

PHYSICS

The structure of musical harmony as an ordered phase of sound: A statistical mechanics approach to music theory

Jesse Berezovsky

Music, while allowing nearly unlimited creative expression, almost always conforms to a set of rigid rules at a fundamental level. The description and study of these rules, and the ordered structures that arise from them, is the basis of the field of music theory. Here, I present a theoretical formalism that aims to explain why basic ordered patterns emerge in music, using the same statistical mechanics framework that describes emergent order across phase transitions in physical systems. I first apply the mean field approximation to demonstrate that phase transitions occur in this model from disordered sound to discrete sets of pitches, including the 12-fold octave division used in Western music. Beyond the mean field model, I use numerical simulation to uncover emergent structures of musical harmony. These results provide a new lens through which to view the fundamental structures of music and to discover new musical ideas to explore.

INTRODUCTION

The ubiquity of music throughout history and across cultures raises a fundamental question: Why is this way of arranging sounds such a powerful medium for human artistic expression? Although there are myriad musical systems and styles, certain characteristics are nearly universal, including emergent symmetries such as a restriction to a discrete set of sound frequencies (pitches). Historically, the theory of music has followed an empirical top-down approach: Patterns are observed in music and generalized into theories. Recent work has aimed to generalize these generalized theories to uncover new potential patterns that can lead to new theories of music (1–3). Here, instead, we observe patterns that emerge naturally from a bottom-up theory. We start from two basic (and conflicting) principles: A system of music is most effective when it (i) minimizes dissonant sounds and (ii) allows sufficient complexity to allow the desired artistic expression. Mathematical statement of these principles allows a direct mapping onto a standard statistical mechanics framework. We can thereby apply the tools of statistical mechanics to explore the phenomena that emerge from this model of music. Just as in physical systems where ordered phases with lower symmetry (e.g., crystals) emerge across transitions from higher-symmetry disordered phases (e.g., liquids), we observe ordered phases of music self-organizing from disordered sound. These ordered phases can replicate elements of traditional Western and non-Western systems of music, as well as suggesting new directions to be explored.

The basis for a bottom-up approach was provided by discoveries in the field of psychoacoustics originating with Helmholtz (4) and further developed in the 20th century, which established a quantitative understanding of how sound is perceived. This leads to the idea that the structure of music is related to a minimization of dissonance D , as explored by Plomp and Levelt (5), Sethares (6, 7), and others. Minimization of D cannot be the only criterion for an effective musical system, however, or we would all listen to “music” composed from just a single pitch. Instead, an effective system of music must have some degree of complexity to provide a sufficiently rich palette from which to compose. A recognition of this idea has led to work on quantifying complexity in music, including by computing the entropy S of music

in the context of information theory (8) or by considering musical systems to be self-organizing via an evolutionary process (9).

The model I present here combines both the minimization of D and the maximization of S . I draw an analogy to thermodynamic systems with energy U and entropy S , whose macrostate is determined by minimizing Helmholtz free energy $F = U - TS$. The fixed temperature T is a parameter that specifies the trade-off between decreasing U and increasing S . Here, I similarly introduce a parameter T that specifies the trade-off between decreasing D and increasing S . A musical system in equilibrium will then be found by minimizing $F = D - TS$, allowing us to exploit the powerful array of tools developed for studying physical systems in statistical mechanics.

The remainder of this paper is organized as follows: I next describe the general model presented here, including how dissonance is quantified. Then, we study the behavior of the model in the mean field approximation and observe phase transitions between disordered sound and ordered distributions of pitches that reproduce commonly used musical systems. Last, we turn to a more realistic model with fewer assumptions and use numerical simulation to explore the patterns that emerge on a lattice of interacting tones.

Music: A system of interacting tones

Music is composed of a set of sounds, commonly referred to as “tones,” arranged in time. Individual tones can be characterized by their fundamental frequency f (“pitch”) with amplitude A , and a set of “partials” with amplitudes $\alpha_n A$ and frequencies $\phi_n f$ that approximately specify the frequency spectrum of the particular sound. For example, $\phi_n = n = 1, 2, \dots$ and $\alpha_n = 1/n$ for a sound produced by a sawtooth wave, as given by its Fourier series. In general, ϕ_n and α_n determine the “color” or “timbre” of the sound and depend on the instrument producing the sound.

Here, we consider music as a system of N_t tones $\{t_i\}$. Each tone t_i has pitch f_i and for present purposes, we take all tones to have the same timbre (the same α_n and ϕ_n). Two tones t_i and t_j interact with each other with strength D_{ij} via a listener’s perception of dissonance between tones. The total dissonance $D_{\text{tot}} = \sum_{\{ij\}} D_{ij}$, where the sum is over all pairs $\{ij\}$.

Dissonance is quantified as $D_{ij} = a_{ij} D(f_i, f_j)$, where $D(f_i, f_j)$ is a function specifying the dissonance between two tones with pitches f_i and f_j and a given timbre that are perceived together to a degree specified by the parameter a_{ij} (e.g., tones overlapping in time would have

Copyright © 2019
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

Downloaded from <http://advances.sciencemag.org/> on September 17, 2019

larger a_{ij}). We can compute $D(f_a, f_b)$ for a given timbre by first positing a function $d(f_a, f_b)$ that quantifies the dissonance (also known as roughness) for a pair of pure tones (that is, tones with pitches f_a and f_b and a single partial with $\phi = \alpha = 1$) and then summing over all pairs of partials. It is found experimentally that $d(f_a, f_b)$ increases with pitch difference $\Delta x = \log_2 f_a/f_b$ up to a critical value $\Delta x = w_c$ and then falls off to zero for larger Δx (see Fig. 1A) (5). (With this logarithmic scale for pitch, $\Delta x = 1$ corresponds to a change of one octave.) The measured values of w_c are found to depend on the lesser of f_a and f_b , varying from $w_c \sim 0.5$ for the lower audible range to $w_c \sim 0.02$ for the higher audible range. The preference for consonant pairs of pure tones is seemingly universal, appearing even in cultures with little exposure to Western music (10). We then make the simple assumption that the dissonance D of two non-pure tones is found by summing up the pairwise contributions from all pairs of partials (see Methods) (7). Figure 1B shows $D(f_a, f_b)$ versus Δx with a fixed $w_c = 0.03$ and the sawtooth timbre. This quantification of dissonance represents one of several competing approaches, but this formalism could also be applied to other models of dissonance perception [e.g., (11, 12)].

As noted by Plomp and Levelt (5) and others, for harmonic timbres (i.e., with $\phi_n = n$), the prominent minima of $D(f_a, f_b)$ occur at or near frequency ratios prominently featured in Western music. For example, the dotted lines in Fig. 1B indicate intervals of an octave, fifth, fourth, major third, and minor third. These intervals represent pitches with frequency ratios $f_a/f_b = 2, 3/2, 4/3, 5/4$, and $6/5$, respectively. One can note that the values of Δx corresponding to these ratios are reasonably approximated by integers 12, 7, 5, 4, and 3 divided by 12 (green dots in Fig. 1B show integers $n/12$). Hence, Western music traditionally divides the octave into 12 distinct pitches. We will see that this pattern and others emerge from the model presented here.

