www.sciencemag.org/cgi/content/full/313/5783/72/DC1

Supporting Online Material for

The Geometry of Musical Chords

Dmitri Tymoczko*

*To whom correspondence should be addressed. E-mail: dmitri@princeton.edu

Published 7 July 2006, *Science* **313**, 72 (2006)
DOI: 10.1126/science.1126287

This PDF file includes:

Materials and Methods

Figs. S1 to S12

Table S1

Movies S1 to S4

Soundfile S1

References

SUPPORTING ONLINE MATERIAL

TABLE OF CONTENTS

1. Overview of the argument. Our goal is to model composers' judgments about the relative sizes of voice leadings. One cannot assume that these judgments will be consistent with any mathematical norm or metric (§3). However, it is reasonable to stipulate that a theoretical model should reflect widely recognized features of Western music. Since musicians treat transposition and inversion as preserving musical distance, I will require that voice-leading comparisons be invariant under transposition and inversion of any of their individual musical voices (§2). Since Western pedagogues instruct composers to minimize voice-leading size while eschewing "voice crossings" (*1–3*), I will require that "crossed" voice leadings be no smaller than their natural, uncrossed alternatives (§§3–4). These minimal requirements suffice to establish a number of interesting results about voice leading (§§6–8).

Sections 2 and 3 articulate mathematical constraints on methods of comparing voice leadings. Section 4 shows that these constraints formalize my two musical requirements. Section 5 derives the voice-leading orbifolds. Sections 6 and 7 relate the structure of a chord to its voice-leading possibilities. Section 8 introduces a further constraint on methods of comparing voice leadings, and uses it to derive a polynomial-time algorithm for finding a minimal voice leading (not necessarily bijective) between arbitrary chords.

2. Pitch and pitch class. Pitches are modeled as real numbers. The distance between two pitches p, q is the absolute value of their difference, $|q - p|$. Pitch classes are modeled as points in the quotient space $\mathbb{R}/12\mathbb{Z}$. They are sets of real numbers $\{p + 12k \mid k \in \mathbb{Z}\}$, with p representing some pitch in the pitch class. We can label these sets using real numbers in the range $0 \leq x < 12$. The elements of $\mathbb{R}/12\mathbb{Z}$ form a group under addition of their labels modulo $12\mathbb{Z}$. The distance between pitch classes a and b, written $\|b - a\|_{12\mathbb{Z}}$, is the smallest nonnegative real number x such that, if p is a pitch belonging to pitch class a, then either $p + x$ or $p - x$ belongs to pitch class b.

Let $(p_1, p_2, ..., p_n) \rightarrow (q_1, q_2, ..., q_n)$ be a voice leading between multisets of pitches, and let $(a_1, a_2, ..., a_n) \rightarrow (b_1, b_2, ..., b_n)$ be a voice leading between multisets of pitch classes. I will say that the two voice leadings are *associated* if the pitches p_i and q_i all belong to the pitch classes a_i and b_i, respectively. The *displacement multiset* of the voice leading $(p_1, p_2, ..., p_n) \rightarrow (q_1, q_2, ..., q_n)$ is the multiset of distances $\{|q_i - p_i|\}$. Similarly, the displacement multiset of the voice leading $(a_1, a_2, ..., a_n) \rightarrow (b_1, b_2, ..., b_n)$ is the multiset of distances $\{||b_i - a_i||_{12\mathbb{Z}}\}$. I will require that a method of comparing voice leadings be invariant under transposition and inversion of its individual voices: in any voice leading, one can replace the voice (p, q) with $(q + x, p + x)$ without changing the voice leading's size. It follows that the size of a voice leading depends only on its displacement multiset. I will also require that the size of a displacement multiset be nondecreasing in each of its members: we can replace any member of a displacement multiset with a smaller number without increasing the size of the resulting voice leading (§3). This ensures that a voice leading between multisets of pitch classes will be the same size as the smallest associated voice leading between multisets of pitches.

3. Comparing voice leadings. A method of comparing voice leadings is a relation "≥" over multisets of nonnegative real numbers that is reflexive, transitive, and total (a total preorder; see Table S1 for more information). The relation must satisfy what I call the *distribution constraint:*

$$\{x_1 + c, x_2, ..., x_n\} \geq \{x_1, x_2 + c, ..., x_n\} \geq \{x_1, x_2, ..., x_n\}, \text{ for } x_1 > x_2, c > 0$$

(**NB**: since multisets are unordered, the numerical subscripts do not have ordinal significance: x_1 is no more "first" than x_2 or x_n.) The first inequality requires that the total preorder not consider an uneven distribution of values to be smaller than a more even distribution with the same total sum: if X is an n-member displacement multiset whose members sum to x, then $\{x, 0, ..., 0\} \geq X \geq \{x/n, x/n, ..., x/n\}$. Thus, x semitones of motion in a single voice yields at least as large a voice leading as x semitones of motion distributed over multiple voices. This requirement is a weakened relative of the triangle inequality (*4*). The distribution constraint's second inequality requires that the size of a multiset be nondecreasing in each of its members: increasing the size of any number in a multiset never decreases that multiset's size. If a total preorder satisfies both inequalities strictly, I will say that it *strictly satisfies* the distribution constraint.

Every music-theoretical method of comparing voice leadings satisfies the distribution constraint.

A. Smoothness. The size of a voice leading is the sum of the objects in the displacement multiset (*5–7*). Thus {2, 2} > {3.999} > {1, 1, 1}. Smoothness is sometimes called the "taxicab norm." It reflects aggregate physical distance on keyboard instruments. Smoothness satisfies the distribution constraint non-strictly.

B. L^p vector norms. Smoothness is analogous to the L^1 vector norm, though the components of vectors are ordered whereas the members of displacement multisets are not. The analogues to the L^p vector norms strictly satisfy the distribution constraint for finite $p > 1$. The Euclidean vector norm L^2 has been used by Callender (*8*).

C. Semitonal and stepwise voice leadings. According to the L^∞ vector norm, the size of a displacement multiset is its largest member. The musical terms "semitonal voice leading" and "stepwise voice leading" refer to this measure of voice-leading size. Semitonal voice leadings have an L^∞ norm of 1; stepwise voice leadings have an L^∞ norm less than or equal to 2. The L^∞ norm measures the largest physical distance moved by any single voice on a keyboard instrument. It satisfies the distribution constraint non-strictly.

D. Parsimony. Parsimony is related to the lexicographic ordering. It generalizes a notion introduced by Richard Cohn and developed by Jack Douthett and Peter Steinbach (*9–10*). Given two voice leadings, α and β, α is smaller (or "more parsimonious") than β iff there exists some real number d such that

1) for all real numbers $c > d$, c appears the same number of times in the displacement multisets of α and β; and
2) d appears fewer times in the displacement multiset of α than β.

Thus, according to Parsimony, $\{3 + \varepsilon\} > \{3, 2, 1\}$ (in this case, $d = 3 + \varepsilon$), and $\{4, 3, 3\} > \{4, 3\}$ (here, $d = 3$). Parsimony strictly satisfies the distribution constraint.

The first three methods of comparing voice leadings resemble familiar mathematical metrics. Parsimony, however, does not. This is because there is no function from multisets to real numbers, such that $f(X) \geq f(Y)$ if and only if $X \geq Y$ according to Parsimony. Nevertheless, Parsimony represents a musically viable way of thinking about voice-leading size.

4. Minimal voice leadings and voice crossings. This section connects the formalism of §§2–3 to the avoidance of voice crossings. I begin by demonstrating that if a method of comparing voice leadings obeys the distribution constraint, then there is a minimal bijective voice leading between any two chords that is crossing free. I then argue in the opposite direction, showing that if a method of comparing voice leadings violates the distribution constraint, some crossed voice leading will be smaller than its natural uncrossed alternative. I conclude with a few remarks about the relation between the avoidance of voice crossings and judgments about voice-leading size.

Let $(p_1, p_2, ..., p_n) \rightarrow (q_1, q_2, ..., q_n)$ be a voice leading between multisets of pitches. The voice leading has no voice crossings (is crossing free) if $p_i > p_j$ implies $q_i \geq q_j$ for all i, j. I will say that it is *strongly* crossing free if $(p_1 + 12k_1, p_2 + 12k_2, ..., p_n + 12k_n) \rightarrow (q_1 + 12k_1, q_2 + 12k_2, ..., q_n + 12k_n)$ is crossing free, for all integers k_i. Thus one cannot introduce crossings into a strongly crossing-free voice leading simply by shifting the octaves in which its voices appear. A voice leading between multisets of pitch classes is crossing free if it is associated with a strongly crossing-free voice leading between multisets of pitches in which no voice moves by more than six semitones. Intuitively, a voice leading is crossing free if the notes of the source chord can be connected to those of the target along minimal-length line-segments that intersect only at their endpoints.

The following theorem shows that one can always find a minimal voice leading that is crossing free. Though I state the argument using chords of pitches, it generalizes straightforwardly to chords of pitch classes.

> **THEOREM 1.** Let P and Q be any two n-note multisets of pitches, and let our method of comparing voice leadings be a total preorder satisfying the distribution constraint. Then there will exist a minimal bijective voice leading from P to Q that is crossing free. If the total preorder strictly satisfies the distribution constraint, then every minimal bijective voice leading from P to Q will be crossing free.

Suppose $(p_1, p_2, ..., p_n) \rightarrow (q_1, q_2, ..., q_n)$ contains a crossing, with $p_1 < p_2$ and $q_1 > q_2$ [Fig. S7(*a*)]. Let (r, r) be the intersection, in $\mathbb{R}^2$, of the line segment $(p_1, p_2) \rightarrow (q_1, q_2)$ with the line $x = y$. Since a line's slope is constant, $|r - p_2|/|r - p_1| = |q_2 - r|/|q_1 - r|$. There are three cases: either both numerators are greater than their denominators, both are equal to their denominators, or both are less than their denominators. In all three cases, the distribution constraint implies that, for any numbers d_i,

$$\{|q_1 - r| + |r - p_1|, |q_2 - r| + |r - p_2|, d_1, ..., d_m\} \geq \{|q_1 - r| + |r - p_2|, |q_2 - r| + |r - p_1|, d_1, ..., d_m\},$$

since the multiset on the left is more uneven than that on the right. By construction $|q_1 - r| + |r - p_1| = |q_1 - p_1|$ and $|q_2 - r| + |r - p_2| = |q_2 - p_2|$. By the triangle inequality,

$|q_1 - p_2| \le |q_1 - r| + |r - p_2|$ and $|q_2 - p_1| \le |q_2 - r| + |r - p_1|$. Thus, using the distribution constraint again, $\{|q_1 - p_1|, |q_2 - p_2|, d_1, ..., d_m\} \ge \{|q_1 - p_2|, |q_2 - p_1|, d_1, ..., d_m\}$ and the voice leading $(p_1, p_2, ..., p_n) \rightarrow (q_1, q_2, ..., q_n)$ is no smaller than $(p_1, p_2, ..., p_n) \rightarrow (q_2, q_1, ..., q_n)$. We do not increase the size of a voice leading when we replace the crossed voices (p_1, q_1) and (p_2, q_2) with the uncrossed (p_1, q_2) and (p_2, q_1).

Suppose now that our total preorder strictly satisfies the distribution constraint. In this case, I claim that the crossed voice leading $(p_1, p_2, ..., p_n) \rightarrow (q_1, q_2, ..., q_n)$ is strictly larger than the uncrossed $(p_1, p_2, ..., p_n) \rightarrow (q_2, q_1, ..., q_n)$. This can be proved by contradiction. Suppose the two voice leadings were the same size. Since the total preorder strictly satisfies the distribution constraint, $|r - p_2| / |r - p_1| = |q_2 - r| / |q_1 - r| = 1$. Furthermore, $|q_1 - r| + |r - p_2| = |q_1 - p_2|$ and $|q_2 - r| + |r - p_1| = |q_2 - p_1|$. Therefore, $|q_1 - p_1| = |q_1 - p_2|$ and $|q_2 - p_1| = |q_2 - p_2|$. In other words, q_1 and q_2 are each equidistant from p_1 and p_2. This contradicts the hypothesis that the voices (p_1, q_1) and (p_2, q_2) cross. We conclude that if a method of comparing voice leadings strictly satisfies the distribution constraint, then removing any voice crossing will make the resulting voice leading strictly smaller.

