Crystals of Sound: Applying the Physics of Phase Transitions to Musical Intonation

RYAN BUECHELE
Case Western Reserve University, Cleveland (OH)

ALEX COOKE
Cleveland Institute of Music, Cleveland (OH)

JESSE BEREZOVSKY [1]
Case Western Reserve University, Cleveland (OH)

ABSTRACT: We explore a physics-based model describing systems of musical intonation. Under the assumption that tuning systems seek to balance minimization of dissonance with maximization of compositional variety, we can use the same methods that are used to describe how transitions between the phases of matter are governed by a balance between minimizing energy and maximizing entropy. In both cases, a parameter – temperature – controls the balance between these factors. We show that as the temperature is raised from low to high, our model generates tuning systems that closely resemble those used throughout the history of Western music, from just intonation, to meantone temperaments, to equal temperament. We demonstrate how our model reflects the way in which tuning systems evolved along with trends in composition by comparing the results of our model vs. temperature to a corpus of 9620 pieces vs. year of composition from 1568-1968. By changing parameters in the model, we also find that other tuning systems, such as 31-tone temperaments, emerge. These results provide a new lens for understanding tuning systems, and a new approach for generating novel systems of intonation.

BEAUTY in music can arise out of a balance between consonance and dissonance, regular order and variety, predictability and surprise (Huron 2006, Meyer 2008). In nature, similar competitions cause beauty to spontaneously emerge out of randomness. Symmetric snowflakes assemble from randomly colliding water molecules. Intricate crystals grow from solutions of ions dissolved in water. To understand how these processes occur, we can apply the ideas and mathematical tools of statistical mechanics (see, for example, Pathria 2016, Chapters 11-13). In the realm of physics, the study of statistical mechanics yields a window into transitions between the phases of matter. As temperature decreases, and the randomly colliding atoms in a gas condense into a liquid, then crystallize into an ordered solid structure, different balances are struck between randomness and cohesive forces. Here, we will apply the same methods to see how the balance between consonance and compositional variety can open a new view onto the development of musical tuning systems.

The problem of assigning frequencies to musical notes, known as temperament, has a rich history. Lindley (2001) notes that despite this history, the 12-tone equal temperament, in which the octave is divided into 12 equal semitones, has become the standard tuning system in Western music today, except for specialists in early music. If one wishes to design a tuning system dividing the octave into 12 pitches, the fundamental problem is that if one tunes notes purely by fifths, the note one returns to once completing the chromatic is not an integer multiple of the original, creating an octave that is out of tune with itself. In other words, the space covered by 12 fifths is ever so slightly larger than that covered by 7 octaves. This difference is known as the Pythagorean comma. Similarly, the space covered by four justly tuned perfect fifths is slightly

larger than that covered by two octaves and a justly tuned major third, a difference known as the syntonic comma.

Various temperaments have arisen over time to correct these issues. Pythagorean temperament was used in the Middle Ages and makes most fifths sound very consonant, but at the cost of dissonant thirds and the inability to modulate to equally well-tuned keys. Pythagorean temperalent was followed by meantone temperament. Meantone temperament was an improvement on these issues, fixing major thirds with slight compromises to fourths and fifths (with one particularly ugly fifth known as the "wolf") while increasing the number of usable triads. Meantone also allows for modulation to a certain degree. Modified meantone later emerged to help reduce the wolf.

Equal temperament, in which every semitone is tuned to the same frequency ratio, eventually took hold. The advantage of equal temperament is the full availability of keys and modulations. The downsides are a loss of key "colors" and the fact that no interval is perfectly in tune (save for the octave), but with each key and interval equally out of tune, it has become the system of choice in a modern landscape, where composers work with frequent and distant modulations and non-diatonic pitch collections.

In this paper, we will illustrate the evolution of Western tuning systems using a mathematical model borrowed from the study of the phases of matter (Berezovsky 2019). Just as water spontaneously reorganizes its constituent atoms to form a gas, liquid, or solid as the temperature is changed, we can view the historical development of tuning systems as a reorganization of the constituents of music in response to changing tastes or stylistic imperatives. The results we present here mirror the construction of scales and tuning systems that Tenney (2008) referred to as "crystal growth" in harmonic space.

To motivate the model we develop here, we consider building a tuning system from scratch. For our purposes, we will require that the tuning system has octave periodicity, but otherwise impose no restrictions, including on the number of pitch classes. We could start by placing a single pitch class somewhere within the octave. In principle, we could stop there and compose music with just a single pitch class. Audiences, however, would soon tire of these compositions (and composers might tire of them as well). So, we might be motivated to add the next most consonant interval, a perfect fifth. A perfect fifth can be obtained by adding another pitch class with root frequency 3/2 above the original pitch class. Another perfect fifth can be introduced in the tuning system by introducing a pitch class a perfect fifth below the original pitch class (or with octave periodicity, a perfect fourth above). The fundamental frequency of this new pitch class is 2/3 of the original (or 4/3, transposed up an octave.) Now, we have three pitch classes within the octave. But now, if there is nothing special about the original pitch class, we should also add perfect fifths above and below the two new pitch classes, then above and below these new pitch classes as well, and so on. We may also wish to apply the same reasoning to other relatively consonant intervals, such as major or minor thirds (or by inversion, minor or major sixths, respectively.) Note that here and throughout, we refer to "sensory" consonance and dissonance arising from the psychoacoustic perception of two simultaneous tones as defined by Terhardt (1984). Other notions of musical consonance and dissonance as cataloged by Tenney (1988) have progressed from linear melodic ideas to vertical harmonic guidelines, to functional harmonic syntax. The modern explanation based on psychoacoustic principles focuses on measurable acoustic properties like beating and roughness and allows the quantitative treatment presented here, but the role of musical learning and enculturation also remains relevant. Huron (1994) tabulates several empirical measurements of perceived sensory consonance.

The proliferation of pitch classes results in a much-expanded palette of pitches from which to compose. It also introduces less consonant intervals. The first addition of a perfect fourth and fifth above the original pitch class introduces a whole tone (major second) between the two new pitch classes with significantly less sensory consonance than the perfect fourth or fifth. A major third and a perfect fourth yield a semitone (minor second) between them. And even more so, the difference between the original pitch class and twelve consecutive justly tuned perfect fifths above it yields the highly dissonant Pythagorean comma (about 24 cents). Likewise, the difference between four consecutive perfect fifths and a justly tuned major third yields the dissonant syntonic comma (about 22 cents). We might then be inclined to adjust the tuning of the pitch classes to eliminate these dissonant commas.

The above considerations are familiar in the study of tuning systems. The choice of which pitch classes to add to a tuning system involves making trade-offs. How much does one favor consonant intervals over dissonant intervals? How much flexibility and complexity does one wish to allow for within this system? Is it safe to assume that certain pairs of pitches will not be sounded together, and thus should be given less weight in considerations of consonance and dissonance? Should different keys have different "colors"?

Balancing these factors in different ways has led to a plethora of tuning systems, but is there a way to balance these factors systematically?

Here, we draw an analogy between the factors that govern how we arrange pitches into different tuning systems and how atoms arrange themselves to form different states of matter (e.g. solid, liquid, and gas). On the strength of this analogy, we can borrow the mathematical techniques (statistical mechanics) used to describe physical phase transitions to derive musical tuning systems. The results we obtain will yield information about the division of an octave into discrete pitches, the intonation of those pitches, and the relative frequency with which those pitches are used.

Like tuning systems, phase transitions between states of matter are governed by trade-offs. In this case, the phase in which a system exists at a particular temperature depends on a trade-off between a system's tendency to lower its energy and a tendency to increase its entropy. An atom typically has lower energy when it is near another atom. The atoms may have still lower energy when they are near an atom at certain angles with respect to other nearby atoms. Consequently, the tendency towards low energy means that atoms tend to attract each other, and stick together in particular geometric arrangements. The tendency towards increased entropy, however, works against this tendency towards order. Loosely speaking, entropy is a quantitative measure of disorder in a system. More precisely, entropy is a measure of the number of ways the individual constituents of a system can be rearranged to yield the same overall state of the system as a whole. (Specifically, entropy is the natural logarithm of this number). For example, our system may consist of a very large number of atoms, and its overall state may be either a solid or a gas. In a solid, the atoms are arranged in a crystal lattice with precise rules that govern its structure. The number of ways the atoms can be arranged to form the same solid crystal structure is relatively low, and the state has low entropy. Conversely, in a gas, the atoms are arranged at random positions. There are many different random positions an atom can occupy, and therefore there is an extremely large number of different configurations of all the atoms that result in the overall state of a gas. Thus a gas is a higher entropy state than a solid.

The tendency towards low energy leads towards atoms ordering into the solid phase, whereas the tendency towards high entropy leads towards atoms existing in the disordered gas phase. The winner between these two opposing tendencies is chosen by the temperature. In fact, this is how temperature is defined in physics: as the trade-off between an increase of energy for an increase of entropy. At high temperatures, a system will allow its energy to increase [2] (against its tendency towards low energy) so that it can achieve a higher entropy state. And at low temperatures, the tendency towards low energy will win, despite leading to a low entropy state. At certain temperatures, the balance suddenly shifts to favor one state over another. At these temperatures, a phase transition occurs from one state to another, as at 32 deg. F (0 deg. C) for water going between the liquid and solid phases.

In analogy with this description of physical phase transitions, we have argued above that musical tuning systems are governed by the trade-off between two factors: 1) the desire for inclusion of more consonant intervals and fewer dissonant intervals and 2) the desire for greater complexity and variation in the music that can be composed within that system.

Table 1. Analogy between the forces driving physical phases of matter and musical tuning systems.

	Physics	**Music**
Minimize:	Energy	Dissonance
Maximize Entropy:	# of ways of arranging atoms	# of ways of arranging pitches

Specifically, the desire for less dissonance in a tuning system is analogous to a physical system's tendency to low energy, and the desire for more complexity (more possible arrangements of notes within the system) is analogous to the tendency towards higher entropy (see Table 1). To extend the analogy further, we can posit a parameter that sets the optimal trade-off between low dissonance and high complexity in a tuning system: the "temperature" of that system.

If the general tendencies of a system to decrease some quantity while maximizing entropy can be defined mathematically, then the tools of statistical mechanics can be used to predict the optimal state of the system as a whole. The power of statistical mechanics comes from its ability to predict the overall properties

of a system composed of a large number of constituents, without having to deal with the details of each constituent. Instead, the constituents (say, atoms in physics, or tones in music) are treated statistically.

To describe a tuning system statistically, we can describe the probability that a given pitch will be used in that system. Technically, since pitch can take on a continuum of values, we can describe a "probability distribution." [3] For example, the 12-TET system could be described by the probability distribution shown in Figure 1. Each peak in the probability distribution represents one of the twelve equally-spaced pitch classes of 12-TET. Because the spacing between all pitch classes is equal, there is nothing to distinguish one pitch class over another, and as such, all the peaks have the same height, meaning that each pitch class is equally probable. For example, in tonal music, the tuning system contains no preference for one key over another, so we can expect all keys to be used equally, resulting in equal usage of each pitch class overall.
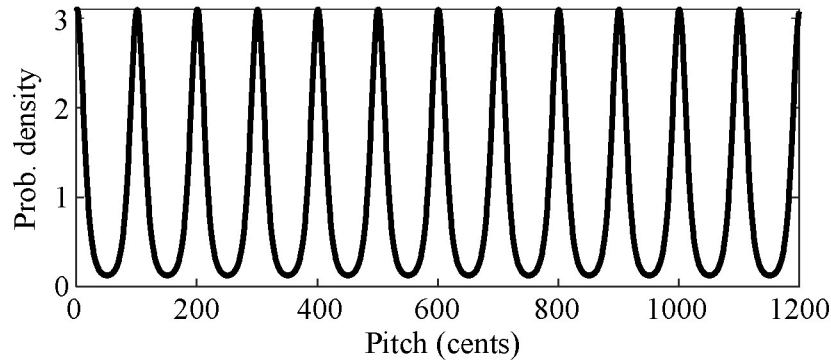

Figure 1. An example of a probability distribution representing a tuning system, in this case, 12 tone equal temperament. Note that one octave on the pitch axis is shown as a range of 1200 cents, but the distribution is normalized to an octave range of 1.