RESULTS AND DISCUSSION

Mean field approximation

We first apply a mean field approximation and study the equilibrium distribution of pitches. In the limit $N_t \rightarrow \infty$ and all $a_{ij} = a_0$ constant, the pitches within the set are described only by their probability distribution $P(x)$, where the pitch $x = \log_2 f/f_{\text{ref}}$ is specified by the ratio of its frequency f with an arbitrary reference frequency f_{ref} . $P(x)$ represents the probability with which a pitch x is used in a system of music. We make two further simplifying assumptions: (i) We take a fixed value of w_c , so D is a function of x only, and (ii) we assume the periodicity $P(x) = P(x + 1)$. That is, the pitch distribution is the same in every octave. With these assumptions, we define $D_p(x) = \sum_{n=-\infty}^{\infty} D(x + n)$ and restrict our attention to a single octave $x \in [0, 1)$. Integrating over $P(x)$, we obtain the total dissonance

$$D_{\text{tot}} = \frac{1}{2} \int_0^1 \int_0^1 P(x) D_p(x - y) P(y) dy dx \quad (1)$$

and entropy

$$S = - \int_0^1 P(x) \ln P(x) dx \quad (2)$$

We now wish to find the distribution $P(x)$ that minimizes the free energy $F = D_{\text{tot}} - TS$, subject to the constraint that $P(x)$ is normalized. Using a variational method, we obtain an integral equation for the extremum condition and find stable minima using an iterative approach (see Methods) (13).

Figure 2A shows the solutions $P(x)$ for several values of T using the first 10 partials of the sawtooth timbre and $w_c = 0.03$ [see fig. S1A for a plot of $P(x)$ over a continuous range of T]. At high T , $P(x) = 1$ is the stable solution, and the system is completely disordered. At low T , $P(x)$ approaches a single delta-function spike at an arbitrary pitch. As T increases from zero, additional peaks emerge, eventually leading to a transition to a $P(x)$ with 12 equal peaks. We will see below that the high- and low- T behaviors are separated by phase transitions from random, disordered sound to the ordered, discrete tuning systems used in music.

Insight can be gained using the Fourier representations $P(x) = \frac{1}{2} \sum_{k=0}^{\infty} p_k e^{2\pi i k x} + p_k^* e^{-2\pi i k x}$ and $D_p(x) = \sum_{k=0}^{\infty} d_k \cos(2\pi k x)$. For example, the real coefficients d_k are shown in Fig. 1C for the $D_p(x)$ calculated from (Fig. 1B). Normalization requires $p_0 = 1$, with the other p_k representing order parameters that describe the degree of k -fold periodic order.

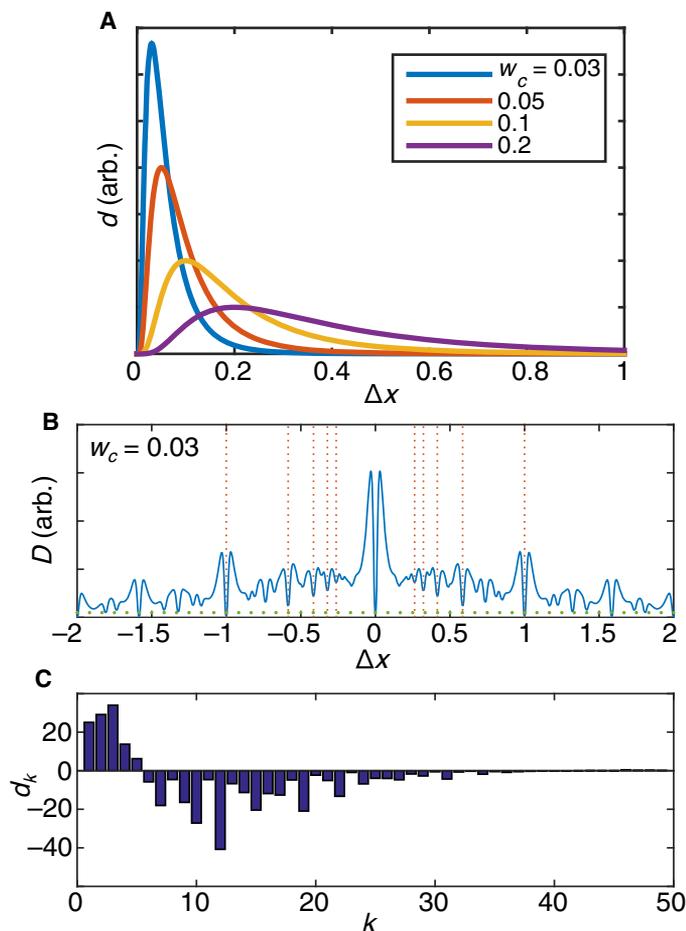


Fig. 1. Quantifying dissonance. (A) Pure-tone dissonance curves $d(\Delta x)$ at different critical widths w_c . (B) Dissonance function $D(\Delta x)$ for tones with a sawtooth timbre and a fixed $w_c = 0.03$. Just-tuned commonly used intervals are indicated by red dotted lines. Green dots show $x = n/12$, for integers n . (C) Fourier components d_k of the periodic dissonance function $D_p(\Delta x)$ corresponding to $D(\Delta x)$ shown in (B).

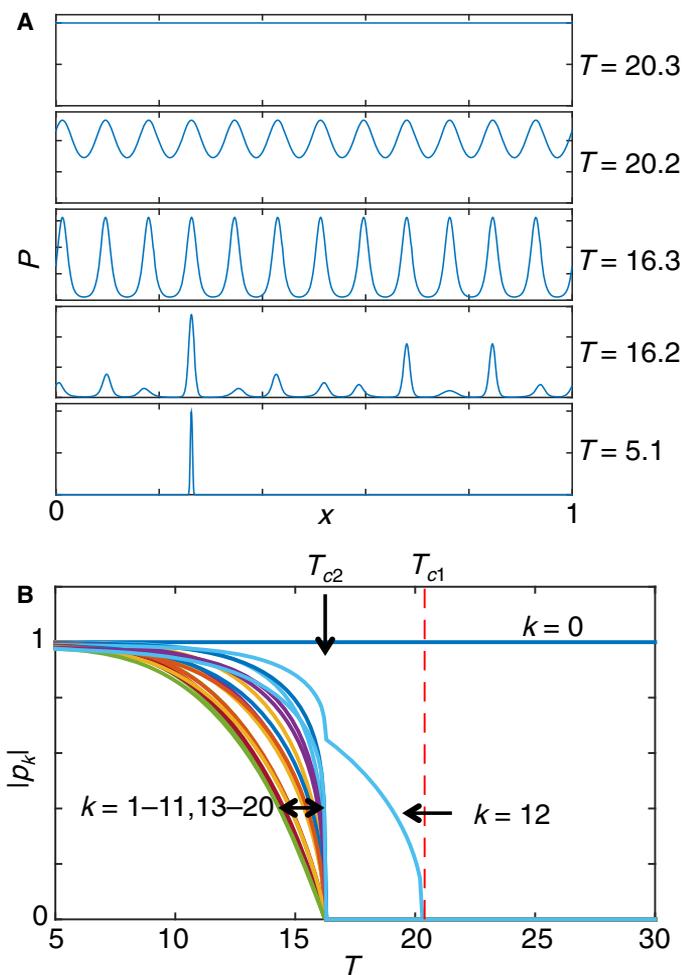


Fig. 2. Mean field results. (A) Solutions $P(x)$ to Eq. 4 for a range of T and $w_c = 0.03$. (B) Modulus $|p_k|$ of order parameters versus temperature T . The dashed line indicates the predicted $T_{c1} = -d_{12}/2$.