It remains to be shown that replacing (p_1, q_1) and (p_2, q_2) with (p_1, q_2) and (p_2, q_1) does not introduce additional crossings into the voice leading. Recall that $p_1 < p_2$ and $q_1 > q_2$. Now suppose some voice (p_3, q_3) crosses (p_1, q_2) [Fig. S7(*b*)]. Then either $p_3 < p_1$ and $q_3 > q_2$, in which case (p_3, q_3) crosses (p_2, q_2); or $p_3 > p_1$ and $q_3 < q_2$, in which case (p_3, q_3) crosses (p_1, q_1). Similarly, if (p_3, q_3) crosses (p_2, q_1) it crosses either (p_1, q_1) or (p_2, q_2). Finally, suppose (p_3, q_3) crosses both (p_1, q_2) and (p_2, q_1). Then, either $p_3 < p_1$ and $q_3 > q_1$, or $p_3 > p_2$ and $q_3 < q_2$; in either case (p_3, q_3) crosses both (p_1, q_1) and (p_2, q_2). The voice (p_3, q_3) therefore crosses at least as many voices in the old, crossed voice leading as it does in the new, uncrossed one. Replacing (p_1, q_1) and (p_2, q_2) with (p_1, q_2) and (p_2, q_1) therefore decreases the total number of crossings in the voice leading.

This establishes Theorem 1. The result immediately generalizes to chords of pitch classes. ■

I now argue in the opposite direction, showing that violations of the distribution constraint imply that some crossed voice leading is smaller than its natural uncrossed alternative.

Suppose a method of comparing voice leadings depends only on the displacement multiset and violates the distribution constraint's first inequality. That is, some voice leading with displacement multiset $\{x_1 + c, x_2\}$ is smaller than some voice leading with displacement multiset $\{x_1, x_2 + c\}$, with $x_1 > x_2$ and $c > 0$. It follows that the crossed voice leading $(p, p + x_1 - x_2) \rightarrow (p + x_1 + c, p + x_1)$ is smaller than the uncrossed voice leading $(p, p + x_1 - x_2) \rightarrow (p + x_1, p + x_1 + c)$ [Figure S8(*a*)]. Removing the crossing therefore increases the size of the voice leading. Additional voices do not affect the logic of the argument.

Now suppose that a method of comparing voice leadings violates the distribution constraint's second inequality: some voice leading with displacement multiset $\{x_1 + c, x_2, x_3\}$ is smaller than some voice leading with displacement multiset $\{x_1, x_2, x_3\}$, with $x_3 \geq x_2 > 0$ and $c > 0$. Consider the uncrossed voice leading

$$(p + x_1 + x_2, p + x_1 + x_3 + c) \rightarrow (p + x_1, p + x_1 + c).$$

Suppose we decide to add pitch p to the first chord, and to map it to some note in the second chord. As shown in Figure S8(b), the crossed alternative

$$(p, p + x_1 + x_2, p + x_1 + x_3 + c) \rightarrow (p + x_1 + c, p + x_1, p + x_1 + c)$$

is smaller than the uncrossed $(p, p + x_1 + x_2, p + x_1 + x_3 + c) \rightarrow (p + x_1, p + x_1, p + x_1 + c)$. Thus we can conclude that, for multisets with three or more nonzero members, total preorders satisfying the distribution constraint are the *only* methods of comparing voice leadings that are transpositionally and inversionally invariant, and that embody the principle that voice crossings do not make a voice leading smaller.

It remains possible that composers considered crossed voice leadings to be smaller than uncrossed voice leadings, yet avoided crossings for reasons unrelated to voice-leading size. There are two reasons to reject this hypothesis, however. First, if there were a conflict between the goal of avoiding voice crossings and the goal of minimizing voice-leading size, there would presumably be some evidence of this in the musical, theoretical, or pedagogical literature. In particular, one would not expect voice-leading preferences to be even approximately invariant under transposition of their individual musical voices: when voices were close together, the goal of avoiding voice crossings would trump the goal of minimizing voice-leading size, while when voices were farther apart minimal voice leadings could be used freely. Western music would therefore manifest two distinct sets of voice-leading preferences, depending on whether voices were close together or far apart. There is no evidence that this is so. No pedagogue or theorist has ever proposed a method of comparing voice leadings that favors voice crossings. Second, methods of comparing voice leadings that favor crossings tend to be undesirable for independent reasons: they cannot, for example, represent aggregate or maximal keyboard distance (§3)—physical quantities that presumably help shape composers' intuitive judgments of voice-leading size. Violations of the distribution constraint are also violations of the triangle inequality (*4*) and hence conflict with basic geometric intuitions about distance. Finally, it is difficult to develop simple, robust, and musically plausible heuristics that help composers to reduce voice-leading size when using total preorders that violate the distribution constraint. By contrast, when a method of measuring voice-leading size obeys the distribution constraint, then a composer can always reduce a voice leading's size simply by removing crossings. We therefore have good reason to require that the goal of avoiding voice crossings not conflict with the goal of minimizing voice-leading size.

5. Derivation of the voice-leading orbifolds. I now turn from methods of comparing voice leadings to the geometry of musical chords. I begin by deriving Figure 2 in the main text. Figure S9(*a*) shows the 2-torus $\mathbb{T}^2$, representing the space of ordered pairs of pitch classes. To form the space of unordered pairs we need to identify all points (x, y) and (y, x). As can be seen from Figure S9(*a*), this involves folding the 2-torus along the diagonal line-segment AB. The result is a triangle with two sides identified, shown in Figure S9(*b*). This figure is a Möbius strip. To see why, cut Figure S9(*b*) along the line-segment CD, creating two detached triangles. Then attach AC on one triangle to CB on the other, turning one piece of paper over so that chords on the shared edge match. The result is the main text's Figure 2. Figure S4 shows how Figure 2 tiles the infinite plane representing ordered pairs of pitches.

I now proceed more abstractly, deriving the orbifolds $\mathbb{T}^n/\mathcal{S}_n$ for arbitrary n. Since $\mathbb{R}^n/12\mathbb{Z}^n$ is the n-torus $\mathbb{T}^n$, we can describe $\mathbb{T}^n/\mathcal{S}_n$ as the quotient of $\mathbb{R}^n$ by the semidirect product $12\mathbb{Z}^n \rtimes \mathcal{S}_n$. We can construct our orbifolds by describing fundamental domains of $12\mathbb{Z}^n \rtimes \mathcal{S}_n$ in $\mathbb{R}^n$, and showing how their boundary points are to be identified.

To construct a fundamental domain of $\mathcal{S}_n$ in $\mathbb{R}^n$, consider only those points with coordinates in nondecreasing order: $\{(x_1, x_2, ..., x_n) \mid x_1 \le x_2 \le ... \le x_n\}$. To incorporate the $12\mathbb{Z}^n$ action, require that $x_n \le x_1 + 12$, and $0 \le \Sigma x_i \le 12$. The resulting fundamental domain is an n-dimensional prism whose base is an $(n–1)$-dimensional simplex. To see why, observe that the inequalities $x_1 \le x_2 \le ... \le x_n \le x_1 + 12$ define an $(n–1)$-simplex in every plane $\Sigma x_i = c$. Every point in the fundamental domain lies in exactly one of these simplexes and every one of these simplexes is either entirely within, or entirely outside of, the fundamental domain. Since addition by $(d, d, ..., d)$ sends the simplex in the plane $\Sigma x_i = c$ to the simplex in the plane $\Sigma x_i = c + nd$, the vector $(1, 1, ..., 1)$ points in the direction of the "height" coordinate of the prism. Figure S6 illustrates three lower-dimensional cases.

It remains to be determined how the two simplicial faces of the prism are to be identified. Define the function

$$\boldsymbol{O}(x_1, x_2, ..., x_n) = (x_2 - 12/n, x_3 - 12/n, ..., x_n - 12/n, x_1 + 12 - 12/n).$$

$\boldsymbol{O}$ is an automorphism of the prism that cyclically permutes the vertices of the simplex in each plane $\Sigma x_i = c$. By repeatedly applying $\boldsymbol{O}$ to a chord X, we can obtain chords X, $\boldsymbol{O}(X)$, $\boldsymbol{O}^2(X)$, ... $\boldsymbol{O}^{n-1}(X)$, all related by transposition, and with pitch classes all summing to the same value. In Euclidean space, $\boldsymbol{O}$ is a rotation when the prism has odd dimension, and a rotation-plus-reflection otherwise. On Figure S6(*a*), $\boldsymbol{O}$ reflects the square around the line at its center. On Figure S6(*b*), $\boldsymbol{O}$ rotates the prism around the central line. On

Figure S6(c), $\boldsymbol{O}$ is a rotation-plus-reflection that cyclically permutes the vertices of each tetrahedron.

Now suppose that $(x_1, x_2, ..., x_n)$ lies in the $\Sigma x_i = 0$ face of the prism (the base). Then $(x_1 + 12/n, x_2 + 12/n, ..., x_n + 12/n)$ is the corresponding point on the opposite face. $\boldsymbol{O}(x_1 + 12/n, x_2 + 12/n, ..., x_n + 12/n) = (x_2, x_3, ..., x_n, x_1 + 12)$ also lies on that same face. When we disregard order and octave, $(x_1, x_2, ..., x_n)$ and $(x_2, x_3, ..., x_n, x_1 + 12)$ represent the same chord, and should be identified. Thus for any point on the base of the prism x and corresponding point on the opposite face y, we identify x with $\boldsymbol{O}(y)$. Figures S5 and S6 illustrate, as does a free computer program written by the author (*11*).

6. Efficient voice leading, chord structure, and symmetry. This section relates the internal structure of a chord to its voice-leading capabilities. Let $\boldsymbol{F}$ be any function over pitch classes. I begin by showing how the internal structure of a chord A determines whether the chords A and $\boldsymbol{F}(A)$ can be linked by efficient bijective voice leading. I then show how a chord's voice-leading capabilities relate to its distance from nearby symmetrical chords. Suppose $\boldsymbol{G}$ is an isometry of pitch class space, and $S_{\boldsymbol{G}}$ is invariant under $\boldsymbol{G}$. I first show that the size of any voice leading from A to $S_{\boldsymbol{G}}$ sets an upper bound on the size of the minimal voice leading(s) from A to $\boldsymbol{G}(A)$. I then argue in the opposite direction, showing that the size of a bijective crossing-free voice leading from A to $\boldsymbol{G}(A)$ sets an upper bound on the size of the smallest of the voice leadings from A to any chord $S_{\boldsymbol{G}}$ that is invariant under $\boldsymbol{G}$. Since there is always a minimal voice leading that is crossing free (§4), these results show that there is a close relationship between the size of the minimal bijective voice leading(s) from A to $\boldsymbol{G}(A)$ and the size of the smallest of the bijective voice leadings from A to the chords $S_{\boldsymbol{G}}$ that are invariant under $\boldsymbol{G}$.

In §§6–7, I write $\mathbf{T}_x(a) = a + x$ to refer to transposition by x semitones, and $\mathbf{I}_x(a) = x - a$ to refer to the inversion that sends 0 to x. If $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ are two functions over pitch classes, and $A = \{a_1, a_2, ..., a_n\}$ is any chord, then I will say that $\boldsymbol{F}_1$ *resembles* $\boldsymbol{F}_2$ (with respect to A) when $\boldsymbol{F}_1(a_i) \approx \boldsymbol{F}_2(a_i)$ for all a_i. Finally, I will say that a chord A is invariant under σ if there is a trivial voice leading $(a_1, a_2, ..., a_n) \rightarrow (a_{\sigma(1)}, a_{\sigma(2)}, ..., a_{\sigma(n)})$.