Individual pieces favor certain keys and thus, certain pitches, but across all repertoire, we expect equal usage. (Of course, design of particular instruments or historical tradition might lead to a preference for particular keys, but that preference is not inherent in the tuning system itself.) Note that while a pitch class in a tuning system is usually specified by one exact pitch, here a pitch class corresponds to a small range of pitches. This implies that there is some flexibility in the intonation of each pitch. Whether or not it is explicitly discussed as part of a tuning system, such small adjustments of intonation are common and add richness to performed music, assuming the instrument (e.g. the piano) doesn't make it impractical.

With the probability distribution of pitches in hand, there is still a problem: even though we know how often each pitch class is used with a particular tuning system, we don't know whether those pitches will be sounded together. For example, B, C and G may be all used with equal likelihood in the corpus of music composed and performed within 12-TET, but the perfect fifth of C and G is typically much more likely to be sounded together than the semitone of B and C. The same problem arises in physics: we can specify a probability distribution for the position of an atom in the system, but to know the total energy, we also need to know what the positions of the atoms are relative to any particular atom. There are many sophisticated mathematical and computational techniques to overcome this problem, but the most basic, and surprisingly successful, method is to just ignore this problem. Instead, we make the assumption that the single-constituent probability distribution contains all the information we need. In this case, we assume that a given pitch will be sounded together with another pitch chosen at random from the probability distribution. That is, two tones that are sounded together are both drawn independently from the same probability distribution. This assumption is called the "mean field assumption" and such a model is a mean field model (see, for example, Pathria 2016, Chapter 11).

To mathematically analyze a mean field model for a musical tuning system, we must establish a method for quantifying both the total average dissonance and the total entropy, given a particular tuning system specified by the pitch probability distribution. First, we will discuss quantifying dissonance, then quantifying entropy. The following description attempts to avoid overwhelming the reader with mathematics; the details are given in Appendix A.

QUANTIFYING DISSONANCE

Much has been written about perceived dissonance in music. Experimentally, the perception of dissonance has been found to subtly depend on many factors, such as culture (McDermott et al. 2016, Milne et al. 2023), musical training (Proverbio et al. 2016), and expectation (Meyer 2008). There are, however, some underlying commonalities. Closely-spaced (but not identical) pitches with perceptible beating (said to be within a "critical band") are almost always perceived to be rough or unpleasant (Berg and Stork 1990). Intervals with fundamental frequencies in small integer ratios (e.g. 2:1 or 3:2) are typically found to be consonant, and intervals slightly deviating from these ratios are found to be dissonant. The detailed psycho-acoustic mechanisms behind these perceptions remain unclear. One explanation is that small integer ratios are favored simply because of the perceived roughness between closely-spaced (but not identical) higher partials of two tones. This idea dates back to the physicist Helmholtz (2013), and was further developed in the 20th century by Plomp and Levelt (1965), and Sethares (1998). This explanation is unlikely to be the whole story. There is, for example, reason to believe that a mental process exists in which the brain attempts to match the partials coming from a sound source to a harmonic series, with better matching resulting in greater perceived consonance. This idea, developed by Paul Erlich, is mathematically described by the concept of harmonic or spectral entropy (which is distinct from the application of entropy here) (Sethares 1998, Milne et al. 2017). Both of these models make similar predictions, particularly in the case of dyads (two-tone chords). If one calculates the predicted dissonance for different intervals, one obtains similar graphs from both models, with prominent peaks and dips in the same places. One might imagine that culture and training would then overlay these innate mechanisms to modify the prominence of certain peaks or dips in the graph of dissonance, though would be unlikely to completely reshape it. For our purposes here, the mechanism behind dissonance perception is not important. All we require is some way of quantifying perceived dissonance between pairs of tones. In principle, one could perform this quantification experimentally by testing the perceptions of a test group, or theoretically, by creating a model including critical band roughness, harmonic entropy, acculturation, and so forth. Here, we follow the tradition of the physics community (as Einstein said, "Everything should be made as simple as possible, but no simpler") and use the simplest model that will suffice for our purposes, based only on critical band roughness, following Sethares.

In the critical band roughness model used here, we need to specify two things to calculate how dissonance depends on the interval between two tones: 1. The width of the critical band, and 2. the set of partials for the tones. Experimentally, the width of the critical band is observed to vary with the range of the two pitches, with lower pitches having a wider critical band. In the spirit of simplification, we fix the critical band width by specifying a reasonable value for the peak roughness and can then test how the results change as we change that width. We also make the simple assumption that the partials follow that of a sawtooth wave, a reasonable approximation to the timbre of a bowed string (Reid 2004). A sawtooth wave contains all harmonics, with the amplitude of each harmonic inversely proportional to its harmonic number.

Figure 2 shows the calculated dissonance plotted against interval size across one octave with peak roughness of 36 cents. The units of dissonance, D_0, are arbitrary, but will be the same as units of temperature as discussed below. Because we will assume that the distribution of pitches is the same in every octave, we have summed the dissonance from each octave in order to collapse the plot to a single octave. This summation results in a symmetric curve in which, for example, the perfect fifth and perfect fourth display the same dissonance. Though it is not strictly necessary to assume octave periodicity, the absence of this assumption requires one to make additional assumptions about how perceived dissonance varies with interval size and pitch range. The graph has prominent valleys that correspond to highly consonant intervals, such as the perfect fifth (3/2), perfect fourth (4/3), and major third (5/4). As the numbers in the ratio become larger, the valleys become less prominent. The valley for the minor third (6/5) is actually slightly higher than the tritone around 600 cents, and while there is a modest valley near the major second (9/8), it arises almost accidentally from how the contributions from nearby ratios add together. There are also high peaks, particularly just above the unison and below the octave, but also just above and below other consonant intervals, that correspond to very dissonant intervals.
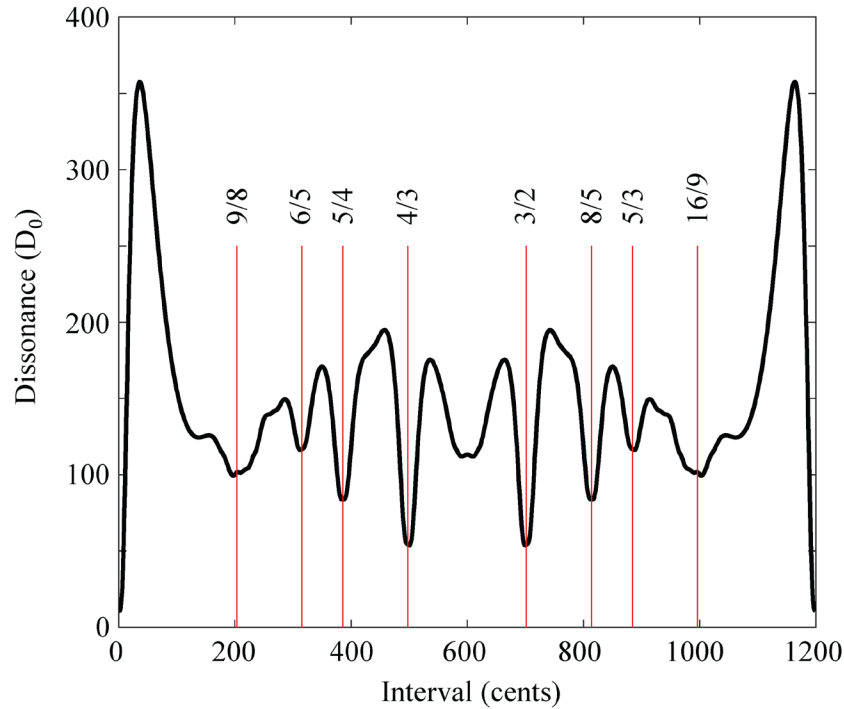
Figure 2. Dissonance function for 10 harmonic partials and critical band maximum of 36 cents. Red lines indicate $1200\log_2 r$ of the indicated ratios r.

QUANTIFYING ENTROPY

We now turn to the quantification of entropy. As described above, we will specify our tuning system by giving the probability of usage for each pitch class. Statistical mechanics then provides a well-established formula for calculating the total entropy for music composed according to these probabilities (see, for example, Pathria 2016, Chapter 3). [4]

Without going deeply into the mathematics, we can get an intuitive understanding of this formula via some examples. Recall that entropy is a measure of the number of ways the constituents of a system can be arranged, while remaining consistent with the overall state of the system. Here, the state of a tuning system is specified by the probabilities for each pitch, and the constituents are the notes that make up music composed within that system. Because we are not concerning ourselves with the rhythmic dimension of music here, imagine we have already composed a rhythm for a piece of music containing 1000 notes, and we want to assign a pitch to each note of that rhythm. The relevant question is then: How many different pieces of music can we create by assigning the pitches differently?

As a first example, let us return to the (very boring) tuning system with just a single pitch. That is, the probability of the one pitch is equal to 1, and all other pitches have probability zero. In this state, there is actually no choice at all – each note is assigned the same pitch. There is just one unique piece of music that can be generated. This represents the lowest possible entropy.

Next, we can imagine a tuning system with just two pitches, each equally likely (with probability 1/2). Now for each note in the rhythm, we have a choice of two pitches that can be assigned to that note. So, for the first note, there are two possibilities. Again, for the second note there are two possibilities, resulting in $2 \times 2 = 4$ possibilities for the first two notes together. Still two more choices for the third note results in $2 \times 2 \times 2 = 8$ possibilities for the first three notes. Following this pattern, we have $2 \times 2 \times 2... \times 2$ (1000 times), or two to the thousandth power, possible compositions for all 1000 notes. This large number of possible compositions represents a higher entropy system.

Finally, we can consider the number of pieces that could be composed on top of this rhythm in 12-TET. That is, 12 pitches, each with probability 1/12. Following the same argument above, we expect an even larger number of possible compositions: 12 raised to the 1000th power. Note that this number is larger than 1 followed by 1000 zeros. This system has even higher entropy than the cases above.

To summarize the examples above, we see that tuning systems with larger numbers of pitch classes will generally have higher entropy. But what if not all pitch classes occur with the same probability? For example, what if a tuning system favors tonal music in a particular key? In that case, we might expect the tonic pitch of that key to occur with higher probability and chromatic pitches to occur with lower probability. It can be shown mathematically that having unequal probabilities for the pitch classes results in lower entropy than with equal probabilities (see for example, Cover 1999, Chapter 12). We can understand this intuitively: if a disproportionate number of notes in a composition are taken up by one pitch class, that will result in a smaller number of possible compositions than if all pitch classes were used with equal probability.
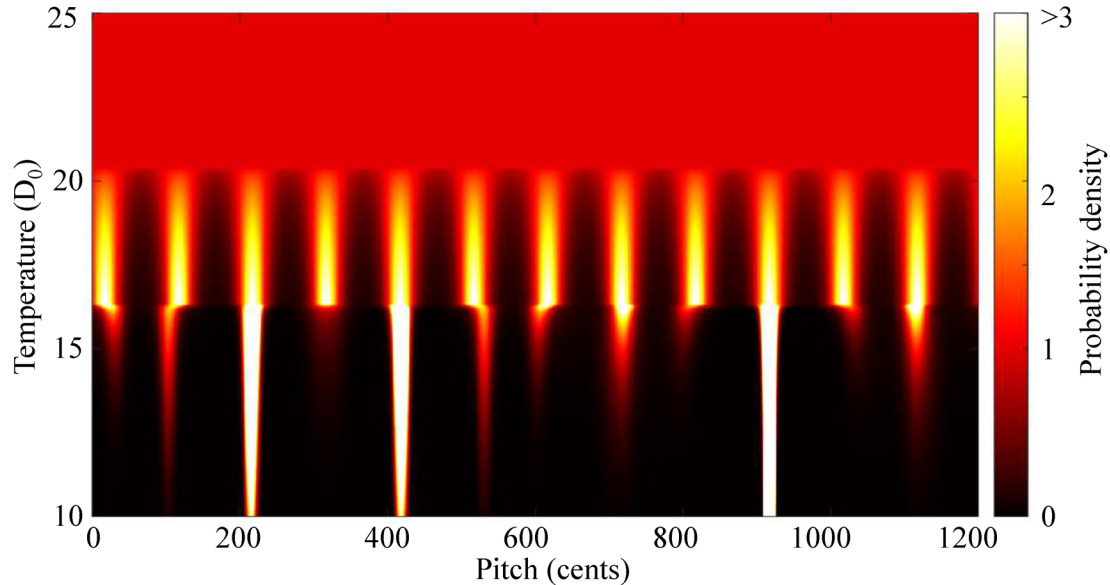

Figure 3. Result of mean field calculation at different temperatures, using the dissonance function in Figure 2. $T > 21\ D_0$ shows disordered sound, between $16.6\ D_0$ and $21\ D_0$ shows 12-TET, $< 15\ D_0$ approximates just intonation, with the range just below $16.5\ D_0$ approximating meantone temperaments.