Figure 2B shows $|p_k(T)|$ for $k = 0$ to 20 for the data shown in (Fig. 1A). As expected, the high temperature solution has $p_0 = 1$ and all other $p_k = 0$. As T decreases, a single component p_{12} is the first to become nonzero at $T_{c1} = 20.2$. After a further decrease of T , the other p_k become nonzero at $T < T_{c2} = 16.2$.

The transition between the high- T disordered solution and the 12-fold solution can be understood from Landau theory. In the neighborhood of T_{c1} , we can expand F in powers of a single order parameter p_{k_0} (see Methods). We find that there is a second-order phase transition at $T_{c1} = -d_{k_0}/2$ from the disordered phase to a phase with k_0 -fold order, where d_{k_0} is the most negative Fourier coefficient. d_{12} is the most negative Fourier coefficient in Fig. 1C, with the transition to the 12-fold phase occurring at $T = -d_{12}/2$, as indicated by the dashed line in Fig. 2B.

The free energy only depends on the modulus of p_{k_0} , with the complex phase randomly chosen by spontaneous symmetry breaking. This symmetry breaking is seen in Fig. 2 as the continuous symmetry of the disordered phase changes to the 12-fold translational symmetry below T_{c1} . The randomly chosen phase of p_{12} gives rise to the particular positions of the 12 peaks in $P(x)$. In musical terms, a benchmark pitch is chosen arbitrarily and is fixed only by convention (e.g., A440

at 440 Hz). In this phase, all 12 peaks are equivalent and equally spaced, calling to mind the equal temperament (ET) tuning that has been commonly used in Western music for the last several hundred years (14).

When more than one $p_k \geq 1 \neq 0$, there are additional terms in the free energy arising from interactions between the modes. The coupling yields more complex behavior where multiple modes undergo transitions at the same temperature T_{c2} , reducing the symmetry to only the assumed octave-wise (“onefold”) symmetry. The spontaneous symmetry breaking randomly chooses one of the 12 peaks to become dominant over the others. This symmetry breaking calls to mind a tuning system that favors a particular key, as was common in medieval music through the time of J. S. Bach. For example, “just intonation” (JI) tuning places pitches at small-integer ratios relative to the root pitch. In Fig. 3, we compare the maxima of $P(x)$ from the mean field model with the pitches used in the JI and ET systems. Above T_{c2} , all pitches are treated equally by symmetry, so we must have the even 12-fold division of ET. When the symmetry is broken below T_{c2} , a single pitch can be designated as the root, and the pitches tend to align with the JI scheme. Just below T_{c2} , the peaks reflect a compromise between ET and JI, which is similar to tuning systems that have also been used historically, such as the sixth-comma meantone system (14).

The dependence of $P(x)$ on both T and the fixed value of w_c can be visualized via a color map of $d|p_k|/dT$, with different values of k represented by superimposing different colors (see Methods). The resulting plot is essentially a phase diagram with transitions made apparent by large values of $d|p_k|/dT$. Figure 4 shows the phase diagram for the sawtooth timbre used above, with T swept down at each value of w_c . [See fig. S1 for plots of $P(x)$ versus T at several values of w_c , as well as a comparison of the phase diagrams with T swept up and down. As expected for a second-order transition, T_{c1} is the same for T swept up and down, up to a small computational error. The lower transition shows significant hysteresis in T_{c2} , illustrating that mode coupling leads to a first-order transition]. The lines on the plot show $-d_k/2$ versus w_c with the value of k indicated by the same color scale. When T is swept down from the disordered phase, the first transition is well described by the single-mode, mean field prediction. The single value of k_0 can be seen clearly just below T_{c1} by a distinct color. As more modes become mixed in, the color fades, suddenly becoming bright white after the transition at T_{c2} . While the 12-fold phase that appears from $0.022 < w_c < 0.042$ reproduces the 12-fold division of the octave used in Western music, other ranges of w_c exhibit phases with other values of k_0 , such as 5, 7, 19, and 31. Suggestively, these values of k_0 are among those used in some non-Western music traditions (15, 16) or in modern music (17). The particular values of k_0 that appear, and their relative stabilities, depend on the choice of timbre (see fig. S1, C and D, for phase diagrams of timbres featured in non-Western music that do not use a 12-fold octave division). This mean field model therefore provides an avenue for optimizing the timbre of instruments to achieve a desired stability (or instability) for an existing or new musical system.

The mean field model assumes a fixed value of w_c , although experimentally w_c is found to depend on the root pitch (lower of the two pitches). The values of w_c that correspond to root pitches from middle C (C4) to the highest C on the piano (C8) are indicated on Fig. 4. Each ordered phase occupies a fairly small range of root pitches, with the 12-fold phase occurring only at relatively high pitch, likely because of the gross simplifications of the mean field approximation.

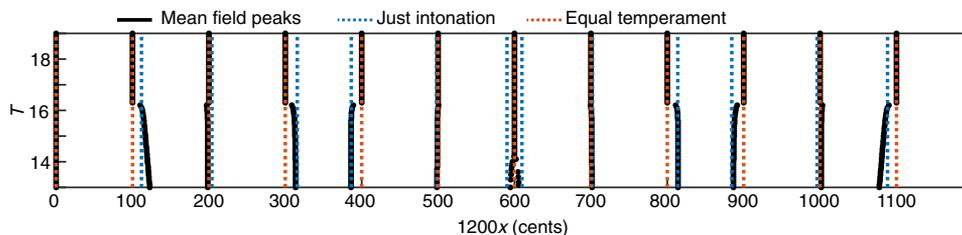


Fig. 3. Comparison of peak positions of $P(x)$ versus T from the mean field model with ET or JI tuning. Two values are shown for the tritone (the pitch near 600 cents) in JI, corresponding to an augmented fourth or diminished fifth. The mean field result shows equal peaks at both positions.