Observation A. Let $\boldsymbol{F}$ be any function over pitch classes. Chords A and $\boldsymbol{F}(A)$ can be linked by efficient, bijective voice leading if and only if some permutation of A's notes resembles $\boldsymbol{F}$ (with respect to A).

Any bijective voice leading from A to $\boldsymbol{F}(A)$ can be written $(a_1, a_2, ..., a_n) \rightarrow (\boldsymbol{F}(a_{\sigma(1)}), \boldsymbol{F}(a_{\sigma(2)}), ..., \boldsymbol{F}(a_{\sigma(n)}))$ where σ is some permutation. This voice leading will have displacement multiset $\{\|\boldsymbol{F}(a_{\sigma(i)}) - a_i\|_{12\mathbb{Z}}\}$, or, equivalently, $\{\|\boldsymbol{F}(a_i) - a_{\sigma^{-1}(i)}\|_{12\mathbb{Z}}\}$. The size of this voice leading depends on the difference between the effects of the function $\boldsymbol{F}$ and the permutation σ^{-1}. The voice leading will be small when $\boldsymbol{F}(a_i) \approx a_{\sigma^{-1}(i)}$, for all i, and will be trivial with strict equality.

Thus, the internal structure of any chord A, as represented by the size of the voice leadings $(a_1, a_2, ..., a_n) \rightarrow (a_{\sigma(1)}, a_{\sigma(2)}, ..., a_{\sigma(n)})$, for all permutations σ, determines the functions $\boldsymbol{F}$ such that A and $\boldsymbol{F}(A)$ can be linked by efficient voice leading. For example, when the notes $\{a_1, a_2, ..., a_n\}$ are all close together, the voice leadings $(a_1, a_2, ..., a_n)\to(a_{\sigma(1)}, a_{\sigma(2)}, ..., a_{\sigma(n)})$ will resemble the trivial voice leading. In this case, A and $\boldsymbol{F}(A)$ can be linked by efficient voice leading only when $\boldsymbol{F}$ resembles the identity. By contrast, when the notes $\{a_1, a_2, ..., a_n\}$ divide the octave nearly evenly, then there will be a voice leading $(a_1, a_2, ..., a_n)\to(a_{\sigma(1)}, a_{\sigma(2)}, ..., a_{\sigma(n)})$ that resembles transposition by $12/n$ semitones. In this case, A and $\mathbf{T}_{12/n}(A)$ can be linked by efficient voice leading.

Observation B. Let $\boldsymbol{G}$ be an isometry of pitch class space, and let $S_{\boldsymbol{G}}$ be invariant under $\boldsymbol{G}$. The size of any voice leading from A to $S_{\boldsymbol{G}}$ sets an upper bound on the size of the minimal voice leading(s) from A to $\boldsymbol{G}(A)$.

Suppose that $\boldsymbol{G}$ is an isometry. It follows that for any voice leading from A to $S_{\boldsymbol{G}}$, with $S_{\boldsymbol{G}}$ invariant under $\boldsymbol{G}$, we can find an equally large voice leading from $S_{\boldsymbol{G}}$ to $\boldsymbol{G}(A)$. If a voice leading from A to $S_{\boldsymbol{G}}$ has displacement multiset $\{d_1, d_2, ..., d_n\}$, then there is some permutation $\boldsymbol{\sigma}$ such that the minimal voice leading from A to $\boldsymbol{G}(A)$ has displacement multiset less than or equal to $\{d_1 + d_{\sigma(1)}, d_2 + d_{\sigma(2)}, ..., d_n + d_{\sigma(n)}\}$. The distribution constraint tells us that no multiset $\{d_1 + d_{\sigma(1)}, d_2 + d_{\sigma(2)}, ..., d_n + d_{\sigma(n)}\}$ can be larger than $\{2d_1, 2d_2, ..., 2d_n\}$. The size of any voice leading from A to $S_{\boldsymbol{G}}$ thus sets an upper bound on the size of the minimal voice leading(s) from A to $\boldsymbol{G}(A)$. Note that the same reasoning can be used when we require that a bijective voice leading from A to A involve the permutation $\boldsymbol{\sigma}$: the size of a voice leading from A to $S_{\boldsymbol{\sigma}}$, with $S_{\boldsymbol{\sigma}}$ invariant under $\boldsymbol{\sigma}$, sets an upper bound on the size of the smallest voice leading of the form $(a_1, a_2, ..., a_n)\to(a_{\sigma(1)}, a_{\sigma(2)}, ..., a_{\sigma(n)})$.

Observation C. Let $\boldsymbol{G}$ be an isometry of pitch class space. The size of a bijective crossing-free voice leading from A to $\boldsymbol{G}(A)$ sets an upper bound on the size of the smallest of the voice leadings from A to any chord $S_{\boldsymbol{G}}$ that is invariant under $\boldsymbol{G}$.

Transposition and inversion are the isometries of pitch class space. When $\boldsymbol{G}$ is an inversion, the reasoning is straightforward. A bijective, crossing-free voice leading from A to $\mathbf{I}_x(A)$ maps pitch class a_i to $x - a_{\sigma(i)}$, where $\boldsymbol{\sigma}$ is a permutation involution. Therefore, it also maps $a_{\sigma(i)}$ to $x - a_i$. Let pitch class b_i be $d/2$ semitones away from both a_i and $x - a_{\sigma(i)}$. Pitch class $x - b_i$ will therefore be $d/2$ semitones away from both $a_{\sigma(i)}$ and $x - a_i$, and we can label it $b_{\sigma(i)}$. It follows that if a bijective, crossing-free voice leading from A to $\mathbf{I}_x(A)$ has displacement multiset $\{d_1, d_2, ..., d_n\}$, then we can find a voice leading from A to $S_{\mathbf{I}_x} = (b_1, b_2, ..., b_n)$ with displacement multiset $\{d_1/2, d_2/2, ..., d_n/2\}$ (*12*).

When $\boldsymbol{G}$ is a transposition, things are more complicated. In this case, we cannot use the size of a bijective crossing-free voice leading from A to $\mathbf{T}_x(A)$ to determine the size of a voice leading from A to $S_{\mathbf{T}_x}$, with $S_{\mathbf{T}_x}$ invariant under $\mathbf{T}_x$. Figure S10 illustrates. The chords {0, 3, 6, 8}, or {C, E♭, G♭, A♭}, and {0, 3, 5, 8}, or {C, E♭, F, A♭}, can be arranged as cycles of pitch classes whose elements are 2, 3, 3, and 4 semitones apart. Consequently, there is a minimal bijective voice leading between each chord and its $\mathbf{T}_3$-form having displacement multiset $\{\|2 - 3\|_{12\mathbb{Z}}, \|3 - 3\|_{12\mathbb{Z}}, \|3 - 3\|_{12\mathbb{Z}}, \|4 - 3\|_{12\mathbb{Z}}\}$, or

{1, 0, 0, 1}. However, one can transform {0, 3, 6, 8} into a $\mathbf{T}_3$-invariant chord by moving only one note, whereas one needs to move two notes to transform {0, 3, 5, 8} into a $\mathbf{T}_3$-invariant chord. Many methods of comparing voice leadings therefore consider {0, 3, 6, 8} to have a smaller minimal bijective voice leading to the nearest $\mathbf{T}_3$-invariant chord than does {0, 3, 5, 8}. But this difference is not reflected in the size of the minimal bijective voice leadings between the two chords and their $\mathbf{T}_3$-forms. Thus, unlike the inversional case considered above, the size of a bijective crossing-free voice leading from A to $\mathbf{T}_x(A)$ does not directly determine the size of a voice leading from A to $S_{\mathbf{T}_x}$, where $S_{\mathbf{T}_x}$ is invariant under $\mathbf{T}_x$.

However, the size of a bijective crossing-free voice leading from A to $\mathbf{T}_x(A)$ does set an upper bound on the size of the smallest of the bijective voice leadings from A to any $S_{\mathbf{T}_x}$. Let A be a chord with n notes, and let α be a bijective voice leading of the form $(a_0, a_1, ..., a_{n-1}) \rightarrow (a_{n-1} + x, a_0 + x, a_1 + x, ..., a_{n-2} + x)$, which combines transposition by x semitones with a permutation consisting of a single cycle. Furthermore, let $nx \equiv 0$ modulo $12\mathbb{Z}$, so that there is an $S_{\mathbf{T}_x}$ that is invariant under $\mathbf{T}_x$. The voice leading α has displacement multiset $\{\|(a_i + x) - a_{i+1 \text{ (mod } n)}\|_{12\mathbb{Z}}\}$. Write $d_i = -(a_i + x - a_{i+1 \text{ (mod } n)})$, so that the displacement multiset is $\{\|d_i\|_{12\mathbb{Z}}\}$, and $a_{i+1 \text{ (mod } n)} - a_i = d_i + x$. Figure S10($c$) shows that there is a voice leading β from A to some $S_{\mathbf{T}_x}$ having displacement multiset

$$\{0, \|d_0\|_{12\mathbb{Z}}, \|d_{n-1}\|_{12\mathbb{Z}}, \|d_0 + d_1\|_{12\mathbb{Z}}, \|d_{n-1} + d_{n-2}\|_{12\mathbb{Z}}, \|d_0 + d_1 + d_2\|_{12\mathbb{Z}}, ...\}.$$

This multiset has one zero member, two members of the form $\|d_i\|_{12\mathbb{Z}}$, two members of the form $\|d_i + d_j\|_{12\mathbb{Z}}$, two members of the form $\|d_i + d_j + d_k\|_{12\mathbb{Z}}$, and so on. By the distribution constraint, this multiset will be smaller than

$$\{0, \|d_0\|_{12\mathbb{Z}}, \|d_{n-1}\|_{12\mathbb{Z}}, \|d_0\|_{12\mathbb{Z}} + \|d_1\|_{12\mathbb{Z}}, \|d_{n-1}\|_{12\mathbb{Z}} + \|d_{n-2}\|_{12\mathbb{Z}}, \|d_0\|_{12\mathbb{Z}} + \|d_1\|_{12\mathbb{Z}} + \|d_2\|_{12\mathbb{Z}}, ...\},$$

where we have used the fact that $\|a + b\|_{12\mathbb{Z}} \leq \|a\|_{12\mathbb{Z}} + \|b\|_{12\mathbb{Z}}$. Now arrange the quantities $\|d_i\|_{12\mathbb{Z}}$ in a sequence from largest to smallest, labeling the elements of this sequence $c_1, c_2, ..., c_n$ such that $c_i \geq c_{i+1}$. By the distribution constraint,

$$\{0, \|d_0\|_{12\mathbb{Z}}, \|d_{n-1}\|_{12\mathbb{Z}}, \|d_0\|_{12\mathbb{Z}} + \|d_1\|_{12\mathbb{Z}}, \|d_{n-1}\|_{12\mathbb{Z}} + \|d_{n-2}\|_{12\mathbb{Z}}, \|d_0\|_{12\mathbb{Z}} + \|d_1\|_{12\mathbb{Z}} + \|d_2\|_{12\mathbb{Z}}, ...\}$$
$$\leq \{0, c_1, c_2, c_1 + c_2, c_3 + c_4, c_3 + c_4 + c_5, ...\}.$$

Return to the voice leading α: $A \rightarrow \mathbf{T}_x(A)$, which has the form $(a_0, a_1, ..., a_{n-1}) \rightarrow (a_{n-1} + x, a_0 + x, a_1 + x, ..., a_{n-2} + x)$. Label the members of its displacement multiset $c_1, c_2, ..., c_n$ such that $c_i \geq c_{i+1}$. The preceding paragraph shows that we can find a voice leading β: $A \rightarrow S_{\mathbf{T}_x}$ with a displacement multiset that is smaller than the first n elements of the sequence $(0, c_2, c_3, c_2 + c_3, c_4 + c_5, c_2 + c_3 + c_4, c_5 + c_6 + c_7, ...)$. (The circular nature of pitch class space allows us to choose a_0 so as to avoid any terms involving c_1.) We conclude that the size of our voice leading from A to $\mathbf{T}_x(A)$ sets a limit on the size of the smallest of the voice leadings from A to any $S_{\mathbf{T}_x}$.