Finally, we should note that the discussion of entropy above assumes that we have some number of discrete pitch classes, whereas the tuning system is actually described by a continuous probability distribution. So instead of a single probability, a pitch class is represented by a peak in the distribution with some width and height. But still, if pitch classes are represented by well-separated peaks with the same width, then more pitch classes mean higher entropy, with the highest possible entropy occurring when all peaks have the same height. If the width of a peak increases, then the entropy increases. Recall that the width of a peak in the probability distribution indicates the flexibility in tuning for that pitch class. It stands to reason that greater flexibility in tuning would increase the variety of possible compositions.

BALANCING DISSONANCE AND ENTROPY

With mathematical expressions for the average dissonance and entropy for a tuning system, we then turn to methods from statistical mechanics to find the tuning system that makes the optimal trade-off between low dissonance and high entropy, at a given temperature (see, for example, Kittel 1998, Chapter 3). [5] Because temperature represents an exchange rate between dissonance with units of D_0 and entropy, which is a unitless number, the temperature also has units of D_0. [6]

Figure 3 shows the calculated pitch distributions over a range of temperatures, for the dissonance curve in Figure 2 with a peak roughness at 36 cents. Black represents low probability, and brighter colors represent higher probability. Three distinct temperature regimes can be clearly seen. At high temperature (at the top), no peaks or valleys appear in the pitch distribution; it is completely disordered, with every different pitch in the continuum occurring with equal likelihood. This regime represents noise or other sounds that would generally be considered non-musical. At a temperature of $21\ D_0$, twelve peaks with equal amplitudes emerge, evenly distributed across the range of pitches. Around a temperature of $16.3\ D_0$, there is another transition where the peak heights suddenly become unequal, and the peaks shift away from the even spread.

The particular values of temperature at which the transitions occur will depend on the particular shape of the dissonance curve - the size and width of the peaks and troughs. The characteristics of the probability distributions, however, show similar behavior for different dissonance curves of this form, though possibly with a different number of pitches per octave. A survey of how different dissonance curves lead to differences in octave divisions is given in Appendix B.

We will now compare the mean field pitch distribution at different temperatures to historical tuning systems. The results are summarized in Figure 4, comparing the peaks in the mean field (M.F.) results to common tuning systems at temperatures where these peaks closely match the pitches of the tuning system. The rows of the table are divided into pairs to indicate a mean field result (darker gray rows) and historical tuning system (lighter gray rows) to compare. Pitches tuned higher than 12-TET are colored red and pitches tuned lower are colored blue. At $T = 16.6\ D_0$, the mean field calculation yields 12 equally spaced pitches, in exact agreement with 12 tone equal temperament (12-TET). As the temperature is lowered below the transition, we see reasonable agreement with meantone temperaments, finally converging towards just intonation at lower temperatures. Trends emerge in the table in Figure 4, such as the increasing detuning of the major and minor third away from equal temperament, as temperature is decreased. Further details are discussed in the following sections.

12-TET	0c	100c	200c	300c	400c	500c	600c	700c	800c	900c	1000c	1100c
M.F: T = 16.6 D_0		+0	+0	+0	+0	+0	+0	+0	+0	+0	+0	+0
1/6-comma meantone		+8	-3	+5	-6	+2	-/+10	-2	+6	-5	+3	-8
M.F.: T = 16.26 D_0		+6	-3	+5	-7	+1	0	-1	+7	-5	+3	-6
¼-comma meantone		+17	-7	+10	-14	+3	-/+20	-3	+14	-10	+7	-17
M.F.: T = 16.1 D_0		+11	-3	+10	-13	+0	+0	+0	+13	-10	+3	-11
5-limit JI		+12	+4	+16	-14	-2	-/+10	+2	+14	-16	-4	-12
M.F.: T = 12 D_0		+26	-3	+15	-14	-2	-/+7	+2	+14	-15	+3	-26

Figure 4. Comparison of mean field (M.F.) pitches to common tuning systems. The first row shows the evenly spaced pitches of 12 tone equal temperament (12-TET). Subsequent light gray rows show the deviations from 12-TET, in cents, of each pitch in three historical tuning systems. The darker gray rows show the same deviations as calculated in the mean field model, at temperatures (T) chosen to have close agreement with the historical systems. Red and blue numbers indicate positive and negative deviations from 12-TET, respectively. Two values are indicated for the augmented fourth and diminished fifth in these tuning systems, and also in the low temperature mean field result.

LOW TEMPERATURE: JUST INTONATION

At low temperatures, we expect the mean field model to strongly favor high consonance and avoid dissonance. As such, the interval with the lowest dissonance (unison/octave) produces a large peak in the pitch distribution giving rise to a particular root pitch, in this case around 920 cents. The particular frequency of the root pitch is arbitrary. An arbitrary choice of the root pitch also occurs in musical tradition, such as tuning to A440, or just intonation based on C Major. With the model favoring strong consonance, we also expect to see peaks in the pitch distribution at consonant intervals relative to the root. If those peaks become large, then additional peaks will appear that are consonant intervals away from them. The following principles summarize how this behavior appears in the locations of peaks in the model:

1. Pitches at a local minimum in the dissonance function relative to the root (1/1) are favored; this tends to include simple whole number ratios, such as 3/2 (perfect 5th), 4/3 (perfect 4th), 5/4 (major 3rd) due to roughness between partials of the root and exclude any pitch that is slightly detuned from those pitches.

2.   When pitches other than the root become prominent, pitches a consonant interval, such as a perfect fifth (3/2), a major 3rd (5/4) or their inversions, away from those are favored.

3.   If multiple close pitches equally satisfy (1) and (2), the mean field model may split the difference and choose a pitch in between (e.g. the tritone at 64/45 or 45/32).
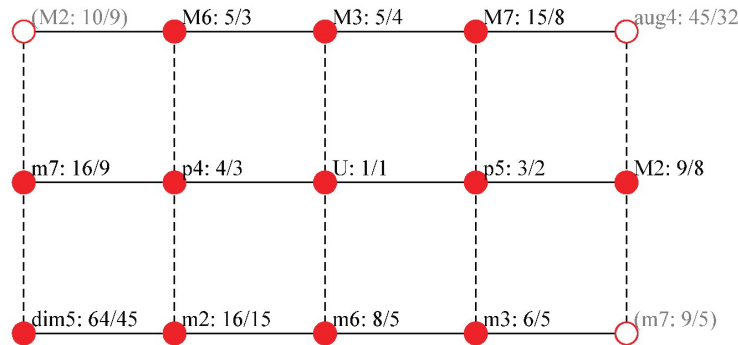

Figure 5. Fokker lattice. Solid lines indicate intervals of perfect fifths (or fourths). Dashed lines indicate intervals of major thirds (or minor sixths). Filled red dots indicate a standard 5-limit just intonation tuning. Alternative tunings for the major second, minor seventh, and tritone are shown as unfilled dots.

These principles are helpfully illustrated by a Fokker lattice, [7] shown in Figure 5, which arranges intervals in rows advancing by perfect fifths and columns advancing by major thirds (Fokker 1969). This lattice is similar to the traditional Tonnetz in which the minor thirds are also connected. For the purposes of generating the pitches in the tuning system, however, the minor thirds are redundant, as they are already included by ascending a perfect fifth and descending a major third. Principle 1 describes pitches in close proximity to the tonic pitch at the center of the lattice. Principle 2 describes pitches that have one or more connections on the lattice to other pitches with significant amplitude in the probability distribution; the more connections a pitch has, and the larger the probabilities of the pitches to which it is connected, the more that pitch will be favored. For example, the minor third does not correspond to a particularly deep valley in the calculated dissonance. The minor third above the tonic is, however, connected to two other favored pitches by a perfect fifth and a major third. Finally, principle 3 is illustrated by three alternative tunings shown by unfilled circles. The major second, minor seventh, and tritone above the tonic have two possible tunings on the Fokker lattice. The model will tend to make some compromise between these options. The pitches shown by filled red circles on the Fokker lattice form one possible system for five-limit just intonation. We will see that the low temperature mean field result tends to agree with this tuning.

Figure 6 shows a comparison between the pitch distribution in the low temperature regime and the pitches in five-limit just intonation. The corresponding numerical values for the peak positions are given in the table in Figure 4. The following describes the agreement between the mean field model and the 12-tone just tuning shown on the Fokker lattice in Figure 5.

At low temperatures, one pitch is more likely than all others, with a large peak in the probability distribution. This pitch, appearing at about 920 cents in Figure 3 and transposed to 0 cents in Figure 6, represents the tonic pitch, labeled 'C'. The next most prominent peaks occur at consonant intervals away from the tonic, including the perfect fifth (G), perfect fourth (F), major third (E), and minor sixth (G♯). These intervals occur very close to the frequency ratios of 3/2 and 4/3, above and below the root pitch, in good agreement with just intonation, according to Principle 1. As listed in Figure 4, the shifts in all these pitches relative to equal temperament match exactly within one cent.
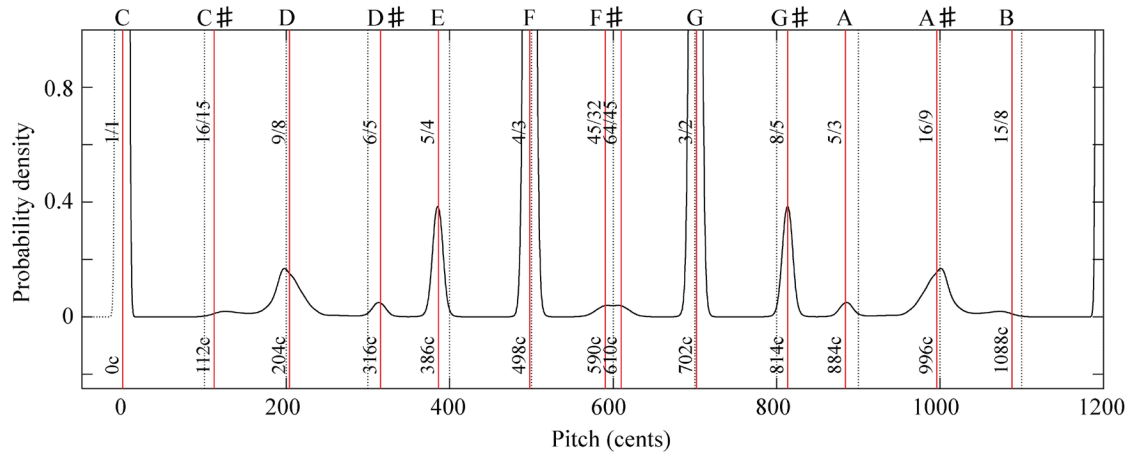
Figure 6. Mean field probability distribution at T = 12 D_0 from Figure 3 (black). Red lines indicate standard 5-limit just intonation, with pitches $1200\log_2 r$ given by the labeled ratios *r*. Vertical dotted lines show 12-TET for comparison.