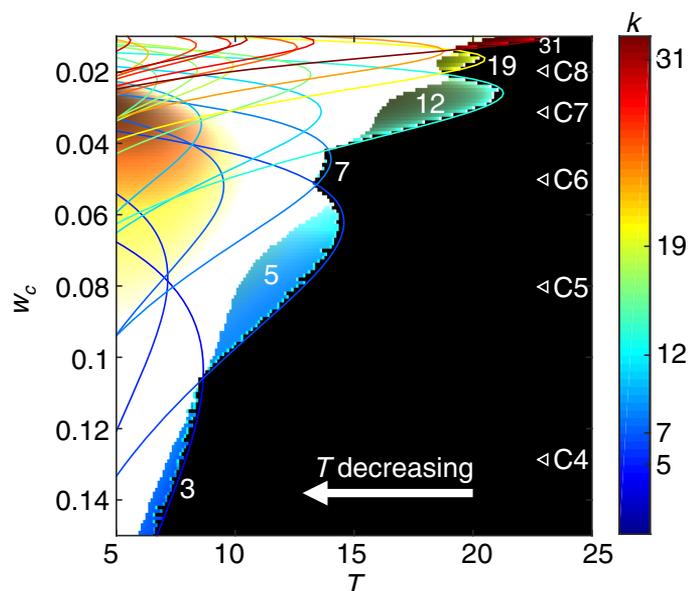


Fig. 4. Phase diagram showing significant order parameters $|p_k|$ versus T and w_c . $d|p_k|/dT$ is plotted with color indicated by the color bar, superimposing all values of k . Regions of a single color indicate phases with one or few dominant p_k ; white regions indicate phases with many significant p_k . The lines show $-d_k/2$ versus w_c for values of k corresponding to the color bar. w_c corresponding to pitches from C4 (middle C) to C8 are indicated.

So far, we have considered a model with a simplified dissonance function and assumed octave-wise periodicity, treated in the mean field approximation. One can explore other assumptions for the periodicity as well, simply by redefining $x = \log_b f/f_{\text{ref}}$ for b other than 2. For example, one finds that for tones with odd harmonic partials ($\phi_n = 1, 3, 5, \dots$), taking $b = 3$ (so “octaves” are separated by a factor of 3 in frequency) yields a 13-fold phase known as the Bohlen-Pierce scale (18) with lower free energy than with $b = 2$ (see fig. S2). One could simplify further in a Potts-type model (19), in which the tones t_i may take on only a discrete set of q pitches, $x_i = n/q$, $n = 0, 1, \dots, q - 1$. This model allows further analytic calculation and lends itself to other methods from statistical mechanics. These models serve to demonstrate that phase transitions can occur from disordered sound to ordered arrangements of pitches that bear notable resemblance to familiar musical systems.

The tone lattice

To go beyond the mean field model, we turn to numerical simulation of a system of interacting tones. Using this method, we will be able to relax the unrealistic assumption that each tone interacts equally with

all other tones. Instead, each tone will interact with a subset of the other tones. This can be accomplished by placing the tones on a lattice, with interaction only with a set of nearest neighbors. This lattice can be interpreted as existing in an abstract space, where tones on nearby lattice sites have a stronger harmonic relationship than more distant ones. Although many lattices with different dimension and different interactions may be studied, here, I choose one of the simplest cases that can be expected to produce rich behavior: tones on a two-dimensional (2D) square lattice, with only short-range interactions. Given octave-wise periodicity of interactions, this system can be described using the well-studied XY model (20, 21), where the pitch x stands in for the angle θ . The XY model on a 2D lattice is known to produce particularly rich behavior. As in superfluid thin films (22) and nematic liquid crystals (23), one expects to observe a Kosterlitz-Thouless transition from a disordered state, with free vortices to an ordered state exhibiting quasi-long-range order (24–26). A vortex in this system is a region surrounding a topological defect, where the pitches along a closed path enclosing the defect traverse one or more octaves. Following a quench of the system from the disordered to ordered state, metastable states are observed with bound vortices and antivortices (27–29). In this case, moreover, interactions tend not only to align neighboring pitches but also to favor sudden jumps by particular intervals $\Delta x \approx 3/12, 4/12, 5/12, 7/12, 9/12$. This type of XY model, where the energy has multiple minima, has found application, for example, in describing stacking domains in bilayer graphene (30), ferroelectric domains in hexagonal manganites (31), and axion-based cosmological models (32). For sufficiently small vortices, we expect not a continuous variation of x around a vortex but, instead, a number of domain boundaries where x changes by one of the more consonant intervals.

To study the 2D XY behavior that emerges from a set of tones on a lattice, we consider interactions only with a set of nearest neighbors and simulate the resulting stochastic dynamics (see Methods). As before, tones with pitches f_i and f_j interact because of their mutual dissonance $D(f_i, f_j)$. For maximum generality, we make no assumption about the periodicity of the pitch distribution and allow the dissonance width w_c to vary with the root pitch $\min(f_i, f_j)$ so that intervals involving lower pitches have broader dissonance curves, as observed experimentally. [Figure S3 shows $D(\Delta x)$ for several root pitches]. We then use a combination of Langevin dynamics and Metropolis Monte Carlo (MC) to simulate the behavior of the tone lattice at temperature T .

Figure 5A shows the histogram of pitch classes $x_i \bmod 1$ following a simulated quench from a disordered state to selected T (see fig. S4A for the full pitch histograms). At $T = 5.5$ and above, no long-range order of the pitches is observed. At $T = 5$ and below, the pitches show clear ordering. At $T = 5$ and $T = 4$, we see five- and sevenfold division of the octave, as was also seen in the mean field result. At $T = 3.5$, the

12-fold division of the octave has emerged, reproducing the Western musical system (for example, see fig. S4B showing the 12-fold division of the octave in both the frequency spectrum of the $T = 3.5$ tone lattice and the spectrum of Bach's *Prelude in D major, BWV850*). Now, we can delve deeper by studying the spatial arrangement of these pitches on the lattice. [See fig. S5 for plots of the distribution $C(\Delta x, r)$ of intervals $\Delta x = x_i - x_j$ versus distance r between sites i and j , at $T = 5.5$ and $T = 3.5$].

Figure 5B shows a visualization of the pitches on the lattice in a metastable configuration following a quench to $T = 3.5$. Each pixel corresponds to one lattice site colored according to a hue/saturation/value triplet. The hue (ranging from blue to red, as shown on the color bar) represents the pitch class $x_i \bmod 1$, with the saturation and value adjusted to represent $\lfloor x_i \rfloor$, the octave in which that pitch lies. Darker (lighter) colors represent lower (higher) octaves. The image is characterized by domains of the same hue, indicating a single pitch class within that domain (pitch classes are labeled by indices 0 to 11). Repeated runs of the simulation result in final states that differ in the details but show the same general behavior, such as the same division of the octave, and similar correlations. Figure S6 shows simulation results at $T = 4, 5$, and 6.

The quenched tone lattice can be viewed as a metastable configuration of bound vortices and antivortices, as expected for the 2D XY model. Because pitches can vary continuously only at a high cost of dissonance, it is more favorable to find domains of nearly constant pitch class separated by domain boundaries, where the pitch class changes by a consonant interval. The most common interval across a boundary is a fourth or fifth (change in pitch index of 5 or 7 mod 12). Major and minor thirds are also common (change in pitch index of 3 or 4 mod 12). Places where two or more domain boundaries meet at a point constitute a topological defect, with the surrounding domains forming a vortex around it. The simplest vortex consists of three domains that meet at a point (see points marked with a triangle in Fig. 5B). The pitches of these domains consist of a root pitch, a major or minor third above the root, and a fifth above the root. This combination of three pitches forms a major or minor triad—the type of chord that forms the backbone of Western music. Vortices with more domain boundaries are indicated by circles. (See the Supplementary Materials for sound clips of selected regions of the tone lattice.)

