords with more extreme consonance/dissonance was "moderate" (single measures ICC = 0.61; calculated on the top and bottom quartiles of the average ratings). Additionally, the reliability of the average consonance ratings calculated across all 30 subjects was again "excellent" (average measures ICC = 0.96), justifying their further use (44). ANOVA analysis indicated that there were significant differences between average ratings of individual chords [$F_{(219,\ 6,380)} = 21.24$, $P < 0.0001$]. Pairwise comparisons indicated that 30% of all possible tetrad pairings (7,206 of 24,090) were perceived as significantly different in consonance. The harmonic similarity analysis correctly predicted the chord perceived as more consonant in 83% (6,013 of 7,206) of these cases. The frequency intervals analysis was applicable in 100% (7,206 of 7,206) of significant cases and correctly predicted the chord perceived as more consonant in 93% (6,669 of 7,206) of them. At least one metric of vocal similarity correctly predicted perceived consonance in 99% of the pairwise comparisons determined to be significantly different at the group level (7,101 of 7,206). See *Supporting Information* for discussion of the 105 significant consonance differences (1%) incorrectly predicted by these metrics.

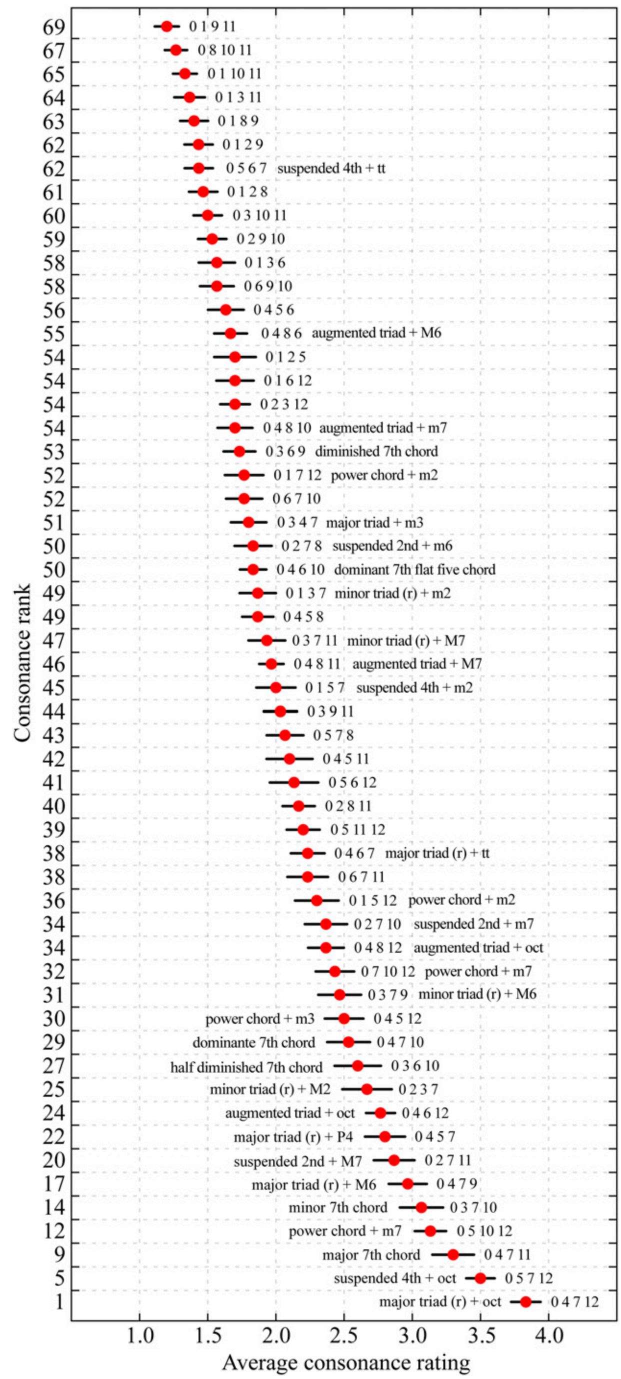

Fig. 3. Tetrad ratings. Mean consonance ratings calculated across all 30 subjects for a subset of the chromatic tetrads sorted from lowest to highest and ranked (see Fig. S2 for complete tetrad ratings). Tetrads with common names and tetrads that are extensions of common triads are labeled accordingly (inversions not labeled). The format is otherwise the same as Fig. 1*A*.