A smaller peak appears at the minor third above the tonic, very close to frequency ratio 6/5 (E♭). As seen in Figure 4, the JI major third is tuned up 16 cents from ET, as compared to up 15 cents in the mean field result. Though there is a small dip in dissonance at this frequency ratio (see Figure 2), it is not enough to produce the observed peak in the probability distribution due to consonance with the tonic alone. Instead, the stronger effect is Principle 2: consonance with the perfect fifth or minor sixth relative to the tonic (G and A♭) already discussed above. These relationships are shown in the Fokker lattice in Figure 5, with the minor third connected to both the perfect fifth and minor sixth. The same reasoning also applies to explain the small peak at frequency ratio 5/3 representing the major sixth above the tonic (A), which is a perfect fifth or minor sixth away from the already added E and F.

The major 2nd (9/8, D) and augmented 6th (enharmonic minor 7th) (16/9, A♯/B♭) above the tonic are slightly off from the indicated just tunings (e.g. the major 2nd at +4 cents for JI vs. -3 cents for mean field), and the peaks in the probability distribution are relatively broad. This occurs because there are several competing factors governing the tuning of these pitches. The frequency ratios 9/8 and 16/9 each have only one connection to a pitch with large probability on the Fokker lattice (Principle 2). Alternatively, these scale degrees could be represented by ratios 10/9 or 9/5 (open circles in Figure 5). Choosing both of these alternatives is not favored: both pitches would have just one connection on the Fokker lattice, and the pitch to which it would be connected would be more distant from the root, and hence would occur with lower probability. Changing just one of them could be advantageous, though. For example, choosing the ratio 10/9 for the major 2nd, instead of 9/8 would yield two connections on the Fokker lattice instead of one, as well as providing an extra connection for the minor 7th still at 16/9.

In practice, particular choices for just intonation of the major 2nd above the tonic might be preferred in different circumstances. The ratio 9/8 allows for just intonation of the V (dominant) triad, which plays an important role in Western harmony. On the other hand, the choice of 10/9 allows just intonation of the minor ii triad with root a major 2nd above the tonic, which also appears in major harmonies. On an instrument with fixed tuning, such as the piano, a choice between these options is required. On an instrument with a free tuning, choices can be made by the performer in response to the particular circumstances. The mean field calculation presented here reflects this range of possible intonations with a relatively broad peak for the major 2nd and minor 7th above the tonic.

The augmented unison (enharmonic minor 2nd) (C♯) and major 7th (B) above the tonic both have two strong connections on the Fokker lattice, favoring the ratios (16/15) and (15/8) respectively. The mean field results, however, show only a small peak, shifted from the justly tuned values. (E.g., in JI, the major 7th above the tonic is tuned down 12 cents vs. down 26 cents in the mean field result relative to ET.) The shift of these peaks occurs because of the very strong dissonance between these pitches and the tonic pitch (or an octave above it). Recall that the mean field approximation says that any pitch may be sounded with any other, with probability given by the probability density near that value of the pitch. So despite consonance between the major 7th with the perfect fifth and the major second above the tonic (to form the V triad), the possibility of the highly dissonant semitone pushes the major seventh lower as a compromise. In practice, this

compromise is usually not necessary because the semitone is traditionally avoided. This could be regarded as a weakness of the mean field assumption, in that it has no way of specifying that some pitches in the system are more likely to be sounded together than others.

Finally, there are two peaks located near 64/45 and 45/32, corresponding to the just augmented 4th and diminished 5th. In five-limit JI, these pitches are tuned up or down 10 cents from the ET intonation, as compared to up or down 7 cents in the mean field result. These two intonations of the tritone are each connected to two other pitches on the Fokker lattice (Principle 2), and are reinforced by a coincidental minimum in the dissonance function around Huygens' tritone 7/5 (Principle 1). At higher temperature, these two features broaden, and merge into a single peak exactly at 600 cents above the tonic.

INTERMEDIATE TEMPERATURE: MEANTONE INTONATION

As temperature increases, the peaks in the mean field model begin to shift away from the low temperature configuration towards a phase transition to equal temperament. In this intermediate regime, the model agrees well with different meantone systems, as seen in the table in Table 4. The reason for the shift in peak position can be understood by drive towards higher entropy as temperature increases. Higher entropy can be achieved by beginning to equalize the amplitudes of the peaks in the pitch distribution. In the lower temperature case in the previous section, the tonic (C), and p4 and p5 above the tonic (F, G) dominate the probability distribution. In the 5-limit just intonation shown in Figure 5, all the other pitches except for the tritone are obtained as justly tuned p4/p5 or M3/m6 away from these three major pitches. Therefore, these less probable pitches are all justly tuned with the high-probability C, F, and G. Many pairs of these less probable pitches, however, form unjustly tuned p4/p5, M3/m6 intervals. If the probabilities of these pitches are low, then the occurrence of these intervals will be rare, and will not greatly increase the total dissonance. But as these peaks grow with increasing temperature, it is increasingly important to minimize the dissonance between a greater number of pitch combinations. This leads to the same compromises that give rise to meantone intonations: a sacrifice of some consonance of justly-tuned intervals to enhance the consonance of a larger number of intervals.

Around $T = 16.26$ D_0, just before the complete transition, the peaks have shifted to locations that line up well with the 1/6th comma meantone system, as seen in the third and fourth rows of the table in Figure 4. The perfect 4th and 5th above the tonic at this temperature are tuned slightly opposite, relative to ET, from their just intonation counterparts at lower temperatures; for example, the perfect 5th here is slightly lower than 700 cents, while a just perfect 5th is slightly higher. The major 2nd and minor 7th relative to the tonic now take on greater importance, as a perfect 5th away from the perfect 5th or 4th above the tonic, and are now precisely given as two perfect fifths above or below the tonic, in agreement with meantone tuning. One of the more noticeable effects of meantone tuning is the less dramatic detuning of the thirds and sixths above the tonic away from ET values. In JI, the major third is tuned down by 14 cents, a significant fraction of a semitone (100 cents). While this produces a consonant interval of 386 cents with the root, it produces quite dissonant intervals elsewhere. For example, with a tonic of C, a major third between E and G♯ would be 428 cents in JI, a full 42 cents above the justly tuned interval! In both meantone systems, and the intermediate-temperature mean field results, we see a compromise between these intervals. For example, in both 1/6th comma meantone, and the mean field results at $T = 16.26$ D_0, the major third from C to E is now 394 and 393 cents respectively, and the interval from E to G♯ is 412 or 414 cents. So the interval with the root has been detuned from the just value of 386 cents, but has moderated the mistuning of other intervals.

As the temperature changes in this transition region, we observe a gradual change of the degree to which consonance between all pitches is favored over consonance of a few significant intervals. At temperature lower than the 1/6th comma meantone example above, we see reasonably good agreement with a 1/4 comma meantone intonation (see the fifth and sixth rows of the table in Figure 4).

HIGHER TEMPERATURE: EQUAL TEMPERAMENT

For temperatures in excess of about $T = 16.3$ D_0 but below $T = 21$ D_0, the twelve peaks in the model are consistently the same height and spaced evenly across the octave pitch distribution, with 100 cents between each pitch. Unlike the lower temperature region where the pitch distribution evolved gradually with temperature, the distribution in this region remains fixed with equal spacing, and equal peak heights. This obviously reflects the 12-tone equal temperament (12-TET) used in much of modern music.

As before, the changes in the pitch distribution occur because of a trade-off between dissonance and entropy. On the one hand, 12-TET yields many intervals significantly away from their just intonations, but the uniformity of the tuning yields intervals that are uniformly mis-tuned. This uniformity yields greater compositional possibilities, without introducing too much additional dissonance. For example, major thirds are now always 400 cents, as compared to the just tuning of 386 cents. But in just intonation, justly tuned major thirds are only available relative to seven of the 12 pitch classes (see Figure 5). Use of the other five major thirds would introduce significantly more dissonance. In 12-TET though, all 12 major thirds can be used equally with the trade-off of a moderate increase in dissonance. With the same reasoning applying to every interval, we see that 12-TET represents a compromise between the consonance of intervals and the range of potential compositions.

The sudden jump from the continuously evolving low-temperature region to the 12-TET region represents a phase transition, analogous, for example, to the transition from ice to water. Different phases represent fundamentally different ways of organizing their constituents, and are often characterized by a distinct symmetry or periodicity. In the low temperature phase, the only symmetry in the pitch distribution is the assumed octave-wise periodicity. That is, the only transposition of the pitch distribution that leaves the tuning system the same is by one or more octaves. This low-symmetry condition remains true even as the distribution evolves from just intonation through meantone systems. But after the sudden transition to 12-TET, there is greater symmetry: a transposition by any number of semitones leaves the pitch distribution unchanged. This symmetry must be enforced at all temperatures in the equal temperament phase, resulting in the same even spacing and peak heights throughout that temperature range. This higher symmetry is directly connected with the increased entropy of the higher-temperature phase. Another hallmark of a phase transition is a change in the type of arbitrary choices that specify the phase. In the just intonation phase, a particular arbitrary tonic pitch is chosen (C, in the examples above), whereas in the equal temperament phase, no tonic pitch is selected, but the absolute tuning of the 12 pitches must still be chosen (for example, by setting an A to 440 Hz.)

HISTORICAL EVOLUTION OF TUNING SYSTEMS

As the temperature is increased in the model, the changing trade-off between consonance and expanded compositional possibilities is generally mirrored by the historical development of tuning systems. Though the historical development of temperaments occurred gradually and without universal consensus (Barbour 2004), some of the broad trends in the evolution of temperament are reproduced here. Likewise, the mechanism at work in our model is similar to the way in which temperaments have evolved to accommodate changing trends of increased modulation, chromaticism, and atonality over time.

At low temperature, the pitch distribution is dominated by the tendency for consonant intervals relative to one arbitrary tonic pitch. Moreover, the distribution is dominated by the tonic, and a justly-tuned p4 and p5 above the tonic. This is suggestive of one of the earliest documented tuning systems, attributed to Pythagoras, which was based on the perfect fifth and octave, utilizing mathematical ratios (2:1 for octave, 3:2 for fifth) to tune scales. This system, which can be mathematically described as 3-limit JI, resulted in major thirds that were sharper than just intonation. The mean field result, while capturing a justly tuned fifth, does not produce a pitch distribution arising from a continuous series of justly tuned fifths. Instead, the pitch distribution more closely matches the 5-limit JI tuning discussed above, in which thirds and sixths are also justly tuned with respect to the tonic. This more closely resembles tunings advocated by Ptolemy, who notably believed that tuning should satisfy both the ear and mathematical ratios. Ptolemy's advocacy of just intonation marked another pole as compared to Pythagorean tuning, finding truth in both ratio and perception. While these 5-limit JI systems did not see widespread use for tuning of fixed-pitch instruments, performers on flexible-pitch instruments, including the human voice, are free to use such intonation at will.

Figure 7. Excerpt from Prelude in C major, T 517, William Byrd (ca. 1613). Published by London & Leipzig: Breitkopf und Härtel, 1899, public domain.