The arrangement of pitch domains on the lattice shown in Fig. 5B reflects elements of musical harmony. The Tonnetz (Fig. 5C), originally described by Euler (33, 34), is a graph in which the nodes are pitch indices and the edges (shown in blue) represent fifths, major thirds, and minor thirds connecting those pitch classes. In the ET approximation, the Tonnetz can be viewed as a graph on a torus, with the connections around the torus indicated by the gray circles. Moving continuously to the right, we increase by fifths, cycling around the “circle of fifths.” Moving along one diagonal direction changes by minor thirds, and the other diagonal direction changes by major thirds. Each set of three adjacent nodes forming a triangle represents a major or minor triad. The pitch classes on the tone lattice in Fig. 5B that share a border can be represented by a similar graph. We construct this graph first by identifying all pitch domains, defined as contiguous regions with at least five lattice sites having the same pitch, rounded to the nearest of the 12 pitch classes. Pitch domains that share a border are then identified as those that have at least two adjacent lattice sites in one domain bordering two adjacent lattice sites in the other domain. The resulting graph of all neighboring domains on the tone lattice (red lines in Fig. 5C) maps directly onto the Tonnetz. All possible fifths are present (for example, the

ellipse in Fig. 5B traverses the circle of fifths), with many of the thirds present as well.

The major or natural minor diatonic scales used in traditional Western music consist of seven contiguous pitch classes around the circle of fifths (or a trapezoid on the Tonnetz), whereas the anhemitonic pentatonic scales used in other musical traditions consist of five contiguous pitches on the circle of fifths. Therefore, all such scales may be constructed from contiguous regions of the tone lattice, with major

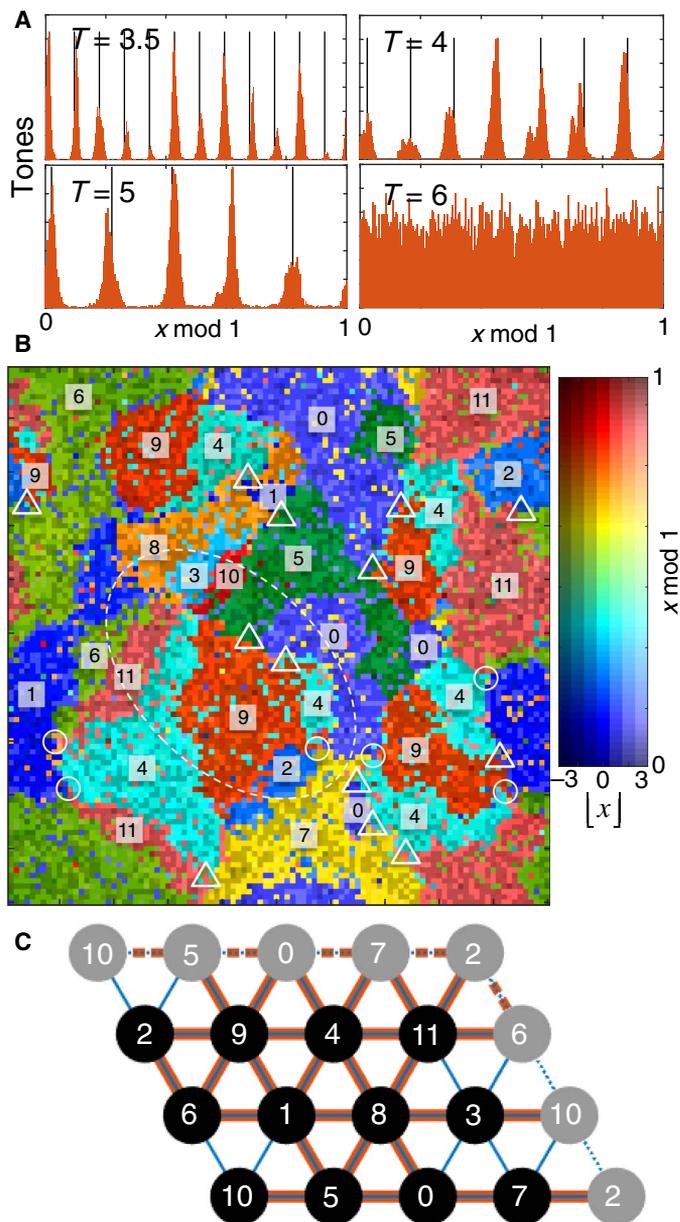


Fig. 5. Tone lattice simulation results. (A) Histograms of pitches $x \bmod 1$ in a metastable configuration of the tone lattice following a quench to temperature T . At $T = 6$, no ordering is observed. At $T = 5, 4$, and 3.5 , ordering is observed, with the octave divided into 5, 7, and 12, respectively. (B) Pitches on the tone lattice at $T = 3.5$. Pitch domains are labeled with pitch indices 0 to 11. Major or minor triads are marked with triangles. Junctions of more than three pitch domains are marked with circles. (C) The Tonnetz (edges shown in blue) with connections between neighboring pitch domains at $T = 3.5$ shown in red.

and minor triads within these scales appearing in these regions. Furthermore, as the tones on a finite lattice relax through the quenched metastable states toward equilibrium, some of the 12 pitch classes begin to grow and others shrink, possibly eventually disappearing. Because the domains are most often separated by fifths or fourths, it is likely that the domains that grow and do not disappear will be contiguous on the circle of fifths. Therefore, the pitches on the tone lattice are expected to converge toward a major or minor diatonic scale (and, upon further relaxation, to an anhemitonic pentatonic scale). Figure 6A shows a histogram of pitch classes from the $T = 3.5$ tone lattice, arranged in order of the circle of fifths. One can see that the first seven pitch classes are all fairly common (in the shaded box), and the three least common pitch classes are contiguous on the circle of fifths and outside the box. This type of pitch distribution is common in Western music, where many of the pitches are drawn from a major or minor diatonic scale (and thus from seven contiguous pitch classes on the circle of fifths), but with some exceptions. For example, the pitch distribution from Bach's *BWV850* is shown in Fig. 6B, in which all 12 pitch classes are present, but with the first 7 pitch classes (corresponding to the D major scale) occurring more frequently than the latter 5. In other runs of the simulation, or at different times within one run, one might observe peaks in the pitch distribution with more equal or less equal height (or some missing), which would agree better with other styles of music.

The arrangement of triads on the tone lattice gives rise to the same elements of music theory that have been extracted from the Tonnetz, such as a geometric interpretation of tonality (35), and progressions of

triads that produce parsimonious voice leading, the concept that allows the combination of melody and harmony known as counterpoint (36). By varying the quench temperature T , pitch distributions with different numbers of pitch classes are observed. Inspecting the arrangement of pitch domains on lattices at different T , one observes certain commonly occurring intervals across domain boundaries. One can then propose "Tonnetzes" in these musical systems by analogy to the usual Tonnetz with 12 pitch classes, where the most commonly occurring interval forms the horizontal edges and the less common intervals are represented by diagonal edges (see fig. S6).