This argument has considered only simple bijective voice leadings whose permutational component consists of a single cycle. However, the reasoning generalizes to bijective voice leadings whose permutational component consists of multiple such cycles, and hence to all bijective voice leadings. (Indeed, the limit we have derived also applies to these voice leadings.) We can also use the argument to show that the size of a bijective voice leading $(a_1, a_2, ..., a_n) \rightarrow (a_{\sigma(1)}, a_{\sigma(2)}, ..., a_{\sigma(n)})$ limits the size of the smallest of the voice leadings from A to any chord S_σ that is invariant under $\boldsymbol{\sigma}$.

7. Maximal evenness and minimal voice leading. I now prove that no bijective voice leading from A to $\mathbf{T}_x(A)$ can be smaller than the minimal bijective voice leading from E to $\mathbf{T}_x(E)$, where E has the same cardinality as A and divides the octave evenly. Thus the perfectly even chord lying at the center of the voice-leading orbifold has the smallest possible minimal bijective voice leading to all of its transpositions. As one moves away from this chord, and toward the singular "boundary" of the orbifold, the size of the voice leadings between chords and their transpositions will in general tend to increase. However, for a given transposition $\mathbf{T}_x$, the rate of increase of these voice leadings depends on the direction in which one is moving, since the size of the minimal voice leading from A to $\mathbf{T}_x(A)$ depends not just on the distance from chord A to the nearest $\mathbf{T}_x$-invariant chord, but also on the internal structure of the chord itself (§6).

> **THEOREM 2.** Let A be any multiset of cardinality n, and let our method of comparing voice leadings be a total preorder satisfying the distribution constraint. For all x, the minimal bijective voice leading between A and $\mathbf{T}_x(A)$ can be no smaller than the minimal bijective voice leading between E and $\mathbf{T}_x(E)$, where E divides pitch-class space into n equal parts.

Since E is invariant under transposition by $12/n$ semitones, the crossing-free voice leadings between E and $\mathbf{T}_x(E)$ will have the form $(e_1, e_2, ..., e_n) \rightarrow (e_1 + c, e_2 + c, ... e_n + c)$, where $c \equiv x$ modulo $(12/n)\mathbb{Z}$. (**NB**: c is congruent to x modulo $(12/n)\mathbb{Z}$, *not* modulo $12\mathbb{Z}$.) Choose c so that $\|c\|_{12\mathbb{Z}}$ is as small as possible. The displacement multiset will have n members each equal to $\|c\|_{12\mathbb{Z}}$. By the distribution constraint, this multiset is as small as any n-member multiset with the same or greater sum.

Now consider any bijective voice leading between transpositionally related n-note chords $(a_1, a_2, ..., a_n) \rightarrow (a_{\sigma(1)} + x, a_{\sigma(2)} + x, ..., a_{\sigma(n)} + x)$. The sum of the members in the displacement multiset is $\sum\|a_{\sigma(i)} + x - a_i\|_{12\mathbb{Z}}$. Since $\|a\|_{12\mathbb{Z}} + \|b\|_{12\mathbb{Z}} \geq \|a + b\|_{12\mathbb{Z}}$, this sum is greater than or equal to $\|nx + \Sigma(a_{\sigma(i)} - a_i)\|_{12\mathbb{Z}} = \|nx\|_{12\mathbb{Z}}$. By construction, $\|nx\|_{12\mathbb{Z}} = \|nc\|_{12\mathbb{Z}}$, since $x \equiv c$ modulo $(12/n)\mathbb{Z}$, and $\|nc\|_{12\mathbb{Z}} = n\|c\|_{12\mathbb{Z}}$, since we can always choose c such that

$\|c\|_{12\mathbb{Z}} \leq 6/n$. Thus the voice leading can be no smaller than the minimal voice leading between E and $\mathbf{T}_x(E)$. ■

I now prove a corollary that applies to collections of k equally spaced pitch classes. Musically, these pitch classes represent an equal-tempered chromatic scale. Such discrete musical universes will not always contain chords that divide the octave into n perfectly even pieces. For example, no five-note chord in twelve-tone equal temperament divides the octave perfectly evenly. But as Douthett and Clough have shown (*13*), equal-tempered scales will always contain a unique collection of "maximally even" chords dividing the octave as evenly as possible (*14*). These maximally even chords are the discrete analogues to the perfectly even chords we have been considering.

In what follows, it will be convenient to relabel the pitch classes so that the equally-spaced points in our chromatic scale have integer co-ordinates. We will therefore temporarily abandon $\mathbb{R}/12\mathbb{Z}$ in favor of $\mathbb{R}/k\mathbb{Z}$, where k is an integer.

> **COROLLARY.** Let A and M be integer-valued n-note submultisets of $\mathbb{R}/k\mathbb{Z}$, and let M be maximally even. Then, for any integer x, the minimal bijective voice leading between A and $\mathbf{T}_x(A)$ can be no smaller than the minimal bijective voice leading between M and $\mathbf{T}_x(M)$.

The proof is very similar to the proof of Theorem 2. Since M is maximally even (*14*), we can find a (not necessarily integer-valued!) chord $\{e_1, e_2, ..., e_n\}$ that divides the octave into n precisely even parts, such that the crossing-free voice-leadings between M and $\mathbf{T}_x(M)$ can be written

$$(\lfloor e_1 \rfloor, \lfloor e_2 \rfloor, ..., \lfloor e_n \rfloor) \rightarrow (\lfloor e_1 + c \rfloor, \lfloor e_2 + c \rfloor, ... \lfloor e_n + c \rfloor), \text{ where } c \equiv x \text{ modulo } (k/n)\mathbb{Z}$$

Choose c so that $\|c\|_{k\mathbb{Z}}$ is as small as possible. The displacement multiset will have n members chosen from the set $\{\|\lfloor c \rfloor\|_{k\mathbb{Z}}, \ \|\lceil c \rceil\|_{k\mathbb{Z}}\}$, and summing to $n\|c\|_{k\mathbb{Z}}$ (*15*). By the distribution constraint, this multiset is as small as any integer-valued n-note multiset with the same or greater sum. As we saw in Theorem 2, the minimal bijective voice leading between A and $\mathbf{T}_x(A)$ will have a displacement multiset summing to at least $n\|c\|_{k\mathbb{Z}}$. Hence it can be no smaller than the minimal bijective voice leading between M and $\mathbf{T}_x(M)$. ■

The preceding corollary can also be applied to scales that do not divide pitch class space evenly. To see why, note that we can always devise a metric on pitch class space such that a given scale evenly divides the octave. For example, given the C major scale (C, D, E, F, G, A, B) we can define a metric according to which the distances C–D, D–E, E–F, F–G, G–A, A–B, and B–C all equal one [Fig. S11(*a*)]. Musicians use the term "scale step" to refer to this scale-dependent unit of distance. Two subsets of the scale are related by *scalar transposition* if they are related by rotation relative to this new metric

[Fig. S11(*b*)]. Since our scale evenly divides the octave according to the new metric, the preceding corollary applies.

Relative to this metric, the familiar diatonic "tertian triads" {C, E, G}, {D, F, A}, {E, G, B}, {F, A, C}, {G, B, D}, {A, C, E}, and {B, D, F} are maximally even and related by transposition [Fig. S11(*b*)]. It follows from our corollary that the minimal bijective voice leadings between these tertian triads will be as small as possible: a bijective voice leading between any pair of transpositionally related three-note diatonic subsets can be no smaller than the minimal bijective voice leading between diatonic tertian triads related by that transposition. Analogous facts hold for the tertian seventh chord {C, E, G, B} and its transpositions, which are also maximally even. Our corollary therefore generalizes a result of Agmon (*16*), who first noted the special voice-leading properties of diatonic tertian triads and seventh chords.

8. A polynomial-time algorithm for finding a minimal voice leading between two chords. Given two chords A and B, how do we find a minimal voice leading between them? The question is nontrivial, since a minimal voice leading need not be bijective. For example, using any of the standard methods of comparing voice leadings, the voiceleading (0, 0, 4, 6)→(10, 0, 6, 6) is smaller than any of the bijective voice leadings between {0, 4, 6} and {6, 10, 0} (*17*). Adding additional voices therefore allows us to decrease the size of the voice leading. The large number of non-bijective voice leadings between any two chords—roughly 2^{mn}, where m and n are their cardinalities—means that an exhaustive search may be impractical, particularly in time-critical applications such as interactive computer music.

Suppose, however, that our method of comparing voice leadings is a total preorder satisfying both the distribution constraint and what I will call *the recursion constraint:*

$$\{x_1, x_2, ..., x_m\} \geq \{y_1, y_2, ..., y_n\} \text{ implies } \{x_1, x_2, ..., x_m, c\} \geq \{y_1, y_2, ..., y_n, c\}$$

The recursion constraint mandates a straightforward relationship between the size of a multiset and the size of its sub-multisets. Every music-theoretical method of comparing voice leadings satisfies this constraint. When a total preorder satisfies both the distribution and recursion constraints, we can use the technique of "dynamic programming," common in computer science, to determine a minimal voice leading between arbitrary chords in polynomial time (order n^2m).

Define the *ascending distance* from pitch class a to pitch class b as the smallest nonnegative real number x such that, if p is a pitch belonging to pitch class a, then $p + x$ belongs to pitch class b. Let(a_1, a_2, ..., a_m, $a_{m+1} = a_1$) order the notes of chord A based on increasing ascending distance from arbitrarily chosen a_1. (Note that I repeat the first element a_1 as the last element of the list.) Similarly for (b_1, b_2, ..., b_n, $b_{n+1} = b_1$). The notation $[a_1, ..., a_i]\rightarrow[b_1, ... b_j]$ will refer to all voice leadings from $\{a_1, a_2, ..., a_i\}$ to $\{b_1, b_2, ... b_j\}$ that can be written with both chords' subscripts in nondecreasing order. Thus $[a_1, a_2]\rightarrow[b_1, b_2, b_3]$ refers to $(a_1, a_1, a_2)\rightarrow(b_1, b_2, b_3)$, $(a_1, a_1, a_2, a_2)\rightarrow(b_1, b_2, b_3, b_3)$, and so on.

If a crossing-free voice leading contains the pair (a_i, b_j), with $i, j > 1$, then it must contain at least one of the following: (a_{i-1}, b_j), (a_i, b_{j-1}), or (a_{i-1}, b_{j-1}). By the recursion constraint, the smallest voice leading of the form $[a_1, ..., a_i]\rightarrow[b_1, ..., b_j]$ will be the voice leading that adds the pair (a_i, b_j) to the smallest voice leading of the form $[a_1, ..., a_{i-1}]\rightarrow [b_1, ..., b_j]$, $[a_1, ..., a_i]\rightarrow[b_1, ..., b_{j-1}]$, or $[a_1, ..., a_{i-1}]\rightarrow[b_1, ..., b_{j-1}]$. Thus, once we have fixed the pair (a_1, b_1) we can recursively compute the smallest voice leading between A and B containing that pair. We do this by creating a matrix whose entries $e_{i,j}$ record the size of the minimal voice leading of the form $[a_1, ..., a_i]\rightarrow[b_1, ..., b_j]$. It is trivial to fill in the first row and column of the matrix. For any other entry $e_{i,j}$, we simply add the voice (a_i, b_j) to the smallest of the voice leadings in the entry's upper, left, and upper-left neighbors.