During the flowering of polyphony in the late Middle Ages and Renaissance, various unequal temperaments on keyboard and fretted instruments arose organically through the gradual modification of Pythagorean tuning. The promulgation of mean tone temperament with its compromise of pure thirds within a delimited tonal spectrum prevailed for two centuries. For a historical example illustrating this style, see the excerpt from Prelude in C major, William Byrd (ca. 1613) in Figure 7. A single accidental is seen in the excerpt shown, and the large majority of pitches in the piece are drawn from the seven pitch classes of this key. To illustrate this, we plot the prevalence of each pitch class in the piece in Figure 8(a). The pitch classes are arranged around the circle of fifths, with the tonic (C) at the top, and the sizes of the blue circles indicate the cumulative time each pitch class is sounded. Clearly, the seven consecutive points corresponding to the diatonic major scale are by far the most prevalent. We also calculate the "center of effect" as an average of the coordinates of the points around the circle of fifths, weighted by the prevalence of that pitch class, and plotted as the purple dot in Figure 8(a). This approach is adapted from the key-finding technique based on the spiral array model developed by Chew (2014). The center of effect will illustrate the degree of modulation or atonality present in a piece, as we will see below.
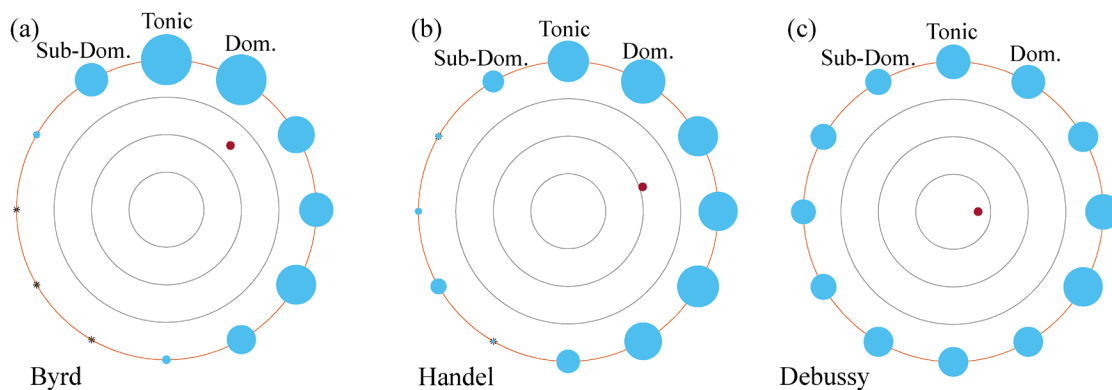

Figure 8. Size of blue circles indicates cumulative time sounded for each pitch class in the pieces excerpted in (a) Figure 7 (b) Figure 9, and (c) Figure 10, with the opening key (C Major, C Major, and E Major, respectively) tonic placed at the top position on the circle of fifths. Purple dots show the center of effect, calculated as the weighted average position of the blue circles.

As the temperature begins to increase, pitches away from the tonic become more prominent and hence intervals relative to those other pitches become more significant. In order to yield reasonably consonant intervals not just with the tonic, but with other pitches as well, compromises in intonation are favored, resulting in somewhat higher dissonance, but with compositional possibilities less restricted to one tonal center in the sense that modulation and chromaticism will still produce relatively acceptable frequency ratios. As the possibilities of modulation continued expanding in the Baroque era, so too did the appetite for a tuning that could accommodate them. This birthed an era of sophisticated irregular temperaments and circulating tunings designed to optimize certain tonal regions. As an example of Baroque music that takes advantage of these changes in temperament, several deviations from the tonic key are seen in the excerpt from Fantasia In C Major HWV 490, Handel (1703-1706) in Figure 9. The distribution of pitch classes for the entire piece (Figure 8(b)) still shows the seven pitch classes of the tonic key appearing most commonly, but with several others also occurring with greater frequency than in the Byrd example in Figure 8(a). Consequently, the center of effect, again plotted as a purple dot in Figure 8(b) is somewhat closer to the center of the circle of fifths.

Figure 9. Excerpt from Fantasia in C Major, HWV 490, Handel (1703-1706). Published in Georg Friedrich Händels Werke. Band 2, Leipzig: Deutsche Händelgesellschaft, 1858. Plate H.W. 2, public domain.

Finally, a transition to equal temperament occurred. At the expense of further dissonance, compositional possibilities were greatly expanded, now treating all tonal centers on an equal footing, permitting unlimited tonal modulation and indeed the abandonment of tonality altogether. As in the model, the 12 tone equal temperament system has remained unchanged for an extended period of time, because its symmetry does not allow the types of shifts seen in the lower temperature regime. The excerpt in Figure 10, from a piano arrangement of Prélude à l'après-midi d'un faune, Debussy (1894), in fact makes use of all 12 pitch classes in a short span. The prevalence of pitches in the entire piece (Figure 8(c)) is now fairly evenly distributed among all 12 pitch classes, and the center of effect is near the center of the circle of fifths.

We can see a general movement toward equalization of pitch between the Byrd, Handel, and Debussy examples in Figure 8. As one moves into the 20th century, further equalization is often noted, such as in the first movement of Schoenberg's Op. 11, which shows pitch-class circulation almost identical to that of an equally long set of random notes and pitch-class distributions with even lower emphasis on any specific pitch (Tymoczko 2010).
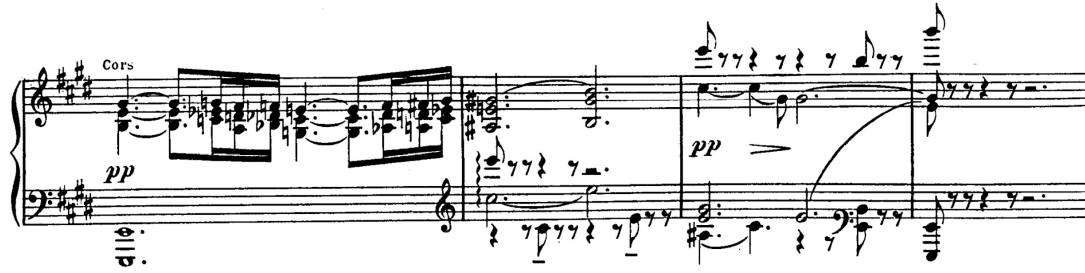
Figure 10. Excerpt from Prélude à l'après-midi d'un faune, Debussy (1894). Piano arrangement by Borwick, score in public domain.

We now apply the above center-of-effect analysis to a broader corpus of pieces, and compare to the results of the mean field model. We used the Yale-Classical Archives Corpus (YCAC) (White 2016) to extract the pitch class prevalence from 9620 pieces and calculate the corresponding center of effect within the circle of fifths. The YCAC contains note onset and durations extracted from MIDI files contributed by users of the website classicalarchives.com. The archive also includes metadata for each piece, including year (or approximate range of years) of composition, and the opening key of the piece as determined by experts. 4.6% of pieces were missing date information, and were excluded from the analyses below. When an approximate range of years was given, the midpoint of that range was used. The three examples shown in Figure 8 were also drawn from this corpus.

We first show the center of effect for all pieces in the corpus composed by Byrd, Handel, Beethoven, and Debussy (Figure 11(a)). For these four composers, chosen to exemplify their time periods, 90, 206, 373, and 39 pieces or separate movements were analyzed, respectively. The center of effect was calculated for each piece or movement, relative to the opening key tonic and plotted in Figure 11(a) as a circle colored according to the legend. The dates in the legend indicate the year ranges in which these pieces were composed, according to the YCAC.
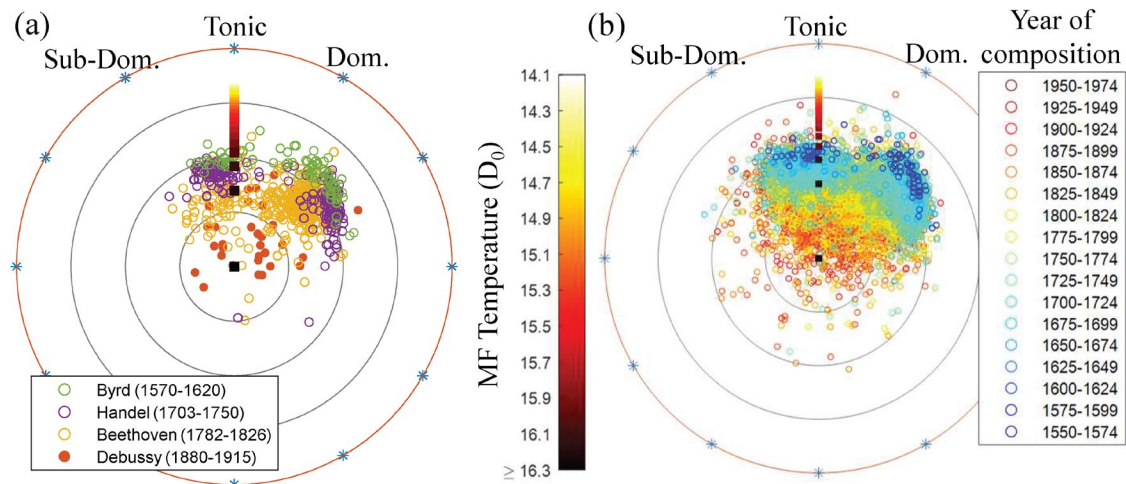

Figure 11. Corpus analysis of pitch class prevalence center of effect on the circle of fifths. (a) Center of effect for pieces by four exemplar composers (circular points), compared to center of effect of pitch class prevalence from the mean field model vs. temperature (square points). (b) Center of effect for 9620 pieces colored by year of composition (circular points), compared to center of effect of pitch class prevalence from the mean field model (square points).

The center-of-effect points show a clear evolution from Byrd, to Handel, to Beethoven, to Debussy. The pieces by Byrd are largely in a single key with limited modulation, and cluster in two groupings corresponding to major and minor modes. Pieces by Handel show similar groupings, but with greater spread towards the center of the circle as modulations became more common. The pieces by Beethoven show greater

modulation still. Though many of the Beethoven points overlap the regions occupied by Byrd and Handel, there is again a shift towards the center, and a greater spread with some compositions reaching very close to the center of the circle. The pieces by Debussy show an even greater spread with many scattered around the center of the circle, indicating frequent and distant modulation, or atonality.

The shift of the center of effect towards the center of the circle of fifths from Byrd to Debussy is also seen in the mean field results as temperature is increased. At each temperature point, we calculate the prevalence of each pitch class as the integrated area under each of the 12 peaks in the mean field result. As with the pieces analyzed above, we then calculate the center of effect within the circle of fifths from the weighted average of the coordinates of each pitch class. The tonic pitch, placed at the top of the circle, is chosen to be the most prevalent pitch. The resulting center of effects are plotted as squares in Figure 11(a) with temperature indicated according the color bar. At lower temperature, the tonic pitch is much more prevalent than all others, and the center of effect approaches the point at the top of the circle of fifths. As the temperature increases, the other pitches become more prevalent and the center of effect shifts towards the center of the circle. Once the transition to equal temperament has occurred, all pitch classes are equally likely, and all subsequent center-of-effect points lie at the center of the circle.

The same trends as observed for the four exemplars shown in Figure 11(a) can be seen when plotting the center of effect for all pieces in the YCAC, color coded by year of composition, as shown in Figure 11(b). Here, we obtain the center of effect for all pieces or separate movements in the YCAC with date information in the same way as above. These center of effects are plotted as circles, color coded according to the legend at the right. The dark blue points at the earliest years are tightly grouped in two regions again corresponding to highly tonal pieces mostly in a single key in the major and minor modes. It should be noted that the large majority of pieces from 1550-1600 in the YCAC are by Byrd. Then as time progresses, we see the spread of center of effect towards the center of the circle. The square points in Figure 11(b) show the same center of effect points from the mean field results as in Figure 11(a).
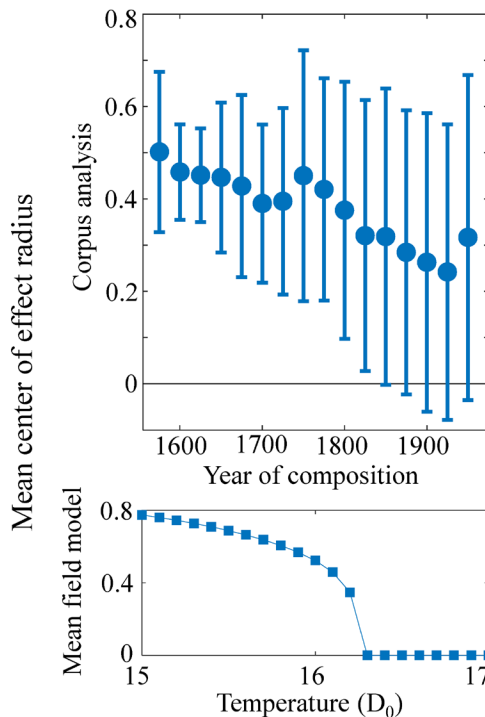

Figure 12. Evolution of average circle-of-fifths center of effect. Top: Data points show the radius of average center of effect in each time period shown in Figure 11(b), with error bars indicating the spread of center of effect radii from the average value. Bottom: Center of effect radius vs. temperature in the mean field model.