The fact that the behavior of the tone lattice in two dimensions replicates many features of traditional Western music raises the question: What happens in other dimensions? No phase transition is expected in 1D (unless the interactions are extended to long range). In 3D, the expected phase transition of a system with XY symmetry has been studied extensively in the context of cosmology. In this case, the phase transition is not a Kosterlitz-Thouless transition but, instead, a normal first- or second-order transition. Still, though, one expects the emergence of topological defects in the ordered phase if the cooling takes place at a finite rate by the so-called Kibble-Zurek mechanism (37). In these topological defects, the point-like vortex cores seen in the 2D lattice are now extended along 1D paths, originally referred to as cosmic strings, and later observed in condensed matter systems (31, 38). It is possible that the extension of the tone lattice to three dimensions may give rise to a 3D Tonnetz-like structure, as has been previously proposed in the spiral array model (35), or in generalized Tonnetzes (3).

The results presented here demonstrate that many patterns in music can emerge from this statistical mechanics framework. Undoubtedly, further insights into the structure of music can be gained by studying related models and bringing all of the tools of statistical mechanics to bear. The historical development of thermodynamics provides a parallel to the present case. Starting in the 17th century, empirical laws (Boyle's law, Charles's law, etc.) were postulated to explain observations of the behavior of gases. This top-down approach to thermodynamics proved quite useful. However, the development of statistical mechanics yielded a bottom-up theory, and this fundamental understanding led to an array of new discoveries in the 20th century and beyond. Likewise, top-down music theory has proven quite useful for composing and understanding music, but a bottom-up theory provides a foundational understanding, perhaps leading to new ways of composing, understanding, and enjoying music.

METHODS

Constructing the dissonance function

The experimentally measured pure-tone dissonance curves can be parameterized by $d(f_a, f_b) = e^{-[\ln(|\Delta x|/w_c)]^2}/w_c$ (as shown in Fig. 1A), where $\Delta x = \log_2(f_a/f_b)$. For the mean field model, I set w_c at a fixed value. For the more realistic dissonance function in the tone lattice simulations, I parameterized $w_c = 0.67 \min(f_a, f_b)^{-0.68}$, with coefficients set by a fit to observed values from (5).

The dissonance of two non-pure tones is then found by summing up the contributions from all pairs of partials

$$D(f_a, f_b) = \sum_{n,m} l_{nm} d(\phi_n f_a, \phi_m f_b) \tag{3}$$

where $l_{nm} = \min(\alpha_n, \alpha_m)^{0.606}$ is the perceived "loudness" of the pair of partials, as approximated by Sethares (7). The general results presented

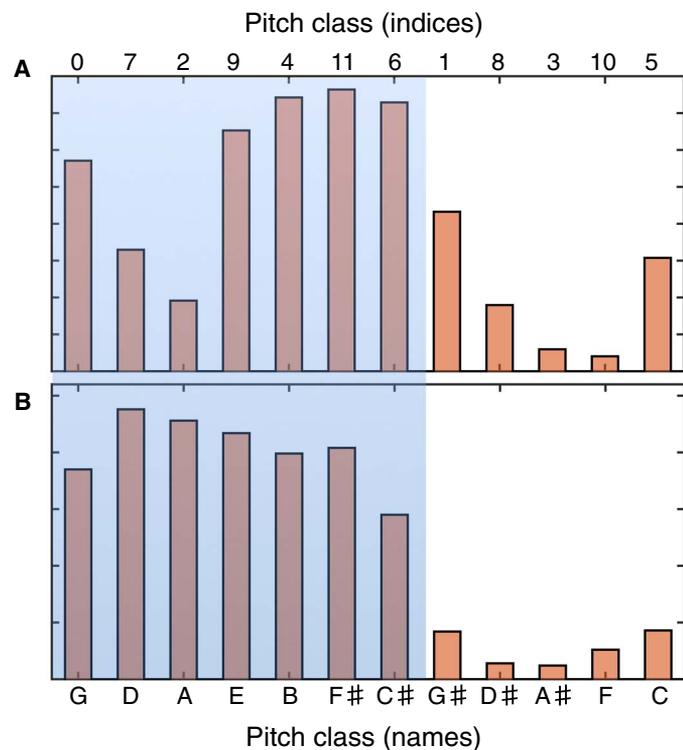


Fig. 6. Comparison of the pitch class distribution on the tone lattice and in a selected piece of music. (A) Histogram of pitch classes $12x \text{ mod } 12$ rounded to the nearest pitch class on the $T = 3.5$ tone lattice and arranged by ascending fifths. **(B)** Histogram of pitch classes appearing in Bach's *Prelude and Fugue in D major, BWV850*, arranged by ascending fifths. The shaded area corresponds to the notes of a diatonic scale.

here are robust to changes in the choice of parameters used to construct the dissonance function.

Mean field model

To find normalized $P(x)$ that are extrema of F , I extremized $F = U - TS - \mu\mathcal{N}$, where $\mathcal{N} = \int_0^1 P(x)dx - 1$ and μ is a Lagrange multiplier. Applying the variational principle, I found the extremum condition

$$P(x) = \frac{\exp\left[-\frac{1}{T}\int_0^1 D_p(x-y)P(y)dy\right]}{\int_0^1 \exp\left[-\frac{1}{T}\int_0^1 D_p(z-y)P(y)dy\right] dz} \tag{4}$$

Stable numerical solutions to this integral equation, which is similar to a Fredholm equation, can be found using an iterative approach (13). With the action of the right hand side of Eq. 4 on $P(x)$ represented by the operator \hat{Q} , the sequence $\hat{Q}^n P_0(x)$ either converges to or oscillates about a solution of the extremum condition $P(x) = \hat{Q}P(x)$, where $P_0(x)$ is an initial trial function. The convergence will be stable against perturbations for minima of F and unstable for maxima. To reduce the chance of oscillations, I instead applied the operator $\tilde{Q} = \gamma\mathbb{I} + (1 - \gamma)\hat{Q}$ with, say, $\gamma = 2/3$ and \mathbb{I} the identity operator. In practice, I discretized $P(x)$ ($0 \leq x < 1$) as an array (typically with 2400 elements) and repeatedly applied \tilde{Q} to an initial P_0 using numerical integration until the mean square difference between successive terms falls below a threshold. P_0 could be an array of random numbers, for example. For sequences of solutions $P(x)$ versus T , I used a random P_0 for the first solution and then used that solution plus small random jitter as the P_0 to solve at the incremented value of T . This speeds up convergence and maintains the same symmetry at each value of T except when a new symmetry is broken. The addition of random jitter ensures that the solution moves away from unstable fixed points.