Figure S12 illustrates the technique, identifying the smallest voice leading between the C and E major-seventh chords, {4, 7, 11, 0} and {4, 8, 11, 3}, such that the voice leading contains the pair (4,4). In constructing this matrix I have used the L^1 "taxicab" norm to measure voice-leading size. The voice leading in the bottom-right entry, $(4, 4, 7, 11, 0)\rightarrow(3, 4, 8, 11, 11)$, is one of the minimal voice leadings between the two chords that contains (4, 4). To remove this last restriction, we would need to repeat the calculation three more times, each time cyclically permuting the order of one of the chords so as to fix a different initial pair. As it happens, however, the voice leading shown in Figure S12 is minimal. This follows from the fact that the mapping in the top-left position, (4, 4), contributes nothing to the overall size of the voice leading; we can therefore add it to any voice leading without increasing its L^1 size.

Figure S12 includes in each entry both the numerical size of the voice leading and the voice leading itself. With the L^1 norm this is unnecessary: we need to keep track of the size, but not the voice leading. To determine the value of entry $e_{i,j}$ we can simply add the distance between the pair (a_i, b_j) to the minimum value in the entries $e_{i-1,\ j}$, $e_{i,\ j-1}$, and $e_{i-1,\ j-1}$. (For the other L^p norms we can calculate the p^{th} power of the voice-leading size in this way, taking the p^{th} root before output.) Having filled in the matrix, we can recover a minimal voice leading between the two chords by "tracing back" a path that moves from the bottom-right entry to the top left, moving only up, left, and diagonally up-and-left, such that the size of the voice leading decreases as much as possible with each step. The entries in boldface indicate the path such a traceback algorithm would take. Due to the circular structure of pitch-class space, the voice leading in the lower right-hand corner of the matrix counts the pair $(a_1, b_1) = (a_{m+1}, b_{n+1})$ twice; this can easily be corrected prior to output. The resulting algorithm is easy to implement and suitable for time-critical applications such as interactive computer music.

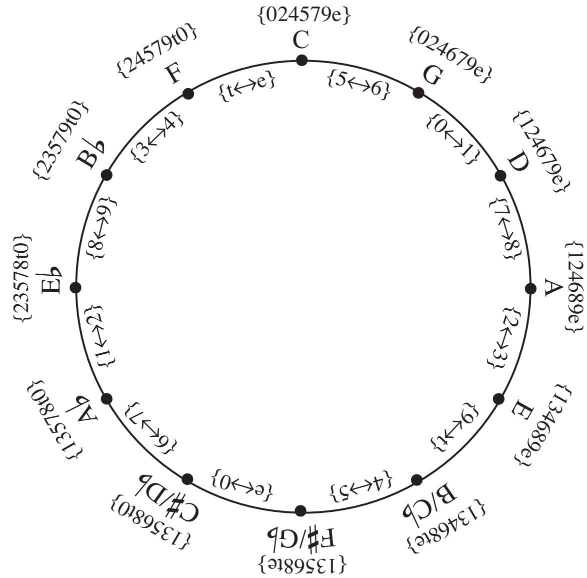


Figure S1. The circle of fifths depicts minimal voice leadings between diatonic collections (major scales). Each diatonic collection can be transformed into its neighbors by moving one pitch class by one semitone. For example, the C major scale can be transformed into the G major scale by moving the pitch class 5 (F) to 6 (F♯). Here and elsewhere, the letters "t" and "e" refer to 10 (B♭) and 11 (B), respectively.

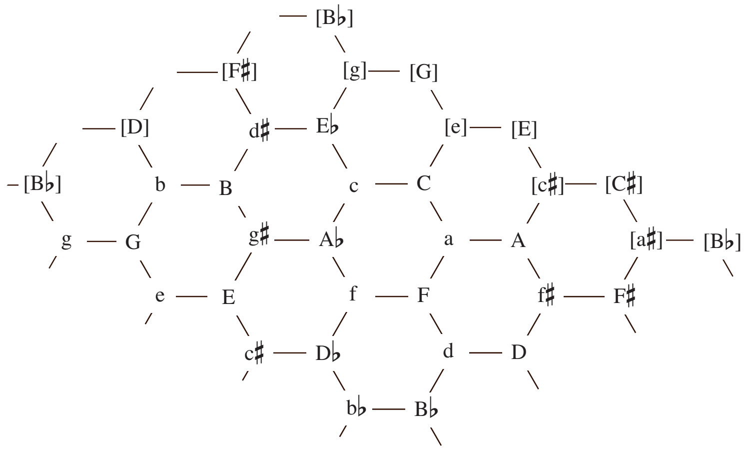


Figure S2. The Tonnetz. Nineteenth-century theorists such as Hostinsky, Oettingen, and Riemann explored a graph that is the geometric dual of the one shown here. The graph displays efficient voice leadings among the 24 familiar major and minor triads. Uppercase and lowercase letters indicate major and minor triads, respectively. Triads connected by horizontal lines share both root and fifth, and can be linked by voice leading in which one note moves by one semitone. (For example, the C major triad can be transformed into a C minor triad by changing E to E♭.) Triads connected by a NE/SW diagonal also share two notes and can be linked by single-semitone voice leading. (For example, the C major triad can be transformed into an E minor triad by changing C to B.) Triads connected by a NW/SE diagonal share two notes and can be linked by voice leading in which one note moves by two semitones. (For example, the C major triad can be transformed into an A minor triad by changing G to A.)

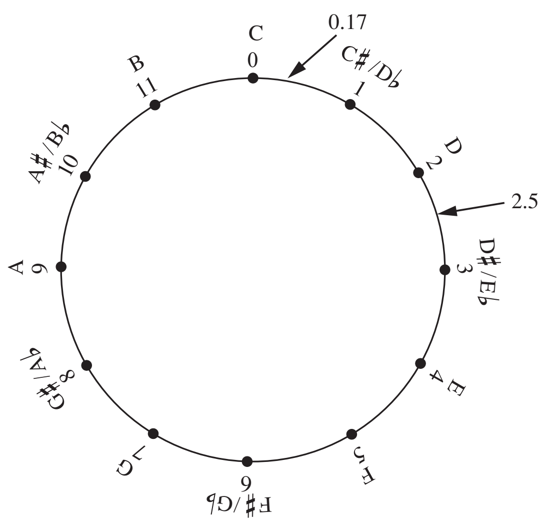


Figure S3. The quotient space $\mathbb{R}/12\mathbb{Z}$ is a circle whose circumference is twelve units long. The twelve familiar pitch classes of Western equal-temperament evenly divide this circle. Since the circle is continuous, it contains a point for every conceivable pitch class. The figure shows the locations of the pitch class 0.17, which is seventeen cents (hundredths of a semitone) above pitch class C, and pitch class 2.5 (D quarter tone sharp), which is halfway between D and E♭.

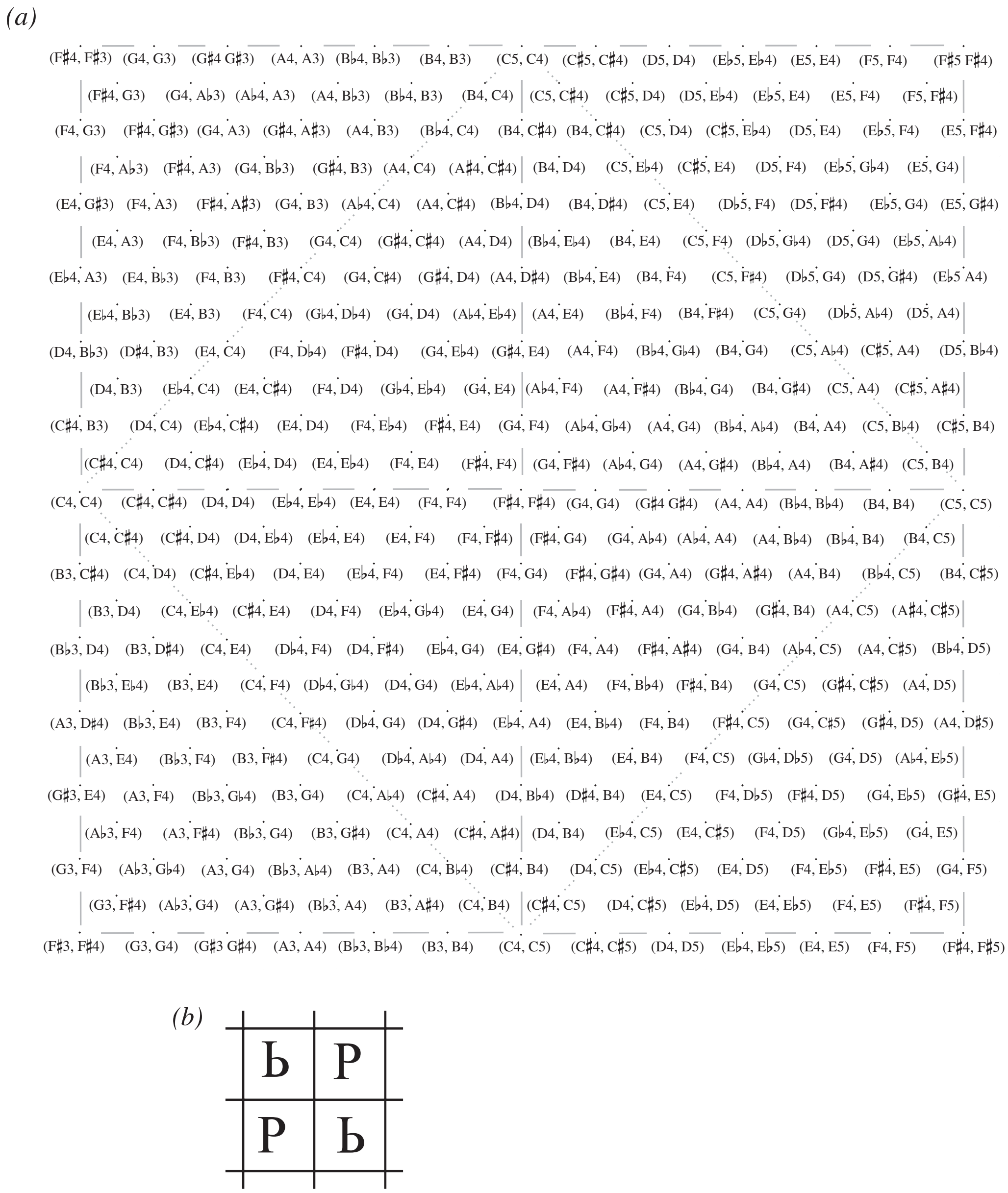


Figure S4. (*a*) A portion of the infinite plane representing ordered two-note chords of pitches. The four quadrants are equivalent to within octave displacement and permutation. Each is equivalent to Figure 2 in the main paper. (*b*) An abstract representation of the symmetries relating the four quadrants. The lower-left quadrant is related to the upper-left by a reflection that preserves their common border. This action permutes each dyad. Translation of any quadrant diagonally up and right transposes the first element of each dyad by an ascending octave. Translation of any quadrant diagonally down and right transposes the second element of each dyad by an ascending octave. These operations suffice to generate the infinite, periodic figure. The space can be described, metaphorically, as wallpaper; Figure 2 in the main paper provides the pattern, Figure S4(*b*) shows how the pattern is to be assembled, and Figure S4(*a*) shows a portion of the result. The diamond in the center of Fig. S4(*a*) corresponds to the 2-torus shown in Figure S9.