The data shown in Figure 11(b) are summarized in Figure 12. Here, we plot the radius of the average center of effect for each time bin. The top plot of Figure 12 shows the average center-of-effect radius (blue dots) calculated from the pieces in the YCAC in each 25-year time bin. A radius of 1 lies on the circle of

fifths, and radius of zero is at the center. The error bars indicate the range of center-of-effect points within that time bin with respect to the average value. The error bars are calculated with a method that takes into account not just the spread of radii, but also of the angle so that two diametrically opposed points with same radius are not considered equivalent (see [8]). Here we can see the trend of center of effect moving towards the center of the circle (smaller radius) as time progresses. The error bars in Figure 12 demonstrate the increasing range of center of effect with time, where tonal pieces largely in a single key remain present (at larger radius) but with increasing spread towards the center of the circle. We compare to the results from the mean field calculation in the lower panel of Figure 12. The square points plot the radii of the square points shown in Figure 11. The temperature range is chosen to cover the range discussed above that reproduces the shift from just intonation, to meantone temperament, to equal temperament. As meantone temperaments emerge, the center of effect moves towards the center of the circle. And after the transition to equal temperament, the center of effect remains at zero radius.

　　　The results in Figure 11 and Figure 12 show that the trend of pitch class prevalence in the mean field model roughly tracks the expanding frontier of increasing modulation and atonality over time. Similar evolutions can be seen in previous corpora studies (Albrecht and Huron 2012). It is worth mentioning that our results do not imply, for example, that all Renaissance music was highly consonant, nor that all music composed today is highly atonal. In the earliest time period shown here, there are certainly more dissonant examples (some of Gesualdo's madrigals, for example, which are not included in the YCAC). But it is worth noting, for example, that the eight Gregorian modes, along with the four added by Glareanus in 1547, show significantly lower levels of chromaticism than, say, a typical late Romantic piece. Likewise, as we approach the present day, we do not simply see a shift of the center of effect towards the center, but a spreading so that pieces cover a range spanning the region mainly occupied in the 16th century all the way to the center of the circle. As such, while 12-TET does indeed place all keys on equal footing, it is not superior for all music in the sense of interval consonance. Rather, as we see in our model, it is a consequence of a need for the optimal expression of a certain compositional language, as well as the practical restrictions of certain instruments. Indeed, earlier temperaments could be deemed superior for not only much music of their time, but for some modern music as well, particularly among some minimalist compositions, where the need for flexibility in pitch language is low and would be arguably superseded by the gains in interval consonance.

　　　Thus far we have not discussed the highest temperature phase, above temperature 21 D_0. Here, the pitch probability distribution is constant, representing equal likelihood for all frequencies of sound. This may be interpreted as non-musical sound, but alternatively may be viewed as a further development of music away from discrete pitches, to music that makes use of continuous frequency spectra. While elements of sound not based on discrete pitches associated with the tonal system have existed in music for centuries, it was in the 20th century that their usage became central in certain pieces and later, in certain genres. Luigi Russolo's experiments with noise music in the 1910s, particularly his self-made Intonarumori, helped to stretch the boundaries of music as organized sound through his intention to "amalgamate the most dissonant, strange, and harsh sounds. In this way, we come ever closer to noise-sound."

　　　Interest in the incorporation of non-musically pitched elements continued with pieces like Varese's *Ionization* in 1931. Henri Pousseur's *Scambi* (1957) is notable in that it used white noise as its basis, arguably the furthest thing from organized collections of discrete pitches. The use of noise expanded beyond classical composition as well, particularly with the advent of electronic music. Artists like Merzbow and Incapacitants use noise as the central compositional tool and helped create the Japanoise genre. Bands like Sonic Youth incorporated the use of noise, feedback, distortion, and unpitched gestures, techniques that would become quite popular in rock and punk music in the 1980s and 1990s. Meanwhile, composers like Xenakis developed techniques for compositional synthesis divorced from the rigidity of the discrete pitch system. Spectralism, although distinct from the use of noise in that it still generally focuses on the use of distinct frequencies, also arose, and in the mathematical sense, sits somewhere between a 12-tone system and pure noise. While a full enumeration of the different compositional threads that employed noise in the 20th century is beyond the scope of the article, one thing is clear: the use of frequencies outside the traditional 12-tone collection, both as a compositional tool and as a primary compositional focus, increased significantly in the aforementioned period.

　　　It is both surprising and satisfying that, by placing no restriction besides octave periodicity on the pitches used in this model, the model results in 12-tone tunings that accurately match tuning systems used for centuries by musicians. By examining why the solutions to the mean field model here match these historical tuning systems, we gain a new window into the reasons behind these systems of intonation.

CONCLUSIONS AND OUTLOOK

The success of this physics-based approach to reproduce tunings previously crafted by strictly musical means motivates us to use this model to investigate other potentially undiscovered tunings which could serve to produce new kinds of music. By changing parameters that specify the dissonance, we can obtain other types of ordering. By changing parameters such as the amplitudes of harmonic partials and the width of the critical band that specifies the interval between pitches at maximum perceived dissonance, we observe different numbers of pitches per octave, such as 5, 7, 19, 22, 31, 41, and 53. These octave divisions have all been explored in different contexts, such as the 31 equal divisions of the octave (EDO) developed by Huygens (Rasch 1986) or a 53-EDO description of classical Turkish music (Yarman 2007). (See appendix B, Figs. B1-B3 for a survey of octave divisions for different parameters.) Alternatively, by changing the timbre to only include odd harmonic partials and partially relaxing the assumption of octave periodicity, we find that Bohlen-Pierce tuning emerges instead of traditional tuning. In Bohlen-Pierce tuning, the "octave" is defined as a tripling of frequency, instead of a doubling, with 13 pitches per "octave" (Bohlen 2001). Even more significant changes are observed for timbres with non-harmonic partials, that is, partials that are non-integer multiples of the fundamental frequency. Such timbres are characteristic of instruments such as drums, xylophones, or metallophones, where the sound is not produced by a linear string or column of air. As described by Sethares (1998), the dissonance curves for non-harmonic timbres generally have minima in different locations than harmonic timbres. This is borne out in the mean-field calculations, where we see that non-harmonic timbres generally do not have a 12-EDO solution. Figure B4 in appendix B shows the octave divisions observed for non-harmonic partials approximating the spectrum of the bonang, an Indonesian metallophone, and the pong lang, a Thai xylophone, as shown in Sethares (1998).
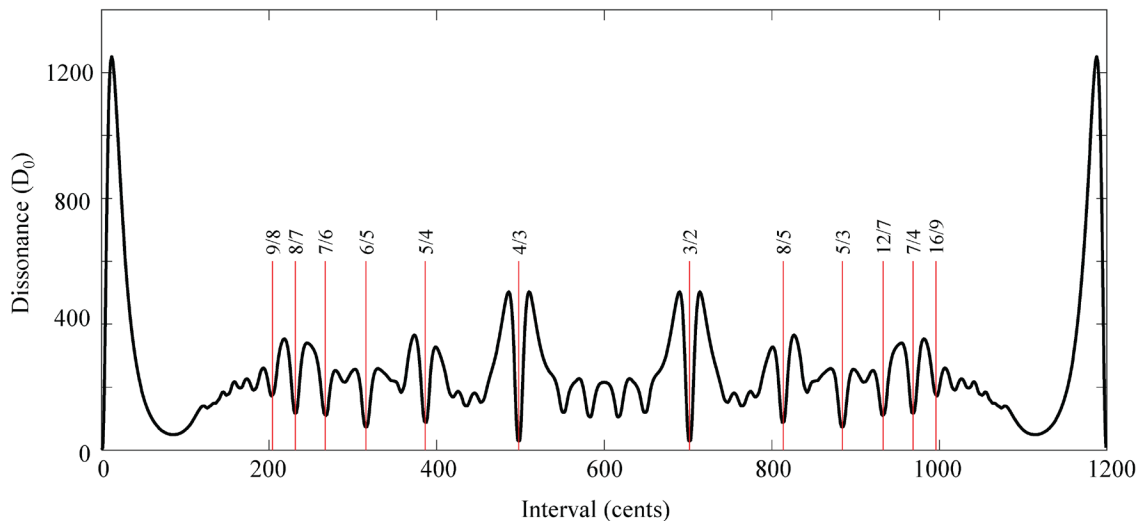

Figure 13. Dissonance function for 14 harmonic partials and critical band maximum of 12 cents.

As an example of another tuning that can emerge from the model, we demonstrate the 31-fold octave arising from a harmonic sawtooth wave with narrower critical bandwidth. With the maximum dissonance at an interval of 12 cents, we obtain a dissonance function with much sharper features (Figure 13). In addition to prominent dips at the perfect intervals and the thirds, the dissonance function now has prominent minima at the so-called supermajor second (8/7) and subminor third (7/6), therefore implying 7-limit just intonation with prime ratios incorporating factors of 7. The mean-field model with this new dissonance function, shown in Figure 14 now yields 31 equal divisions of the octave (31-EDO) at high temperature, with a transition to unequal spacing at lower temperatures. 31-ET saw some popularity among Dutch composers in the mid-20th century.
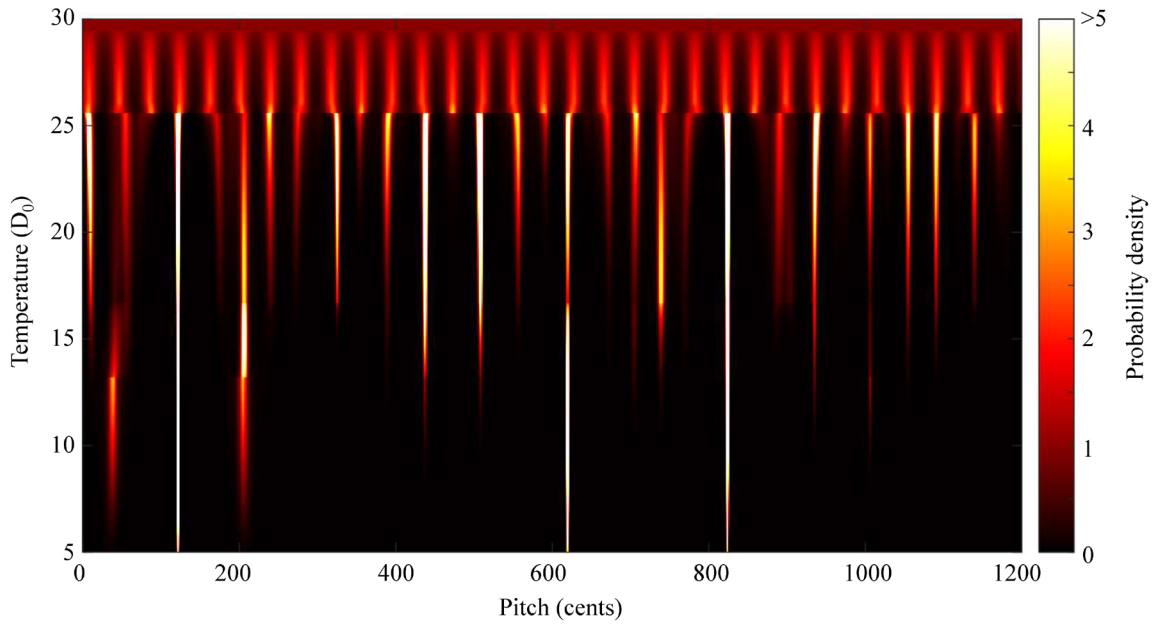
Figure 14. Result of mean field calculation at different temperatures, using the dissonance function in Figure 13. $T > 29.5$ D_0 shows disordered sound, between 25.6 D_0 and 29.5 D_0 shows 31-TET, < 25.5 D_0 represents unequal 31-tone temperaments.