Discussion

Ratings of consonance for every possible chromatic dyad, triad, and tetrad within a single octave were obtained from 30 subjects comprising both musicians and nonmusicians. Statistical analyses showed significant differences in average perceived consonance for all three chord types, with a total of 8,321 of 26,301 pair-wise comparisons (32%) being identified as reliably evoking different consonance percepts at the group level. For the vast majority of these (98%), the chord perceived as more consonant was correctly predicted by at least one of the two metrics used to evaluate vocal similarity. Indeed, a large majority (78.6%) was predicted by both methods. This outcome implies that, like other tonal features of music (12, 19, 24–31), the consonance of musical chords can be rationalized in terms of vocal similarity.

The metrics of vocal similarity used here are based on two fundamental aspects of vocal spectra: that is, their harmonic structure and the minimum frequency intervals between harmonics. The importance of harmonic structure for understanding tonal aesthetics in music has been appreciated since Rameau (49). More recently, the importance of harmonic structure has also been emphasized by neuroscientists, psychoacousticians, and psychologists (12, 18, 20, 21, 25, 40, 50–53). Although these authors have taken different approaches to evaluating the harmonic structure of tone combinations, the most detailed approach has been that of Parncutt (40). In his psychoacoustical model of harmony, Parncutt combined estimates of masking with a harmonic template matching procedure to calculate the "complex tonalness" of a chord, which he then used to describe the extent to which it evoked the perception of a single harmonic series [equal to the perceptual weight associated with the virtual pitch best supported by the chord's spectra (40)]. However, it is not clear that the added complexity of Parncutt's model corresponds to an increase in the predictive power demonstrated here. For example, Cook and Fujisawa (39) point out that Parncutt's model predicts that the augmented triad (semitone intervals = 0 4 8) is more consonant than the first and second inversions of the major triad (0 3 8 and 0 5 9, respectively) as well as all inversions of the minor triad (0 3 7, 0 4 9, and 0 5 8). These predictions are incorrect based on the psychoacoustical data obtained here and in other studies (33, 38). In acknowledgment of these discrepancies, Parncutt (54) points out that "... so far, no psychoacoustical model has succeeded in predicting the relative perceived consonance of common musical chords such as the major, minor, augmented, and diminished triads." In contrast, the harmonic similarity metric used here correctly predicts that augmented triads are perceived as less consonant than all inversions of the major and minor triads (*Supporting*

Information). A final point regarding the harmonic similarity metric is that although the calculations we describe are designed to assess chords tuned using just intonation, it would be straightforward to adapt them for less harmonically precise tuning systems (e.g., 12-tone equal temperament) by introducing a tolerance window for judgments of overlap with the GCD harmonic series.

The rationale for the frequency-intervals metric is that the size of the frequency intervals between harmonics in human vocalizations is limited by the range of F0s our larynx can produce. This important aspect of vocalization is not captured by the harmonic similarity metric, which only assesses the overall harmonic pattern of a chord. The frequency-intervals metric addresses harmonic spacing by predicting chords with harmonics that are closer together than those in human vocalizations (less than ∼50 Hz) to be less consonant. In principle, chords with harmonics farther apart than those in human vocalizations would also be predicted to be less consonant, but this principle did not apply here because the largest interval between tone F0s in the chords we tested was only 174 Hz, far below the upper limit of human phonation. Although conceptually different, the frequency-intervals metric bears some similarity to the "roughness" calculations made by many previous models of consonance, which also treat chords with closely spaced harmonics as dissonant (14, 37, 40, 50, 55–60). This metric has several advantages over analyses of roughness. First, it avoids the flawed assumption that consonance is equal to an absence of roughness (12, 20, 51, 60, 61). Second, it preempts historical disagreements about how to estimate perceived roughness. For example, roughness models usually assume that maximum roughness occurs at some proportion of the critical bandwidth, but disagree about what this proportion is (40, 55, 62). It is also unclear how to combine the roughness resulting from different harmonic interactions into a single value that accurately represents the associated percepts, particularly for chords with more than two tones (40, 55, 57, 63, 64). In sum, compared with previous models that have sought to estimate consonance by an assessment of harmonic structure or by roughness calculations, the approach taken here accords more closely with the available empirical data, is conceptually and computationally simple, and is embedded within a theoretical framework that provides a clear biological rationale for why we are attracted to particular tone combinations.