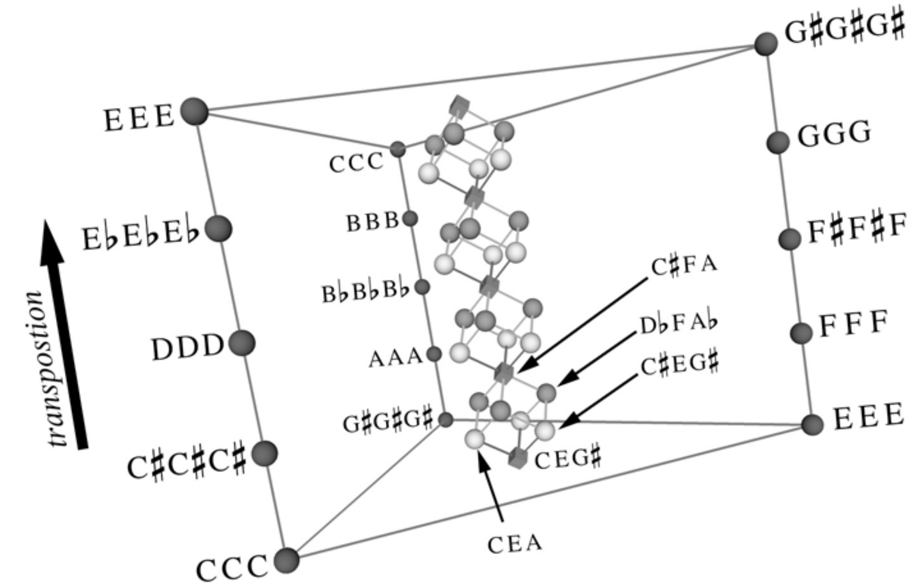


Figure S5. The orbifold $\mathbb{T}^3/\mathcal{S}_3$ is a prism whose two triangular faces are identified by way of a 120° rotation. Several familiar equal-tempered chords are depicted on the figure. Augmented triads, which divide the octave into three equal parts, are shown as dark cubes. Minor chords are light spheres and major chords are dark spheres. Lines connecting augmented, minor, and major chords indicate that they can be linked by voice leading in which a single voice moves by a single semitone. Since minor and major chords divide the octave nearly evenly, they are clustered near the center of the orbifold. Triple unisons, which contain only one pitch class, are found on the edge of the figure.

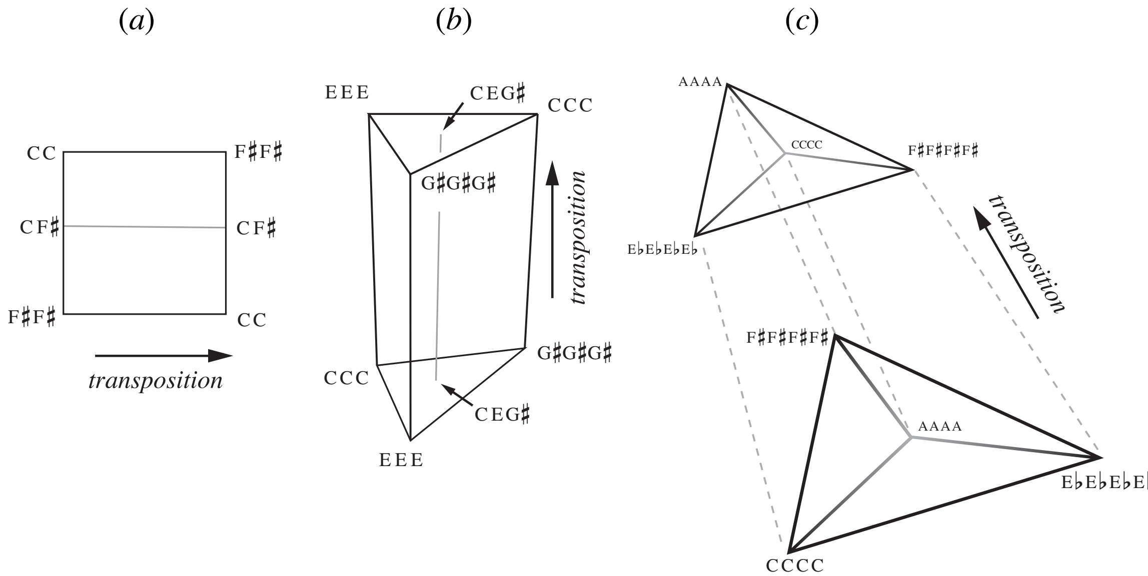


Figure S6. (*a*) The orbifold $\mathbb{T}^2/\mathcal{S}_2$ is a two-dimensional prism (parallelogram) whose base (a line segment) is glued to the opposite face. Before gluing, the base must be twisted so that chords on the left edge match those on the right. This twist is a reflection that can be represented as a rotation in three Euclidean dimensions. The line at the center of the figure contains chords that divide the octave evenly. (*b*) The orbifold $\mathbb{T}^3/\mathcal{S}_3$ is a three-dimensional prism whose two triangular faces are glued together. Before gluing, rotate one face by 120°, so that the chords match. The result is the bounded interior of a twisted triangular 2-torus. Augmented triads, which divide the octave into three equal parts, lie on the line at the center of the figure. Major and minor chords are close to this line, as shown in Figure S5. Rotating the prism around the central line by 120° transposes every chord by major third. (*c*) The orbifold $\mathbb{T}^4/\mathcal{S}_4$ is a four-dimensional prism whose two tetrahedral faces are glued together. The dashed lines extend into the fourth dimension. Before identifying the two faces, twist one so that the chords match. The twist is a reflection, as in the two-dimensional case. Diminished seventh chords, which divide the octave into four equal pieces, lie at the center of the orbifold. Familiar four-note tonal chords lie close to this chord.

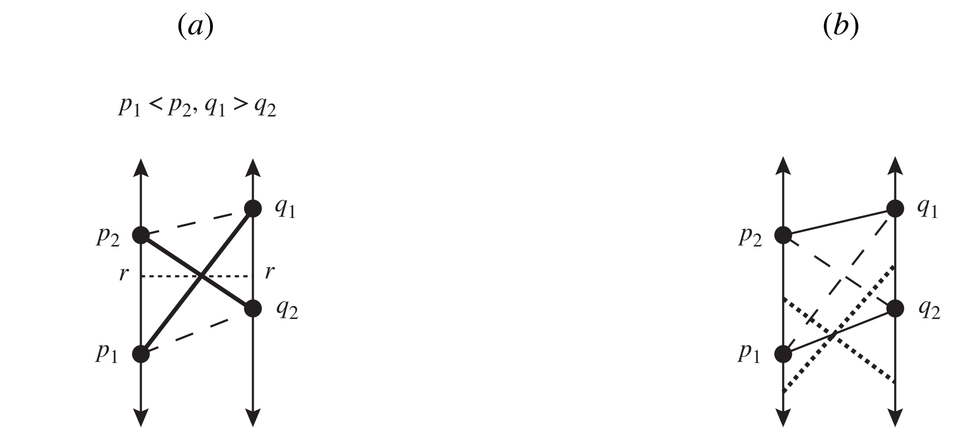


Figure S7. (a) $|r - p_2|/|r - p_1| = |q_2 - r|/|q_1 - r|$. ($b$) Any line segment that crosses (p_1, q_2) crosses either (p_1, q_1) or (p_2, q_2).

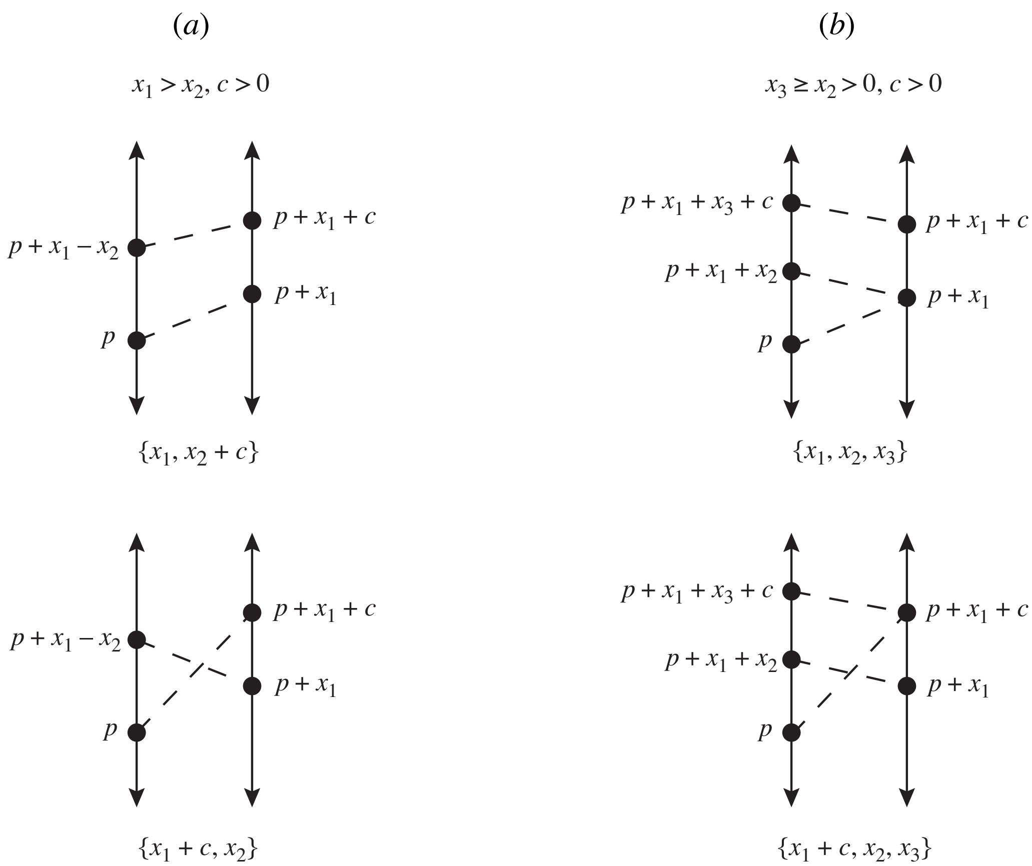


Figure S8. For any violation of the distribution constraint, we can find a crossed voice leading that is smaller than its natural uncrossed alternative. The crossed voice leading at the bottom of each column is smaller than the uncrossed voice leading above it.

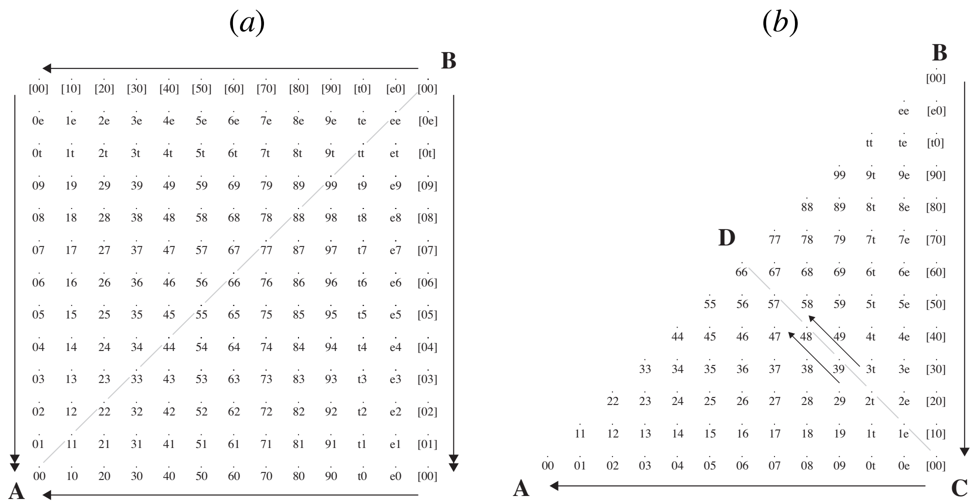
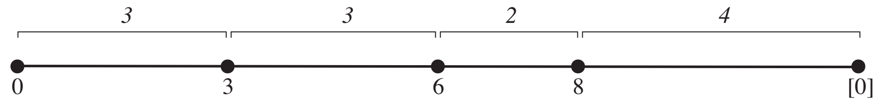
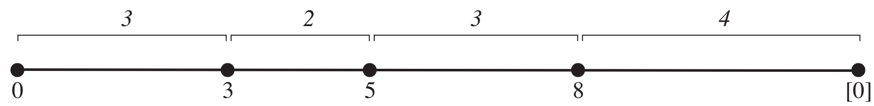


Figure S9. (*a*) The space of ordered two-note chords of pitch-classes is a 2-torus. To identify points (x, y) and (y, x), we need to fold the torus along the AB diagonal. The resulting figure, shown in (*b*), is a triangle with two of its sides identified. This is a Möbius strip. To see why, cut figure (*b*) along the line CD and glue AC to CB. (To make this identification in Euclidean 3-space, you will need to turn over one of the pieces of paper.) The result is a square with opposite sides identified, as in Figure 2 of the main paper.