In terms of the Fourier representations $P(x) = \frac{1}{2}\sum_{k=0}^{\infty} p_k e^{2\pi i k x} + p_k^* e^{-2\pi i k x}$ and $D_p(x) = \sum_{k=0}^{\infty} d_k \cos(2\pi kx)$, we have dissonance

$$D_{\text{tot}} = \frac{1}{2}d_0 + \frac{1}{8}\sum_{k=1}^{\infty} d_k |p_k|^2 \tag{5}$$

Using the Fourier representation and a Taylor series expansion for the logarithm, we have entropy

$$S = \sum_{n=2}^{\infty} \frac{(-1)^{n-1}}{n(n-1)2^n} \int_0^1 dx \left(\sum_{k=1}^{\infty} p_k e^{2\pi i k x} + p_k^* e^{-2\pi i k x} \right)^n \tag{6}$$

The transition between the high- T disordered solution and the 12-fold solution can be understood from Landau theory. Sufficiently close to T_{c1} , all $p_k \geq 1$ are small, and only the leading terms of S are required. Up to second order in p_k , we have free energy

$$F \approx \frac{1}{2}d_0 + \frac{1}{4}\sum_{k=1}^{\infty} \left(\frac{d_k}{2} + T\right) |p_k|^2 \tag{7}$$

A phase transition can occur when a second-order term changes sign, assuming that the highest-order terms are positive. Therefore, as T is reduced from the disordered phase, the term with the most negative $d_k = d_{k_0}$ will be the first to change sign, and a single-mode k_0 will take on a nonzero p_{k_0} . Because there is only a single $p_k \geq 1 \neq 0$ just below T_{c1} , we can use a single-mode approximation to find

$$F \approx F_{k_0} = \frac{1}{2}d_0 + \frac{1}{4}\left(\frac{d_{k_0}}{2} + T\right) |p_{k_0}|^2 + \frac{T}{32} |p_{k_0}|^4 + \frac{T}{96} |p_{k_0}|^6 \dots, \tag{8}$$

and with all subsequent terms also positive. From Landau theory, we conclude that there is a second-order phase transition at $T_{c1} = -d_{k_0}/2$, from the disordered phase to a phase with k_0 -fold order.

When more than one $p_k \geq 1 \neq 0$, there are additional terms in the free energy arising from interactions between the modes. Mathematically, these terms arise from cross terms in Eq. 6 whose exponential factors cancel out, thereby not integrating to zero. In the absence of this mode coupling, one would just expect a series of transitions where p_k becomes nonzero at $T = -d_k/2$ for all negative values of d_k . Instead, the mode coupling yields more complex behavior where multiple modes undergo transitions at the same temperature. These transitions may be either first or second order and can display hysteresis.

Phase diagrams of the mean field solution versus T and w_c were constructed by finding $P(x)$ versus swept T as above and then repeating at various fixed values of w_c . Fourier amplitudes $|p_k|(T)$ were obtained from the series of solutions $P(x)$, and the derivative $d|p_k|/dT$ was calculated by a finite difference. The color of a pixel at each (T, w_c) pair was determined as follows. Thirty-two colors spanning a spectrum color map (MATLAB’s “jet” color map) are specified by red-green-blue triplets (r_k, g_k, b_k) ($k = 1 - 32$). The pixel’s color was then set to $c(T, w_c) = \sum_k - (d|p_k|/dT)(r_k, g_k, b_k)$, scaled to adjust the saturation of the image.

Tone lattice simulation

Here, I placed the tones on a 100×100 2D square lattice with periodic boundary conditions at the edges and with interactions only with the eight nearest and next-nearest neighbors. I set $a_{ij} = 1$ for these neighboring tones t_i and t_j , and $a_{ij} = 0$ otherwise. The dissonance function was calculated with w_c depending on the root pitch, as described in the “Constructing the dissonance function” section. Instead of the sawtooth timbre used in the mean field model, I used harmonic partials given by Helmholtz (4) for a plucked string: $\alpha_{1-6} = 1, 0.812, 0.561, 0.316, 0.13, 0.028$. With the lack of periodic boundary conditions for the pitch distribution, we must now impose additional boundary conditions to prevent the pitches from diverging to $x = \pm \infty$. This was accomplished via quadratic or quartic terms added to the dissonance function to artificially increase the dissonance of very low or high pitches or very large intervals. Specifically, I add $\delta D = 0.2(\Delta x)^4 + 0.3125x_r^2$, where the interval $\Delta x = \log_2(f_a/f_b)$ and the root pitch $x_r = \log_2(\min(f_a, f_b)/300 \text{ Hz})$. That is, an interval of one octave is barely suppressed at all with $\delta D = 0.2$, and intervals of two octaves are suppressed somewhat by $\delta D = 3.2$ (as compared to typical temperature in the simulation $T = 3.5 - 6$). Likewise, root frequencies two octaves above or below 300 Hz are suppressed by $\delta D = 1.25$. (See fig. S3 for plots of these dissonance functions.)

Via the analogy between dissonance and energy, the gradient of dissonance can be viewed as a force, pulling intervals toward lower

dissonance. Assigning the tones a mass m and damping constant ζ , we can write a stochastic equation of motion

$$m\ddot{x}_i + \zeta\dot{x}_i + dD_i/dx = \sqrt{2T\zeta}\eta_i$$

where D_i is the total dissonance of tone i with its neighbors and η_i is a Gaussian white noise scalar. Evolving this system of equations for all i yields the dynamics of the system at temperature T . I chose m and ζ such that the dynamics are typically overdamped so that the system rapidly thermalizes.

Simulating the Langevin dynamics of the system explores the phase space around one local minimum of the total dissonance. To fully sample the phase space, I also included MC steps, chosen via the Metropolis algorithm, resulting in larger jumps in phase space. Here, a lattice site i was chosen at random, and I calculated the change in total dissonance δD if the pitch x_i were changed to a new, randomly selected value. If $\exp(-\delta D/T) < r$, with $0 \leq r < 1$ a uniformly distributed random number, then the change was rejected; otherwise, the change was accepted.

The set of tones was initialized into a disordered state with pitches x_i drawn from a normal distribution with variance $\sigma^2 = 1$. The simulation proceeded by performing a number (in this case, 3000) of MC steps followed by thermalization via Langevin dynamics at a fixed temperature T . This sequence was then repeated 2500 times, essentially simulating quenching from the disordered state.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/5/5/eaav8490/DC1>

Fig. S1. Mean field solutions versus T at selected w_c and phase diagrams for different timbres and sweep directions.

Fig. S2. Comparison of F versus T for different timbres and octave periodicities.

Fig. S3. Dissonance functions $D(\Delta x)$ for different root pitches (and hence different w_c) in legend, calculated as described in Methods for the lattice simulations.

Fig. S4. Pitch histograms and spectra from the tone lattice simulation.

Fig. S5. Distribution of intervals (pitch differences $\Delta x = x_i - x_j$) $C(\Delta x, r)$ as a function of distance r between sites i and j .

Fig. S6. Metastable or disordered configurations of the tone lattice following a quench to $T = 4$, $T = 5$, and $T = 6$.

Movie S1. Major diatonic scale on the quenched tone lattice.

Movie S2. Major and minor triads on the quenched tone lattice.

Description of MATLAB scripts.