Apart from showing that vocal similarity can account for the consonance of chords, the main contribution of this work is the empirical derivation of average consonance ratings for all possible dyads, triads, and tetrads within a single octave. The results are relevant to the design of future experiments. They show that not all differences in consonance assumed by music theory are empirically verifiable (at least not with 30 subjects and the response scale used here). For example, the major triad in root position (semitone intervals: 0 4 6) was not perceived as significantly different in consonance from the minor triad in root position (0 3 7) (average consonance ratings = 3.8 vs. 3.4 respectively, $P = 0.995$). This observation is particularly important because studies of triadic consonance and other higher-order chords have tended to limit their focus to these and other popular chords (see earlier). Because the popularity of chords in music is related to their aesthetic appeal, testing only popular chords creates a bias toward attractive tone combinations, reducing contrast between stimuli and requiring subjects to make what may be unreasonably subtle distinctions. When attempting to measure tonal preferences in subjects with very little musical experience (e.g., infants and nonhuman animals), or people with limited exposure to chords, using stimuli with reduced contrast decreases the likelihood of detecting consonance preferences, simply because the subjects are being asked to discriminate between very similar stimuli (4). The average consonance ratings and associated statistics derived here (provided in the *Supporting Information*) offer an empirical basis for selecting chords that would be most appropriate for such experiments.

Cross-species studies offer a way to test the generality of vocal similarity theory. For species that rely on harmonic vocalizations for social communication, vocal similarity theory predicts some form of attraction to consonant compared with dissonant tone combinations. Experiments assessing tonal preferences in animals have typically used an acoustic place preference paradigm in which consonant/dissonant chords are played through speakers and the subject's proximity to those speakers is the main dependent variable. We are aware of studies in four species, all of which have some harmonic calls in their repertoires (65–68). The results are mixed, with evidence in support of a preference for consonance in chickens [*Gallas gallas*, $n = 81$ (69)] and chimpanzees [*Pan Troglodytes*, $n = 1$ (70)], and evidence against consonance preferences in Cotton-top tamarins [*Saguinus oesdipus*, $n = 6$ (71)] and Campbell's mona monkeys [*Cercopithecus campbelli*, $n = 6$ (72)]. Further studies are thus required to resolve this issue. If such studies aim to test vocal similarity theory, it is essential that the stimuli be customized to reflect the acoustical properties of vocalizations produced by the species in question, both in terms of vocal range as well as other acoustic parameters, such as duration, intensity, and timbre. Attention should also be paid to minimizing stress associated with being exposed to novel/stressful circumstances: for example, by avoiding aversively loud noise and encouraging voluntary participation.

Another key prediction of vocal similarity theory is that the auditory system is more effectively stimulated by tone combinations with spectra resembling harmonic vocalizations. Evidence in support of this prediction comes from two recent neural models of consonance perception. In the "neural pitch salience model," consonant chords stimulate stronger periodic activity at early stations of the auditory pathway, increasing the salience of particular pitches and enhancing their cortical processing (73–76). In the "neurodynamic model," consonant chords stimulate more stable patterns of resonant activity in neural oscillators through mode-locking between populations with sympathetic intrinsic frequencies (21). In both models, the key aspect of consonant chords is that their spectra comprise harmonically related frequencies. Because this aspect of consonance is also captured by the metrics of vocal similarity used here, both models (nonexclusively) represent potential mechanistic realizations of vocal similarity. A related point is that although we use two metrics to assess vocal similarity here, they do not necessarily represent distinct neural processes. Indeed, it seems more likely that the neural response is unitary, responding to harmonic similarity only when harmonics are appropriately spaced.