(*a*) *the dominant-seventh chord {0, 3, 6, 8}*



(*b*) *the minor-seventh chord {0, 3, 5, 8}*



(*c*)

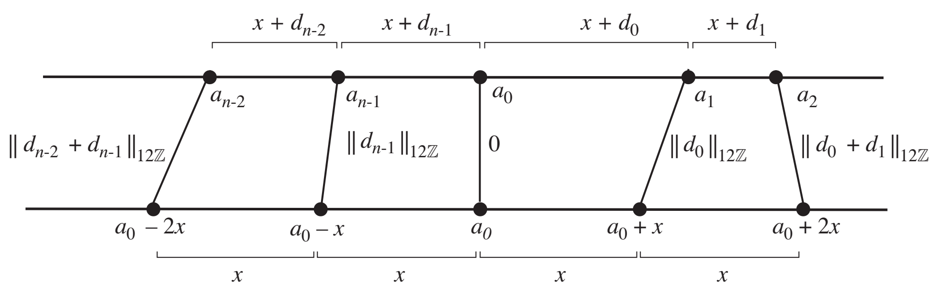


Figure S10. (*a–b*) The dominant-seventh chord {0, 3, 6, 8} and minor-seventh chord {0, 3, 5, 8} can both be arranged as cycles whose notes are 2, 3, 3, and 4 semitones apart. To transform the dominant-seventh chord into a chord that divides the octave evenly, one need only move pitch class 8 to pitch class 9. However, to transform the minor-seventh chord into a chord that divides the octave evenly, one must move at least two pitch classes. (For example, one can move 5 and 8 to 6 and 9, respectively.) Consequently, although both chords have equally small minimal voice leadings to their $\mathbf{T}_3$ forms, the first is closer to the nearest $\mathbf{T}_3$-invariant chord according to many metrics. (*c*) To find a bijective voice leading from any chord to a chord whose intervals are all equal to x, fix a_0, move note a_1 by d_0 semitones, note a_{n-1} by d_{n-1} semitones, note a_2 by $d_0 + d_1$ semitones, and so on. Since pitch-class space is circular, the term $d_{\lfloor n/2 \rfloor}$ will not be involved in this voice leading.

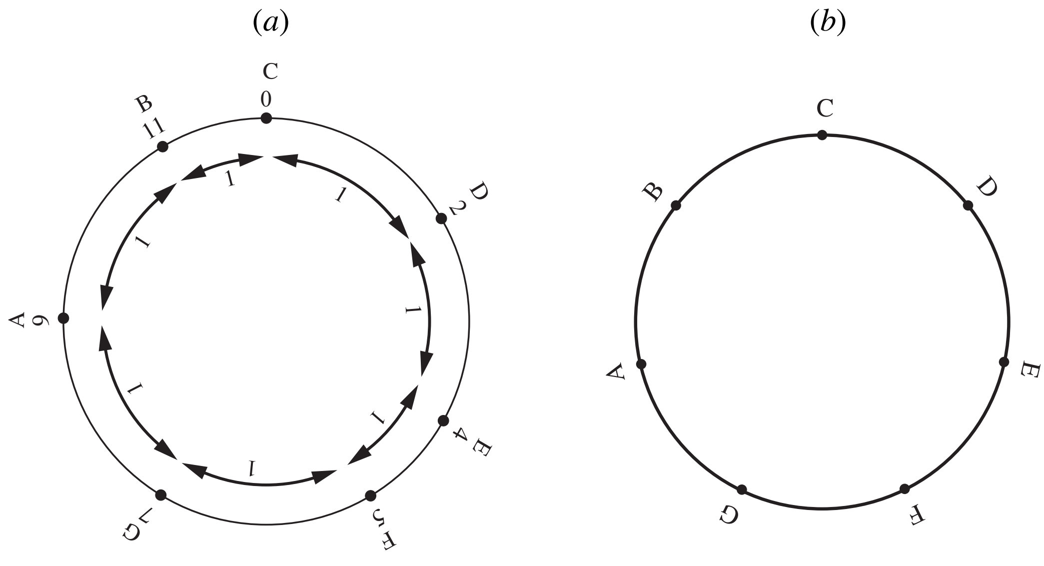


Figure S11. For any scale, we can define a metric such that the scale's notes divide the octave evenly. (*a*) shows the C major scale, as it appears in circular pitch class space. In (*b*), we apply a new metric, so that the scale's notes are equally spaced. The new unit of distance is called a "scale step." Relative to this new metric, the C major triad {C, E, G} and the D minor triad {D, F, A} are related by rotation, which musicians call "scalar transposition."

	4	**8**	**11**	**3**	**4**
4	**(4)→(4)** ***Size: 0***	(4, 4)→(4, 8) *Size: 4*	(4, 4, 4)→ (4, 8, 11) *Size: 9*	(4, 4, 4, 4)→ (4, 8, 11, 3) *Size: 10*	(4, 4, 4, 4, 4)→ (4, 8, 11, 3, 4) *Size: 10*
7	(4, 7)→(4, 4) *Size: 3*	**(4, 7)→(4, 8)** ***Size: 1***	(4, 7, 7)→ (4, 8, 11) *Size: 5*	(4, 7, 7, 7)→ (4, 8, 11, 3) *Size: 9*	(4, 7, 7, 7, 7)→ (4, 8, 11, 3, 4) *Size: 12*
11	(4, 7, 11)→ (4, 4, 4) *Size: 8*	(4, 7, 11)→ (4, 8, 8) *Size: 4*	**(4, 7, 11)→** **(4, 8, 11)** ***Size: 1***	(4, 7, 11, 11)→ (4, 8, 11, 3) *Size: 5*	(4, 7, 11, 11, 11)→ (4, 8, 11, 3, 4) *Size: 10*
0	(4, 7, 11, 0)→ (4, 4, 4, 4) *Size: 12*	(4, 7, 11, 0)→ (4, 8, 8, 8) *Size: 8*	**(4, 7, 11, 0)→** **(4, 8, 11, 11)** ***Size: 2***	(4, 7, 11, 0)→ (4, 8, 11, 3) *Size: 4*	(4, 7, 11, 0, 0)→ (4, 8, 11, 3, 4) *Size: 8*
4	(4, 7, 11, 0, 4)→ (4, 4, 4, 4, 4) *Size: 12*	(4, 7, 11, 0, 4)→ (4, 8, 8, 8, 8) *Size: 12*	(4, 7, 11, 0, 4)→ (4, 8, 11, 11, 11) *Size: 7*	**(4, 7, 11, 0, 4)→** **(4, 8, 11, 11, 3)** ***Size: 3***	**(4, 7, 11, 0, 4, 4)→** **(4, 8, 11, 11, 3, 4)** ***Size: 3***

Figure S12. Using dynamic programming to find a minimal voice leading between {4, 7, 0, 11} and {4, 8, 11, 3}.

SYMBOL OR TERM	DEFINITION
multiset, object, member, element	A multiset is an unordered collection in which duplications are permitted. {0, 1, 1} is the same multiset as {1, 0, 1} but is different from {0, 1}. The multiset {0, 1, 1} contains three objects and has three members. However, it has only two elements, 0 and 1.
$\{a, b, c\}$	The multiset with members a, b, c, some of which may be identical.
(a, b, c)	An ordered list. (a, b, c) and (b, c, a) are distinct.
$i \in \boldsymbol{I}$	Object i belongs to set or multiset $\boldsymbol{I}$.
group	A group is a set whose elements can be combined so as to satisfy certain axioms. See a group theory textbook for details.
$\mathbb{R}$	The real numbers.
$\mathbb{R}^n$	The set of ordered n-tuples $(x_1, x_2, \ldots x_n)$ such that each $x_i \in \mathbb{R}$.
$\mathbb{Z}$	The integers.
$n\mathbb{Z}$, where $n \in \mathbb{R}$	The set $\{nk \mid k \in \mathbb{Z}\}$. Thus $12\mathbb{Z}$ is the set $\{\ldots, -24, -12, 0, 12, 24, \ldots\}$. The elements of this set form a group under addition.
$m\mathbb{Z}^n$, where $m \in \mathbb{R}$ and $n \in \mathbb{Z}$	The set of ordered n-tuples $(x_1, x_2, \ldots x_n)$ such that each $x_i \in m\mathbb{Z}$. These n-tuples form a group under vector addition.
$\mathcal{S}_n$	The symmetric group of degree n, consisting of the group of permutations of n objects.
quotient space	A quotient space is formed by identifying (or "gluing together") points in another space.
$\boldsymbol{A}/\mathcal{G}$, where $\mathcal{G}$ is some group of transformations acting on $\boldsymbol{A}$	The quotient space that identifies all points a and ga, where $a \in \boldsymbol{A}$ and $g \in \mathcal{G}$.
$\mathbb{R}/12\mathbb{Z}$	The circular quotient space in which all real numbers x and $x + 12$ are identified. Points in this space are infinite sets of the form $\{\ldots, x - 24, x - 12, x, x + 12, x + 24, \ldots\}$, where $x \in \mathbb{R}$. These sets can be labeled using real numbers in the range $0 \leq x < 12$. The elements of $\mathbb{R}/12\mathbb{Z}$ form a group under addition of their labels modulo $12\mathbb{Z}$.
$\mathbb{T}^n$	The n-torus, or the Cartesian product of n circles. Since $\mathbb{R}/12\mathbb{Z}$ is a circle, $(\mathbb{R}/12\mathbb{Z})^n$ is an n-torus.
topological equivalence	Two spaces are topologically equivalent if one can be continuously deformed into the other.
manifold	A space that is locally topologically equivalent to $\mathbb{R}^n$.
global-quotient orbifold	A global-quotient orbifold (or orbit manifold) is a quotient space $\boldsymbol{M}/\mathcal{G}$, where $\boldsymbol{M}$ is a manifold and $\mathcal{G}$ is a group acting discontinuously on $\boldsymbol{M}$. Global-quotient orbifolds inherit many geometrical properties from their parent spaces.