The behavior of the 31-EDO mean field results can be understood using the same concepts as the 12-fold case. At the lowest temperature, one root pitch is more probable than the rest. Other pitches are most likely to occur at pitches that occur at small integer ratios, now including numbers up to seven, such as 8/7 or 7/6. As the temperature increases, and pitches other than the root become more probable, the consonance of intervals that do not include the root becomes more significant, leading to a shift to meantone tunings. Finally, the increased weight given to compositional variety (entropy) at higher temperature results in a phase transition to 31-tone equal temperament.
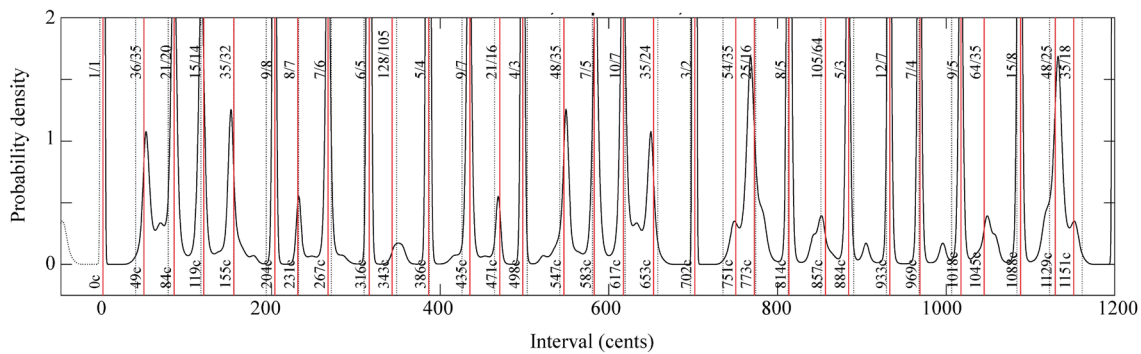

Figure 15. Mean field probability distribution (black) at $T = 23$ D_0 from Figure 14. Red lines indicate a possible 7-limit just intonation, with pitches $1200\log_2 r$ given by the labeled ratios r. Vertical dotted lines show 31-TET for comparison.

The pitch distribution at a low temperature is shown in Figure 15, where the distribution is mainly reflecting just intonation. As above, a Fokker lattice can be used to understand the ratios present in this tuning system. In the seven-limit case, however, the lattice is now represented in three dimensions, and there are many more alternative selections of 31 points on the lattice that lead to reasonable tuning systems. We see that the mean field calculation chooses one particular set of ratios on the seven-limit Fokker lattice, indicated by the labeled ratios in Figure 15. This tuning, like the 12-fold just intonation, has only octavewise translational symmetry. Interestingly, whereas the 12-fold JI had mirror symmetry about the tonic and the

tritone, the 31-fold tuning shown here is symmetric about points near 350c and 950c. We have observed that the particular mean field solutions in the 31-fold just intonation region depend sensitively on the particular details of the timbre and critical band width, resulting in many interesting potential tunings to explore.

It is important to note again that the mean field assumption is not without weaknesses. The primary weakness is that it does not account for the fact that certain combinations of pitches may be used more or less frequently than would be predicted by their independent probabilities. For example, the major 7th scale degree (B) is, in practice, used frequently as a leading tone or within the V triad. The mean field result in Figure 6, however, shows only a very small peak for the major 7th because the mean field approximation does not admit the possibility that B would be sounded together with G (forming a major third) more frequently than the more probable C (forming a semitone). As noted above, this is also why the small peaks at the minor 2nd and major 7th are tuned anomalously high and low, respectively. Going forward, more sophisticated methods in statistical mechanics may be employed to avoid this issue. For example, Din and Berezovsky (2023) use numerical simulations of tones on a lattice to study ordering of pitches with only nearest-neighbor interactions.

These results demonstrate that the application of statistical mechanics methods both provides a new view of the historical development of tuning systems, and sheds light on new tuning systems to be explored. Throughout history, humans have arranged sounds to make music by balancing the competing desires for pleasing consonant combinations of tones, and for complexity and variation. In doing so, certain systems of intonation have emerged from the disordered universe of sounds. The results here suggest that the systems emerging from this process are not arbitrary inventions, but the result of the same type of natural process that causes unique, yet similarly ordered snowflakes to spontaneously form, or that results in the slow growth, deep in caves, of intricate crystals that also awe us with their ordered beauty.

ACKNOWLEDGEMENTS

NOTES

[1] Correspondence can be addressed to: Jesse Berezovsky, Case Western Reserve University, Department of Physics, 2076 Adelbert Rd., Cleveland, OH 44106. E-mail: jab298@case.edu.

[2] Technically, the energy of the system changes by exchanging energy with its environment.

[3] Probabilities can be assigned to discrete outcomes, such as probability 0.5 (50%) for a coin landing on heads or tails. When the outcome is not discrete, but instead a continuous spectrum, we instead describe how probabilities are distributed across that spectrum, using a "probability distribution." The probability distribution can be plotted against the range of outcomes. A peak in the probability distribution indicates that the range of outcomes where the peak is large is more likely than another range where the probability distribution is smaller.

[4] The entropy for each note is equal to the sum of each probability times the natural logarithm of its inverse. The total entropy is then obtained by multiplying by the total number of notes.

[5] Specifically, we multiply the entropy by the temperature, then subtract it from the average dissonance to obtain a quantity known in physics as "free energy." We then find the pitch probability distribution that minimizes that quantity.

[6] In physics, temperature and energy are traditionally expressed in different units (e.g. Kelvin and Joules in SI units). This is, however, a historical convention and necessitates a conversion factor (the Boltzmann constant) to convert units of temperature to units of energy. Here we avoid this unnecessary complication.

[7] In this case, the Fokker lattice is a so-called "5-limit" lattice, in which the ratios change by factors of 3 in one direction and factors of 5 in the other direction, always scaled by factors of 2 (octaves) to remain within one octave.

[8] The average center of effect and error bars in Figure 12 were obtained as follows. Given a set of center of effects vectors $\vec{c}_i = (x_i, y_i)$, we obtain the mean center of effect $\langle \vec{c} \rangle = (\langle x \rangle, \langle y \rangle)$, as the average of the x and y coordinates. The plotted radius is then the magnitude $|\langle \vec{c} \rangle|$. Error bars are calculated as 95% confidence intervals of the projection of each center of effect onto the mean center of effect. We obtain these projections $p_i = \vec{c}_i \cdot \langle \vec{c} \rangle / |\langle \vec{c} \rangle|$. Note that these p_i will be negative if the angle between $\langle \vec{c} \rangle$ and $\vec{c}_i$ is greater than 90°. The total range of the 95% confidence interval is then obtained using the conversion factor from standard error to 95% confidence interval of 3.92 times the standard deviation of the p_i.

REFERENCES

Albrecht, J. D. and D. Huron (2012). A statistical approach to tracing the historical development of major and minor pitch distributions, 1400-1750. *Music Perception 31*(3), 223–243. https://doi.org/10.1525/mp.2014.31.3.223

Barbour, J. M. (2004). Tuning and Temperament: A Historical Survey. Dover Publications.

Berezovsky, J. (2019). The structure of musical harmony as an ordered phase of sound: A statistical mechanics approach to music theory. *Science Advances 5*(5), eaav8490. https://doi.org/10.1126/sciadv.aav8490

Berg, R. E. and D. G. Stork (1990). *The Physics of Sound*. Pearson Education India.

Bohlen, H. (2001). 13 tone steps in the twelfth. *Acta Acustica united with Acustica 87*(5), 617–624.

Chew, E. (2014). Mathematical and computational modeling of tonality. AMC, 10(12), 141. https://doi.org/10.1007/978-1-4614-9475-1

Cover, T. M. (1999). *Elements of information theory*. John Wiley & Sons.

Din, H. and J. Berezovsky (2023). Critical behavior and the Kibble-Zurek mechanism in a musical phase transition. *PLOS ONE*, *18*(1), e0280227. https://doi.org/10.1371/journal.pone.0280227

Fokker, A. D. (1969). Unison Vectors and Periodicity Blocks in 3-Dimensional (3-5-7-) Harmonic Lattice of Notes. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen Series B - Physical Sciences 72*(3), 153.

Helmholtz, H. (2013). *On the sensations of tone (trans.)*. (Original work published 1877): Courier Corporation.

Huron, D. (1994). Interval-class content in equally tempered pitch-class sets: Common scales exhibit optimum tonal consonance. *Music Perception*, *11*(3), 289-305. https://doi.org/10.2307/40285624

Huron, D. (2006). Sweet anticipation–Music and the psychology of expectation. A Bradford book. https://doi.org/10.7551/mitpress/6575.001.0001

Kittel, C., & Kroemer, H. (1998). *Thermal physics*.

Lindley, M. (2001). Temperaments. *The New Grove Dictionary of Music and Musicians 16*, 205–206. https://doi.org/10.1093/gmo/9781561592630.article.27643

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

Meyer, L. B. (2008). *Emotion and meaning in music*. University of Chicago Press.

Milne, A. J., Bulger, D., and Herff, S. A. (2017). Exploring the space of perfectly balanced rhythms and scales. Journal of Mathematics and Music, 11(2-3), 101–133. https://doi.org/10.1080/17459737.2017.1395915

Milne, A. J., Smit, E. A., Sarvasy, H. S., and Dean, R. T. (2023). Evidence for a universal association of auditory roughness with musical stability. *PLOS ONE*, *18*(9), e0291642. https://doi.org/10.1371/journal.pone.0291642

Pathria, R. K. (2016). *Statistical mechanics*. Elsevier.

Plomp, R. and W. J. M. Levelt (1965). Tonal consonance and critical bandwidth. *The Journal of the Acoustical Society of America 38*(4), 548–560. https://doi.org/10.1121/1.1909741

Proverbio, A. M., A. Orlandi, and F. Pisanu (2016). Brain processing of consonance/dissonance in musicians and controls: a hemispheric asymmetry revisited. *European Journal of Neuroscience 44*(6), 2340–2356. https://doi.org/10.1111/ejn.13330

Rasch, R. (Ed.). (1986). Christiaan Huygens, Le cycle harmonique (Rotterdam 1691), Novus cyclus harmonicus (Leiden 1724).

Reid, G. (2004). Practical bowed-string synthesis. Retrieved from https://www.soundonsound.com/techniques/practical-bowed-string-synthesis

Sethares, W. A. (1998). *Tuning, timbre, spectrum, scale*. Springer. https://doi.org/10.1007/978-1-4471-4177-8

Tenney, J. (1988). *A history of consonance and dissonance*. Excelsior.

Tenney, J. (2008). On 'Crystal Growth' in harmonic space (1993–1998). *Contemporary Music Review*, *27*(1), 47–56. https://doi.org/10.1080/07494460701671525

Terhardt, E. (1984). The concept of musical consonance: A link between music and psychoacoustics. *Music perception*, *1*(3), 276–295. https://doi.org/10.2307/40285261

Tymoczko, D. (2010). *A geometry of music: Harmony and counterpoint in the extended common practice*. Oxford University Press.

White, C. W., & Quinn, I. (2016). The Yale-classical archives corpus. *Empirical Musicology Review*, *11*(1). https://doi.org/10.18061/emr.v11i1.4958

Yarman, O. (2007). A comparative evaluation of pitch notations in Turkish makam music. *Journal of Interdisciplinary Music Studies*, *1*(2), 43–61.