Finally, given ongoing controversy over the roles of biology and culture in determining consonance perception (4, 42), it is important to clarify the implications of vocal similarity theory in this context. It seems fair to reject the attempt to treat biology and culture (nature and nurture) as separate influences on tone perception. For example, Parncutt (40) argues that nature and nurture can be usefully opposed in terms of innate versus acquired, arguing that the physiology of sensory organs is innate, whereas the guiding principles of particular musical traditions are arbitrary. Similarly, McDermott et al. (42) recently concluded that consonance is primarily a result of exposure to Western music rather than auditory system neurobiology (see also ref. 16). This approach is problematic, not only because culture is itself a biological phenomenon, but because auditory neurobiology is shaped by experience. Accordingly, it is misleading to characterize the influence of biology on tone perception as "innate," or the influence of culture as arbitrary. Genes do not encode auditory percepts; they make proteins that interact in complex environmentally modulated networks to build and maintain nervous systems (77). Similarly, trends with no biological appeal seldom enjoy widespread popularity. Vocal similarity theory assumes that consonance perception arises through the evolutionary and developmental interaction of auditory neurobiology with tonal stimuli in the environment, including primarily speech and music.

Conclusion

The vast majority of significant differences in musical chord preferences are predicted by simple metrics that evaluate spectral similarity to human vocalizations. These results support the hypothesis that tonal preferences in music are linked to an inherent attraction to conspecific vocalizations and the biological rewards that follow.

ACKNOWLEDGMENTS. The authors thank Isabella De Cuntis for running subjects in Vienna. This work was funded in part by a grant from the Austrian Science Fund (M 1773-B24).