Table S1, part 1. A glossary of mathematical terms and symbols.

singularity, singular locus	A point in an orbifold at which the local geometry or topology is not that of the corresponding point in the parent space. If $\boldsymbol{M}/\mathcal{G}$ is a global-quotient orbifold, then its singularities are fixed points of the group $\mathcal{G}$. The singularities of the orbifolds $\mathbb{T}^n/\mathcal{S}_n$ act like mirrors.
boundary (of the orbifolds $\mathbb{T}^n/\mathcal{S}_n$)	The singularities of the orbifolds $\mathbb{T}^n/\mathcal{S}_n$, which enclose the nonsingular points (see Figs. 2 and S6).
simplex, simplicial	A simplex is a generalized triangle. The 1-simplex is a line segment, the 2-simplex is a triangle, the 3-simplex is a tetrahedron, and so forth.
prism, height coordinate of the prism, base	In this paper, the term "prism" refers to the n-dimensional region formed when an $(n–1)$-dimensional polyhedron (the base) is dragged along a line segment extending into the remaining dimension. The direction of the dragging is the "height coordinate" of the prism.
$\|\|b - a\|\|_{12\mathbb{Z}}$	The distance between two points of $\mathbb{R}/12\mathbb{Z}$, equal to the smallest nonnegative real number x such that, if p belongs to a, then either $p + x$ or $p - x$ belongs to the set b.
$a \equiv b$ modulo $n\mathbb{Z}$	The real numbers a and b are congruent modulo $n\mathbb{Z}$: that is, there exists an integer k such that $a = b + kn$.
$a \approx b$	a is approximately equal to b.
iff	If and only if.
total preorder reflexive relation transitive relation total relation	A total preorder is a relation "≥" over set A that is reflexive ($a \geq a$, for all $a \in A$), transitive ($a \geq b$ and $b \geq c$ imply $a \geq c$, for all a, b, $c \in A$) and total (either $a \geq b$ or $b \geq a$, or both, for all a, $b \in A$). $a < b$ means "$a \leq b$ but not $b \leq a$." $a \equiv b$ means "$a \leq b$ and $b \leq a$." Note that $a \equiv b$ does not imply that a and b are identical.
fundamental domain	A fundamental domain of the group $\mathcal{G}$ acting on the space S is a region R that tiles S under the action of the group: S is the union of the regions gR, for all $g \in \mathcal{G}$, and any two regions gR and hR, for $g \neq h$, intersect only at their boundaries.
$\rtimes$ semidirect product	Let $\mathcal{G}$ be a group with subgroup $\mathcal{F}$ and normal subgroup $\mathcal{N}$. $\mathcal{G}$ is the semidirect product $\mathcal{N} \rtimes \mathcal{F}$ iff every element in $\mathcal{G}$ can be written in one and only one way as the product fn, with $f \in \mathcal{F}$ and $n \in \mathcal{N}$. See a group theory textbook for details.
involution	A function $\boldsymbol{F}$ is an involution iff $\boldsymbol{F}(\boldsymbol{F}(a)) = a$.
isometry	A function $\boldsymbol{F}$ is an isometry iff the distance between $\boldsymbol{F}(a)$ and $\boldsymbol{F}(b)$ is equal to the distance between a and b, for all a and b.
$\lfloor x \rfloor$ and $\lceil x \rceil$	The greatest integer $\leq x$ and the smallest integer $\geq x$, respectively.

Table S1, part 2. A glossary of mathematical terms and symbols.

pitch	Pitch is the perceptual correlate of fundamental frequency. Pitches can be modeled as real numbers such that middle C is 60, the octave has size 12, and semitones have size 1.
pitch class	A set consisting of all pitches separated by an integral number of octaves. A220 and A440 both belong to the pitch class A. Pitch classes can be modeled as elements of the quotient space $\mathbb{R}/12\mathbb{Z}$.
chord	A multiset of either pitches or pitch classes.
transposition, translation	In both pitch and pitch-class space, transposition (or translation) corresponds to addition by a constant value. If a is a pitch or pitch class then $a + x$ is the transposition of a by x semitones.
$\mathbf{T}_x(A)$	The transposition of the chord A by x semitones.
transpositionally related	Two chords are transpositionally related if one is the transposition of the other.
inversion, reflection	In both pitch and pitch-class space, inversion (reflection) corresponds to subtraction from a constant value. If a is a pitch or pitch class then $x - a$ is the inversion of a with "index number" x.
$\mathbf{I}_x(A)$	The inversion that maps a to $x - a$. $\mathbf{I}_x$ maps 0 to x and vice versa.
inversionally related	Two chords are inversionally related if one is the inversion of the other.
voice leading	A voice leading between two multisets $\{a_1, a_2, ..., a_m\}$ and $\{b_1, b_2, ..., b_n\}$ is a multiset of ordered pairs (a_i, b_j), such that every member of each multiset is in some pair.
$(a_1, a_2, ..., a_n)\rightarrow(b_1, b_2, ..., b_n)$	The voice leading that contains all and only the pairs (a_i, b_i). Voice leadings do not uniquely determine the chords they connect. For example, the voice leading $(0, 0, 4, 7)\rightarrow(11, 2, 5, 7)$ is both a bijective voice leading from $\{0, 0, 4, 7\}$ to $\{11, 2, 5, 7\}$ and a non-bijective voice leading from $\{0, 4, 7\}$ to $\{11, 2, 5, 7\}$.
distance	The distance between two pitches p and q is the absolute value of their difference, $\|q - p\|$. The distance between two pitch classes a and b, written $\|\|b - a\|\|_{12\mathbb{Z}}$, is the smallest nonnegative real number x such that, if p is a pitch belonging to pitch class a, then either $p + x$ or $p - x$ belongs to pitch class b.
ascending distance	The ascending distance from pitch class a to pitch class b is the smallest nonnegative real number x such that, if p is a pitch belonging to pitch class a, then $p + x$ belongs to pitch class b.

Table S1, part 3. A glossary of musical terms and symbols.

associated voice leading	$(p_1, p_2, ..., p_n)\rightarrow(q_1, q_2, ..., q_n)$ is associated with $(a_1, a_2, ..., a_n)\rightarrow(b_1, b_2, ..., b_n)$ if pitches p_i and q_i all belong to the pitch classes a_i and b_i, respectively.
displacement multiset	The multiset of distances between notes in the source chord and their images in the target chord.
distribution constraint	$\{x_1 + c, x_2, ..., x_n\} \geq \{x_1, x_2 + c, ..., x_n\} \geq \{x_1, x_2, ..., x_n\}$ for $x_1 > x_2$ and $c > 0$
recursion constraint	$\{x_1, x_2, ..., x_m\} \geq \{y_1, y_2, ..., y_n\}$ implies $\{x_1, x_2, ..., x_m, c\} \geq \{y_1, y_2, ..., y_n, c\}$
trivial voice leading	A voice leading containing only pairs of the form (x, x).
bijective voice leading	A voice leading from A to B such that every member of A is mapped to exactly one member of B, and vice versa.
independent voice leading	A voice leading that cannot be written in the form $(a_1, a_2, ..., a_n)\rightarrow(a_1 + x, a_2 + x, ..., a_n + x)$.
parallel voice leading	A voice leading that is not independent, and hence acts as a transposition.
crossing free, strongly crossing free	Intuitively, a voice leading is crossing free if the notes of the source chord can be connected to those of the target along minimal-length line-segments intersecting only at their endpoints. A voice leading between multisets of pitches $(p_1, p_2, ..., p_n)\rightarrow(q_1, q_2, ..., q_n)$ has no voice crossings if $p_i > p_j$ implies $q_i \geq q_j$, for all $i, j \leq n$. It is *strongly crossing free* if $(p_1 + 12k_1, p_2 + 12k_2, ..., p_n + 12k_n)\rightarrow(q_1 + 12k_1, q_2 + 12k_2, ..., q_n + 12k_n)$ is crossing free, for all integers k_i. Thus one cannot introduce crossings into a strongly crossing-free voice leading simply by shifting the octave in which its voices appear. A voice leading between multisets of pitch classes is crossing free if it is associated with a strongly crossing-free voice leading between multisets of pitches in which no voice moves by more than six semitones.
voice crossing	Two voices (p_1, q_1) and (p_2, q_2) cross if the voice leading $(p_1, p_2)\rightarrow(q_1, q_2)$ is not crossing free.
maximally even	Let $\{e_0, e_1, ..., e_{n-1}\}$ be a subset of $\mathbb{R}/k\mathbb{Z}$ that divides the octave into n precisely even parts. $\{m_0, m_1, ..., m_{n-1}\}$ is "maximally even" if $m_i = \lfloor e_i \rfloor$. For any maximally even chord $\{m_0, m_1, ..., m_{n-1}\}$, and any integer d, the set $\{m_{i + d \text{ (mod } n)} - m_i\}$ will contain either a single integer-valued point in $\mathbb{R}/k\mathbb{Z}$ or a pair of adjacent integer-valued points in $\mathbb{R}/k\mathbb{Z}$. The n-note maximally even subsets of a chromatic scale are related by transposition.

Table S1, part 4. A glossary of musical terms and symbols.

REFERENCES

1. R. Gauldin, *Harmonic Practice in Tonal Music* (Norton, New York, 2004).
2. S. Laitz, *The Complete Musician* (Oxford, New York, 2003).
3. D. Huron, *Music Perception* **19**, 1 (2001).
4. It is possible to show that the triangle inequality implies the distribution constraint, given a few minimal assumptions. Let P, Q, and R be points in $\mathbb{R}^n$, representing ordered series of pitches, and let $P \to Q$ be the voice leading that maps the i^{th} component of P to the i^{th} component of Q. Suppose our method of comparing voice leadings is a metric that assigns a (real-valued) size to every displacement multiset. Let $\boldsymbol{D}(P \to R)$ refer to the size of the displacement multiset associated with the voice leading $P \to R$. Suppose that $\boldsymbol{D}(P \to R) \leq \boldsymbol{D}(P \to Q) + \boldsymbol{D}(Q \to R)$, for all P, Q, and R (the triangle inequality), and that $\boldsymbol{D}(P \to R) = \boldsymbol{D}(P \to Q) + \boldsymbol{D}(Q \to R)$ whenever Q lies on the line segment $P \to R$. It follows that the distribution constraint's first inequality will be satisfied.
5. D. Lewin, *Journal of Music Theory* **42**, 15 (1998).
6. R. Cohn, *Journal of Music Theory* **42**, 283 (1998).
7. J. Straus, *Music Theory Spectrum* **25**, 305 (2003).
8. C. Callender, *Music Theory Online* **10** (2004).
9. R. Cohn, *Journal of Music Theory* **41**, 1 (1997).
10. J. Douthett, P. Steinbach, *Journal of Music Theory* **42**, 241 (1998).
11. http://music.princeton.edu/~dmitri/ChordGeometries.html.
12. Consequently, for all of the standard voice-leading metrics, we can always find minimal bijective voice leadings from A to $\mathbf{I}_x(A)$ and A to $S_{\mathbf{I}_x}$ with displacement multisets equal to $\{d_1, d_2, ..., d_n\}$ and $\{d_1/2, d_2/2, ..., d_n/2\}$.
13. J. Clough, J. Douthett, *Journal of Music Theory* **35**, 93 (1991).
14. Let $\{e_0, e_1, ..., e_{n-1}\}$ be a subset of $\mathbb{R}/k\mathbb{Z}$ that divides the octave into n precisely even parts. $\{m_0, m_1, ..., m_{n-1}\}$ is "maximally even" if $m_i = \lfloor e_i \rfloor$. For any maximally even chord $\{m_0, m_1, ..., m_{n-1}\}$, and any integer d, the set $\{m_{i+d \pmod n} - m_i\}$ will contain either a single integer-valued point in $\mathbb{R}/k\mathbb{Z}$ or a pair of adjacent integer-valued points in $\mathbb{R}/k\mathbb{Z}$. The n-note maximally even subsets of a chromatic scale are related by transposition. In twelve-tone equal temperament, the maximally even chords are {0, 6}, {0, 4, 8}, {0, 3, 6, 9}, {0, 2, 4, 5, 7}, {0, 2, 4, 6, 8, 10}, their transpositions, and the complements of these chords.
15. There is an integer d such that $e_{i+d \pmod n} + x = e_i + c$. So $n\|c\|_{k\mathbb{Z}} = \Sigma\|e_{i+d \pmod n} + x - e_i\|_{k\mathbb{Z}} = \Sigma\|\lfloor e_{i+d \pmod n} + x \rfloor - \lfloor e_i \rfloor\|_{k\mathbb{Z}} = \Sigma\|\lfloor e_i + c \rfloor - \lfloor e_i \rfloor\|_{k\mathbb{Z}}$, where we use the fact that $\|c\|_{k\mathbb{Z}} \leq k/2n$.
16. E. Agmon, *Musikometrica* **3**, 15 (1991).
17. Note that voice leadings do not uniquely determine the chords they connect: the voice leading $(0, 0, 4, 6) \to (10, 0, 6, 6)$ is both a non-bijective voice leading between {0, 4, 6} and {6, 10, 0} and a bijective voice leading between {0, 0, 4, 6} and {6, 6, 10, 0}.