APPENDIX A
Derivation of mean field pitch distribution

Throughout this derivation, a pitch with fundamental frequency f is denoted on a logarithmic scale where $x = \log_2 f/f_0$, with f_0 an arbitrary constant. We parameterize the roughness between a pair of pure tones with frequency f_1 and f_2 as

$$d(f_1, f_2) = \left(\frac{1}{w_c}\right) e^{-\left[\ln\left(\frac{|\Delta x|}{w_c}\right)\right]^2},$$

where $\Delta x = \log_2 f_1/f_2$ and w_c is the critical bandwidth. This choice of d has a similar shape as those described in Sethares (1998), while also having a smooth derivative at $\Delta x = 0$. The function takes on a value of zero at $\Delta x = 0$ and $\Delta x \to \infty$, with maxima at $\Delta x = \pm w_c$. We then take non-pure tones to be characterized by a fundamental with frequency f_0 and amplitude A_0, and a set of partials of frequency $f_n = \phi_n f_0$ and amplitude $A_n = \alpha_n A_0$. We sum over all pairs of partials to obtain the dissonance of two non-pure tones as

$$D(f_1, f_2) = \sum_{n,m} l_{nm} d(\phi_n f_1, \phi_m f_2) = D(\Delta x)$$

where $l_{nm} = \min(\alpha_n, \alpha_m)^{0.606}$ accounts for the loudness of the pair of partials, as given by Sethares (1998). With a fixed w_c, D depends only on Δx, not on the absolute frequencies.

The mean field equilibrium pitch distribution $P(x)$ is obtained by finding the $P(x)$ that minimizes $F = D_t - TS$, where D is the mean dissonance between pairs of pitches, S is the entropy arising from the distribution $P(x)$ and T is the temperature. We assume that $P(x)$ is periodic with period of 1 (an octave). Thus, instead of integrating over all real x, we define $D_p(\Delta x) = \sum_{n=-\infty}^{\infty} D(\Delta x + n)$ and integrate only over one octave. Specifically, we obtain the mean dissonance

$$D_t = \frac{1}{2} \int_0^1 \int_0^1 P(x) D_p(x - y) P(y) dx dy,$$

and entropy

$$S = -\int_0^1 P(x) \ln P(x)\, dx.$$

We then use a variational method to find extrema of $F = D_t - TS - \mu N$, where $N = \int_0^1 P(x) dx - 1$, and μ is a Lagrange multiplier. This additional third term will be zero at the extrema, and ensures that the $P(x)$ is properly normalized. A standard variational method is implemented by calculating $F_0 + \delta F$ by evaluating F for $P(x) = P_0(x) + \delta P(x)$ in the limit of small $\delta P(x)$. An extremum is obtained when $\delta F = 0$. This yields the integral equation

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

Solutions to this equation are found numerically, using an iterative approach described by Berezovsky (2019).

A python Jupyter notebook that implements the above calculations is available at https://github.com/rtbuechele17osu/Crystals-of-Sound.git, as well as data files for the results described in this paper.

APPENDIX B
Summary of results for other critical bandwidths and timbres

Here we present additional results for different values of the critical bandwidth w_c and different timbres specified by the frequency and amplitude of their partials.

To demonstrate the robustness of the octave orderings, we first modify the harmonic sawtooth timbre used in the main discussion above. A sawtooth tone is defined by harmonic partials $\phi_n = n$ with $\alpha_n = \frac{1}{n}$, where ϕ_n and α_n are defined in Appendix A, and $n = 1,2,\dots n_{max}$. We compare to two other prototypical waveforms, the square wave and the triangle wave. The square wave also characterized by $\phi_n = n$ and $\alpha_n = \frac{1}{n}$ but only for odd n. Similarly, the triangle wave also has odd harmonic partials $\phi_n = n$ for odd n and $\alpha_n = 1/n^2$. We use the calculation presented in Appendix A to find the N-TET ordering that first appears as the temperature is lowered from the disordered state, at a particular w_c. Figure B1 shows number lines depicting a range of w_c with regions numbered by the N-TET ordering in that region of w_c. In all cases $n_{max} = 10$. We see similarity in all three waveforms, with the boundaries shifting slightly, and the disappearance of 7-TET for the square waveform.

Figure B2 illustrates the robustness of the results against random variations in the partial amplitudes. Here, we start with the sawtooth waveform, but then randomly alter the amplitude of each partial by a factor uniformly distributed between 0.5 and 1.5. We run the calculation for four separate choices of these random factors. The results are quite consistent, with only small shifts in the boundaries between different orderings.

In Figure B3, we show the effect of changing the number of partials n_{max} using the sawtooth waveform. As n_{max} changes, the boundaries between different N-TET orderings fluctuate in w_c with most of the same values present. The most prominent changes are the occasional disappearances of 10-TET or 22-TET for some values of n_{max}.

Finally, we present results for two non-harmonic timbres, where the partials are non-integer multiples of the fundamental frequency. As described by Sethares (1998) non-harmonic timbres are often encountered in non-Western music in which instruments such as xylophones and metallophones are common. Figure B4 shows the resulting octave divisions for partials estimated from spectra given by Sethares (1998) for the bonang, a metallophone used in gamelan music, and the pong lang, a xylophone used in traditional Thai music. Specifically, the bonang is modeled with partials $\phi_n = \{1, 1.52, 3.46, 3.92\}$ and $\alpha_n = \{1, 0.8, 0.7, 0.6\}$, and the pong lang is modeled with partials $\phi_n = \{1, 1.47, 2.85, 5.48, 8.88\}$ and $\alpha_n = \{1, 0.2, 0.3, 1, 0.2\}$. The results in Figure B4 indeed show different octave divisions than in the harmonic examples above with the notable absence of 12-TET. Instead, we see prominent ranges of 5-TET and 7-TET which are in fact used in gamelan and traditional Thai music.
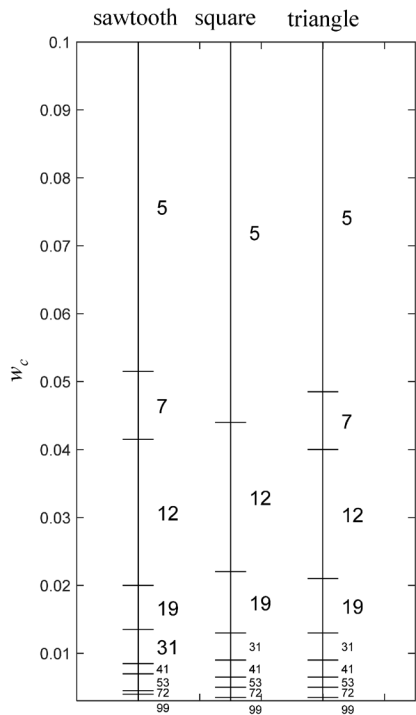
Figure B1. N-TET octave divisions observed in different ranges of w_c for three characteristic waveforms.
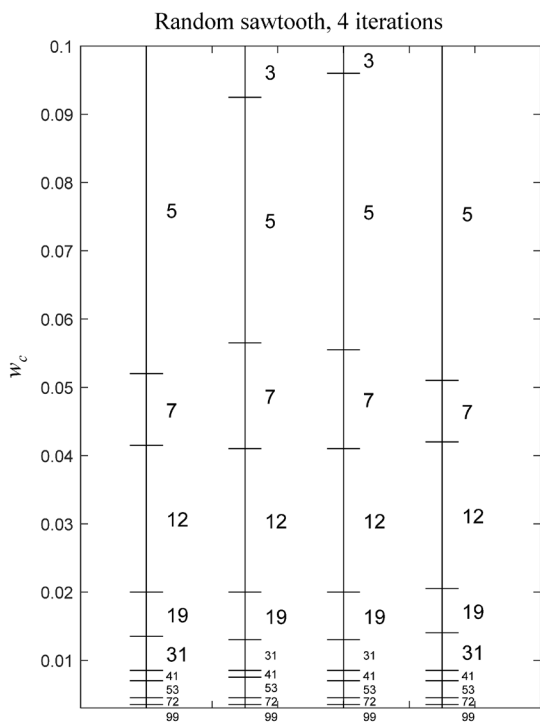

Figure B2. N-TET octave divisions observed in different ranges of w_c for four realizations of sawtooth wave partials with the amplitude of each partial randomly adjusted from 50-150% of its original value.
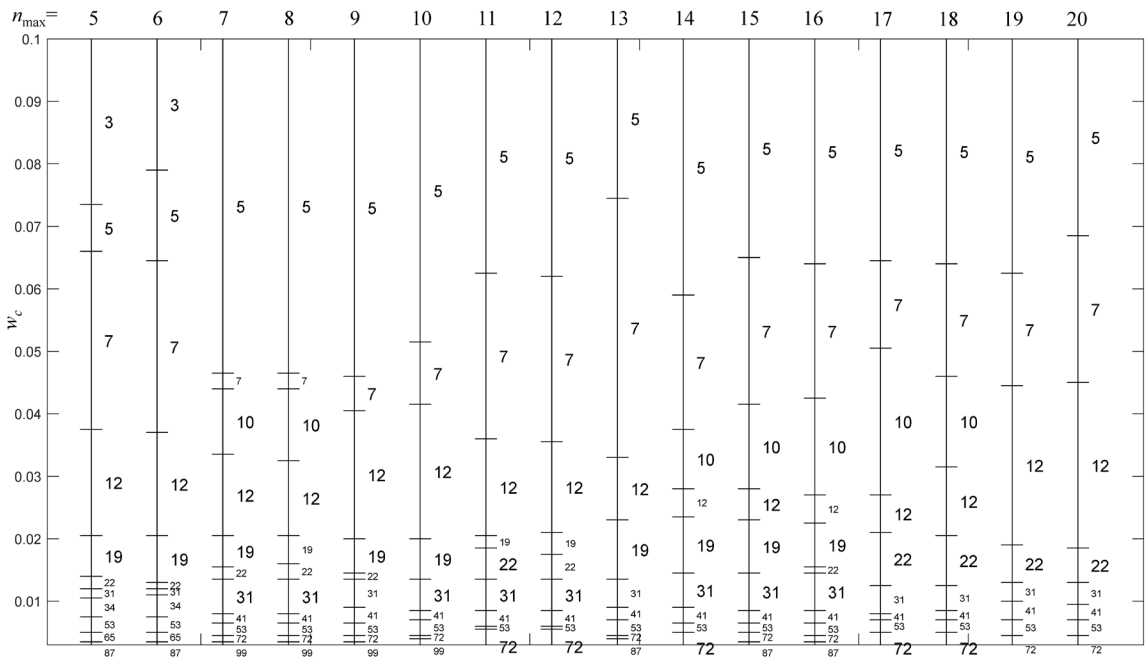
Figure B3. N-TET octave divisions observed in different ranges of w_c for a sawtooth waveform with partials cut off above n_{max}.
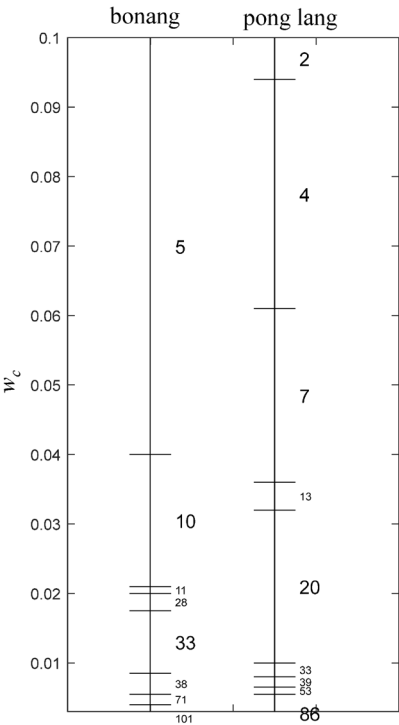

Figure B4. N-TET octave divisions observed in different ranges of w_c for two non-harmonic sets of partials.