1. Brown S, Jordania J (2013) Universals in the world's musics. *Psychol Music* 41:229–248.
2. Burns EM (1999) Intervals, scales, and tuning. *The Psychology of Music*, ed Deustch D (Academic, San Diego), 2nd Ed, pp 215–264.
3. Forster C (2010) *Musical Mathematics* (Chronicle Books, San Francisco).
4. Bowling D, Hoeschele H, Gill K, Fitch W (2017) The nature and nurture of musical consonance. *Music Percept* 35:188–121.
5. Sambamoorthy P (1960) *History of Indian Music* (The Indian Music Publishing House, Chennai, India).
6. Chen JCY (1996) *Early Chinese Work in Natural Science* (Hong Kong Univ Press, Hong Kong).
7. Crocker RL (1963) Pythagorean mathematics and music. *J Aesthet Art Crit* 22:189–198.
8. Zhang J, Harbottle G, Wang C, Kong Z (1999) Oldest playable musical instruments found at Jiahu early Neolithic site in China. *Nature* 401:366–368.
9. Conard NJ, Malina M, Münzel SC (2009) New flutes document the earliest musical tradition in southwestern Germany. *Nature* 460:737–740.
10. Münzel S, Seeberger F, Hein W (2002) The Geißenklösterle flut—Discovery, experiments, reconstruction. *Studies in Music Archaeology III*, Orient-Archäologie Series, eds Hickmann E, Kilmer AD, Eichmann R (VML Maria Leidorf GmbH, Rahden, Germany), pp 107–118.
11. Tenney J (1988) *A History of "Consonance" and "Dissonance"* (Excelsior Music, New York).
12. Bowling DL, Purves D (2015) A biological rationale for musical consonance. *Proc Natl Acad Sci USA* 112:11155–11160.
13. Carterette EC, Kendall RA (1999) Comparative music perception and cognition. *The Psychology of Music*, ed Deustch D (Academic, San Diego), pp 725–791.
14. von Helmholtz H (1885) *On the Sensations of Tone* (Dover, New York), 4th Ed.
15. Stumpf K (1898) Konsonanz und Dissonanz. *Beiträge zur Akust und Musikwissenschaft* 1:1–108.
16. Lundin RW (1947) Toward a cultural theory of consonance. *J Psychol* 23:45–49.
17. Boomsliter P, Creel W (1961) The long pattern hypothesis in harmony and hearing. *J Music Ther* 5:2–31.
18. Terhardt E (1984) The concept of musical consonance: A link between music and psychoacoustics. *Music Percept* 1:276–295.
19. Schwartz DA, Howe CQ, Purves D (2003) The statistical structure of human speech sounds predicts musical universals. *J Neurosci* 23:7160–7168.
20. McDermott JH, Lehr AJ, Oxenham AJ (2010) Individual differences reveal the basis of consonance. *Curr Biol* 20:1035–1041.
21. Large EW, Kim JC, Flaig NK, Bharucha JJ, Krumhansl CL (2016) A neurodynamic account of musical tonality. *Music Percept* 33:319–331.
22. Huron D (2002) A new theory of sensory dissonance: A role for perceived numerosity. *Proceedings of the Seventh International Conference for Music Perception and Cognition*, eds Stevens C, Burnham D, McPherson G, Schubert E, Renwick J (Causal Productions, Adelaide, Australia), pp 273–276.
23. Langner G, Ochse M (2006) The neural basis of pitch and harmony in the auditory system. *Music Sci* 10:185–208.
24. Ross D, Choi J, Purves D (2007) Musical intervals in speech. *Proc Natl Acad Sci USA* 104: 9852–9857.
25. Gill KZ, Purves D (2009) A biological rationale for musical scales. *PLoS One* 4:e8144.
26. Bowling DL, Gill K, Choi JD, Prinz J, Purves D (2010) Major and minor music compared to excited and subdued speech. *J Acoust Soc Am* 127:491–503.
27. Han Se, Sundararajan J, Bowling DL, Lake J, Purves D (2011) Co-variation of tonality in the music and speech of different cultures. *PLoS One* 6:e20160.
28. Bowling DL, Sundararajan J, Han S, Purves D (2012) Expression of emotion in Eastern and Western music mirrors vocalization. *PLoS One* 7:e31942.
29. Bowling DL (2013) A vocal basis for the affective character of musical mode in melody. *Front Psychol* 4:464.
30. Bowling DL, Gingras B, Han S, Sundararajan J, Opitz ECL (2013) Tone of voice in emotional expression: Relevance for the affective character of musical mode. *J Interdiscip Music Stud* 7:29–44.
31. Purves D (2017) *Music as Biology* (Harvard Univ Press, Cambridge, MA).
32. Malmberg CF (1918) The perception of consonance and dissonance. *Psychol Monogr* 25:93–133.
33. Roberts LA (1986) Consonance judgements of musical chords by musicians and untrained listeners. *Acustica* 62:163–171.
34. Rasmussen M, Santurette S, Macdonald EN (2014) Consonance perception of complex-tone dyads and chords. *Proceedings of the Seventh Forum Acusticum*. Available at orbit.dtu.dk/files/100043232/Rasmussen2014.pdf. Accessed April 14, 2017.
35. Lahdelma I, Eerola T (2014) Single chords convey distinct emotional qualities to both naive and expert listeners. *Psychol Music* 44:37–54.
36. Fujisawa TX, Cook ND (2011) The perception of harmonic triads: An fMRI study. *Brain Imaging Behav* 5:109–125.
37. Hutchinson W, Knopoff L (1979) The significance of the acoustic component of consonance in Western triads. *J Musicol Res* 3:5–22.
38. Cook ND (2001) Explaining harmony. The roles of interval dissonance and chordal tension. *Ann N Y Acad Sci* 930:382–385.