REFERENCES AND NOTES

1. F. Lerdahl, R. Jackendoff, *A Generative Theory of Tonal Music* (MIT Press, 1983).
2. D. Tymoczko, The geometry of musical chords. *Science* **313**, 72–74 (2006).
3. C. Callender, I. Quinn, D. Tymoczko, Generalized voice-leading spaces. *Science* **320**, 346–348 (2008).
4. H. Helmholtz, *On the Sensations of Tone*, A. J. Ellis, Transl. (orig. publication, 1877; Courier Corporation, 2013).
5. R. Plomp, W. J. M. Levelt, Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* **38**, 548–560 (1965).
6. W. A. Sethares, Local consonance and the relationship between timbre and scale. *J. Acoust. Soc. Am.* **94**, 1218–1228 (1993).
7. W. A. Sethares, *Tuning, Timbre, Spectrum, Scale* (Springer, 1998).
8. J. E. Youngblood, Style as information. *J. Music Theory* **2**, 24–35 (1958).
9. J.-J. Aucouturier, E. R. Miranda, The hypothesis of self-organization for musical tuning systems. *Leonardo Music J.* **18**, 63–69 (2008).
10. J. H. McDermott, A. F. Schultz, E. A. Undurraga, R. A. Godoy, Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature* **535**, 547–550 (2016).
11. N. D. Cook, T. X. Fujisawa, The psychophysics of harmony perception: Harmony is a three-tone phenomenon. *Empir. Musicol. Rev.* **1**, 106–126 (2006).

12. D. L. Bowling, D. Purves, A biological rationale for musical consonance. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11155–11160 (2015).
13. C. Wagner, On the solution of Fredholm integral equations of second kind by iteration. *J. Math. Phys.* **30**, 23–30 (1951).
14. J. M. Barbour, *Tuning and Temperament: A Historical Survey* (Courier Corporation, 2013).
15. W. Surjoningrat, A. Susanto, P. J. Sudarjana, *Tone Measurements of Outstanding Javanese Gamelans in Jogjakarta and Surakarta* (Gadjah Mada Univ. Press, 1972).
16. D. Morton, C. Duriyanga, *The Traditional Music of Thailand* (University of California Press, 1976).
17. J. Mandelbaum, *Multiple Division of the Octave and the Tonal Resources of 19-Tone Temperament* (Indiana Univ., 1961).
18. M. V. Mathews, J. R. Pierce, A. Reeves, L. A. Roberts, Theoretical and experimental explorations of the Bohlen-Pierce scale. *J. Acoust. Soc. Am.* **84**, 1214–1222 (1988).
19. F. Y. Wu, The Potts model. *Rev. Mod. Phys.* **54**, 235–268 (1982).
20. E. Lieb, T. Schultz, D. Mattis, Two soluble models of an antiferromagnetic chain. *Ann. Phys. Rehabil. Med.* **16**, 407–466 (1961).
21. E. Barouch, B. M. McCoy, M. Dresden, Statistical mechanics of the XY model. *I. Phys. Rev. A* **2**, 1075–1092 (1970).
22. D. Bishop, J. Reppy, Study of the superfluid transition in two-dimensional ^4He films. *Phys. Rev. Lett.* **40**, 1727–1730 (1978).
23. A. N. Pargellis, S. Green, B. Yurke, Planar XY-model dynamics in a nematic liquid crystal system. *Phys. Rev. E* **49**, 4250–4257 (1994).
24. V. Berezinsky, Destruction of long range order in one-dimensional and two-dimensional systems having a continuous symmetry group. I. Classical systems. *Sov. Phys. JETP* **32**, 493–500 (1971).
25. J. M. Kosterlitz, D. J. Thouless, Ordering, metastability and phase transitions in two-dimensional systems. *J. Phys. C Solid State Phys.* **6**, 1181–1203 (1973).
26. J. M. Kosterlitz, The critical properties of the two-dimensional xy model. *J. Phys. C Solid State Phys.* **7**, 1046–1060 (1974).
27. M. Mondello, N. Goldenfeld, Scaling and vortex dynamics after the quench of a system with a continuous symmetry. *Phys. Rev. A* **42**, 5865–5872 (1990).
28. F. Rojas, A. D. Rutenberg, Dynamical scaling: The two-dimensional XY model following a quench. *Phys. Rev. E* **60**, 212–221 (1999).
29. A. Jelić, L. F. Cugliandolo, Quench dynamics of the 2d XY model. *J. Stat. Mech.* **2011**, P02032 (2011).
30. J. S. Alden, A. W. Tsen, P. Y. Huang, R. Hovden, L. Brown, J. Park, D. A. Muller, P. L. McEuen, Strain solitons and topological defects in bilayer graphene. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11256–11260 (2013).
31. S. M. Griffin, M. Lilienblum, K. T. Delaney, Y. Kumagai, M. Fiebig, N. A. Spaldin, Scaling behavior and beyond equilibrium in the hexagonal manganites. *Phys. Rev. X* **2**, 041022 (2012).
32. P. Sikivie, Axions, domain walls, and the early universe. *Phys. Rev. Lett.* **48**, 1156–1159 (1982).
33. L. Euler, *Tentamen Novae Theoriae Musicae ex Certissimis Harmoniae Principiis Dilucide Expositae* (ex typographia Academiae scientiarum, 1739).
34. L. Euler, *De harmoniae veris principiis per speculum musicum repraesentatis. Opera Omnia* **1**, 568–586 (1774).
35. E. Chew, "Towards a mathematical model of tonality," thesis, Massachusetts Institute of Technology (2000).
36. R. Cohn, Neo-Riemannian operations, parsimonious trichords, and their "Tonnetz" representations. *J. Music Theory* **41**, 1–66 (1997).
37. A. del Campo, W. H. Zurek, Universality of phase transition dynamics: Topological defects from symmetry breaking. *Int. J. Mod. Phys. A* **29**, 1430018 (2014).
38. I. Chuang, R. Durrer, N. Turok, B. Yurke, Cosmology in the laboratory: Defect dynamics in liquid crystals. *Science* **251**, 1336–1342 (1991).

Acknowledgments: I acknowledge important contributions from H. Mathur and R. Petshek on methods and models in statistical mechanics. **Competing interests:** The author declares that he has no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the author.

Submitted 25 October 2018

Accepted 2 April 2019

Published 17 May 2019

10.1126/sciadv.aav8490

Citation: J. Berezovsky, The structure of musical harmony as an ordered phase of sound: A statistical mechanics approach to music theory. *Sci. Adv.* **5**, eaav8490 (2019).

The structure of musical harmony as an ordered phase of sound: A statistical mechanics approach to music theory

Jesse Berezovsky

Sci Adv 5 (5), eaav8490.
DOI: 10.1126/sciadv.aav8490

ARTICLE TOOLS	http://advances.sciencemag.org/content/5/5/eaav8490
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2019/05/13/5.5.eaav8490.DC1
REFERENCES	This article cites 29 articles, 5 of which you can access for free http://advances.sciencemag.org/content/5/5/eaav8490#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the [Terms of Service](#)

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Advances* is a registered trademark of AAAS.