39. Cook ND, Fujisawa TX (2006) The psychophysics of harmony perception: Harmony is a three-tone phenomenon. *Empir Musicol Rev* 1:106–126.
40. Parncutt R (1989) *Harmony: A Psychoacoustical Approach* (Springer, Berlin).
41. Apple (2009) Logic Pro 9. (Apple Inc., Cupertino, CA).
42. McDermott JH, Schultz AF, Undurraga EA, Godoy RA (2016) Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature* 535:547–550.
43. Maxwell SE, Delaney HD (2004) *Designing Experiments and Analyzing Data: A Model Comparison Perspective* (Lawrence Erlbaum Associates, Mahwah, NJ), 2nd Ed.
44. Shrout PE, Fleiss JL (1979) Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull* 86:420–428.
45. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15:155–163.
46. Hollien H, Michel JF (1968) Vocal fry as a phonational register. *J Speech Hear Res* 11: 600–604.
47. Henrich N (2006) Mirroring the voice from Garcia to the present day: Some insights into singing voice registers. *Logoped Phoniatr Vocol* 31:3–14.
48. Risberg A (1961) Statistical studies of fundamental frequency range and rate of change. *STL-QPSR* 2:007–008.
49. Rameau J-P (1722) *Traité de l'harmonie* (Jean-Baptiste-Christophe Ballard, Paris).
50. Terhardt E (1974) Pitch, consonance, and harmony. *J Acoust Soc Am* 55:1061–1069.
51. Cousineau M, McDermott JH, Peretz I (2012) The basis of musical consonance as revealed by congenital amusia. *Proc Natl Acad Sci USA* 109:19858–19863.
52. Plack CJ (2010) Musical consonance: The importance of harmonicity. *Curr Biol* 20:R476–R478.
53. Bruckert L, et al. (2010) Vocal attractiveness increases by averaging. *Curr Biol* 20:116–120.
54. Parncutt R (2006) Commentary on Cook & Fujisawa's "The psychophysics of harmony perception: Harmony is a three-tone phenomenon." *Empir Musicol Rev* 1:204–209.
55. Plomp R, Levelt WJ (1965) Tonal consonance and critical bandwidth. *J Acoust Soc Am* 38:548–560.
56. Kameoka A, Kuriyagawa M (1969) Consonance theory part I: Consonance of dyads. *J Acoust Soc Am* 45:1451–1459.
57. Setharas WA (2005) *Tuning, Timbre, Spectrum, Scale* (Springer, London), 2nd Ed.
58. Mashinter K (2006) Calculating sensory dissonance: Some discrepancies arising from the models of Kameoka & Kuriyagawa, and Hutchinson & Knopoff. *Empir Musicol Rev* 1:65–84.
59. Kameoka A, Kuriyagawa M (1969) Consonance theory part II: Consonance of complex tones and its calculation method. *J Acoust Soc Am* 45:1460–1469.
60. Vassilakis PN (2005) Auditory roughness as means of musical expression. *Sel Reports Ethnomusicol* 12:119–144.
61. Nordmark J, Fahlén LE (1988) Beat theories of musical consonance. *Q Prog Status Rep* 29:111–122.
62. Greenwood DD (1991) Critical bandwidth and consonance in relation to cochlear frequency-position coordinates. *Hear Res* 54:164–208.
63. Vassilakis PN (2001) Perceptual and physical properties of amplitude fluctuation and their musical significance. PhD dissertation (University of California, Los Angeles).
64. Terhardt E (1974) On the perception of periodic sound fluctuations (roughness). *Acustica* 30:201–213.
65. Collias NE (1987) The vocal repertoire of the red junglefowl: A spectrographic classification and the code of communication. *Condor* 89:510–524.
66. Marler P (1969) Vocalizations of wild chimpanzees, an introduction. *Proceedings of the Second International Congress of Primatology, Atlanta, 1968* (Karger, Basel), pp 94–100.
67. Cleveland J, Snowdon CT (1982) The complex vocal repertoire of the adult cotton-top tamarin (*Saguinus oedipus oedipus*). *Z Tierpsychol* 58:231–270.
68. Lemasson A, Richard J-P, Hausberger M (2004) A new methodological approach to context analysis of call production. *Bioacoustics* 14:111–125.
69. Chiandetti C, Vallortigara G (2011) Chicks like consonant music. *Psychol Sci* 22: 1270–1273.
70. Sugimoto T, et al. (2010) Preference for consonant music over dissonant music by an infant chimpanzee. *Primates* 51:7–12.
71. McDermott J, Hauser M (2004) Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. *Cognition* 94:B11–B21.
72. Koda H, et al. (2013) Validation of an auditory sensory reinforcement paradigm: Campbell's monkeys (*Cercopithecus campbelli*) do not prefer consonant over dissonant sounds. *J Comp Psychol* 127:265–271.
73. Bidelman GM, Krishnan A (2009) Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *J Neurosci* 29:13165–13171.
74. Cariani PA, Delgutte B (1996) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J Neurophysiol* 76:1698–1716.
75. Bidelman GM, Krishnan A (2011) Brainstem correlates of behavioral and compositional preferences of musical harmony. *Neuroreport* 22:212–216.
76. Cariani P (2004) A temporal model for pitch multiplicity and tonal consonance. *Proceedings of the Eighth International Conference on Music Perception and Cognition (ICMPC)*, ed Lipscomb S (Causal Productions, Adelaide, Australia), pp 310–314.
77. Fisher SE (2006) Tangled webs: Tracing the connections between genes and cognition. *Cognition* 101:270–297.