Cross-Type Transformations and the Path Consistency Condition

JULIAN HOOK

Two kinds of extensions to David Lewin's transformation theory are proposed. First, *cross-type transformations* transform one sort of object to another: for instance, mappings from triads to seventh chords. Special cases include *GIS homomorphisms* and *interscalar transpositions and inversions*. The second extension relaxes Lewin's requirement that the products of transformations along two paths from one node to another in a transformation graph must match. It is argued that this *path consistency* condition is unnecessarily restrictive; several alternatives are considered. Examples from Bach to Bartók illustrate musical relationships that can be most effectively modeled through the use of cross-type transformations and non-path-consistent graphs.

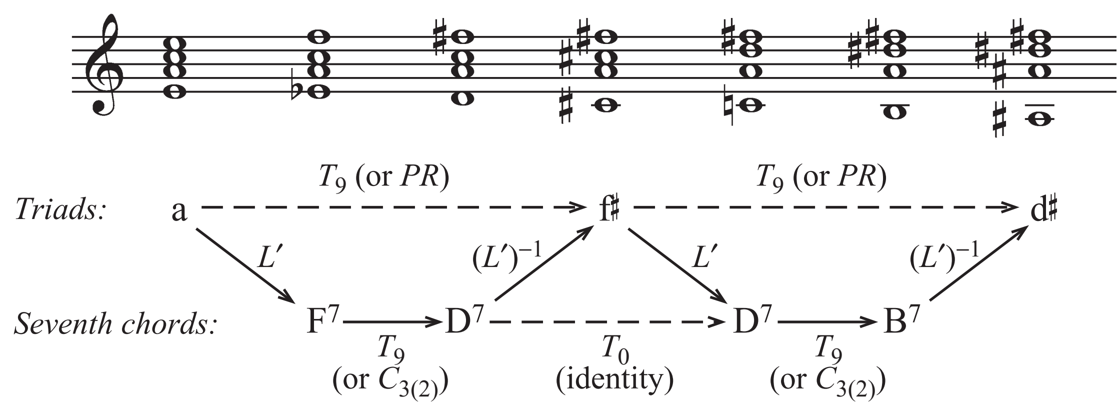
Keywords: Transformational theory, cross-type transformations, Generalized Interval Systems, interscalar mappings, path consistency

Transformational music theory, spearheaded by David Lewin, has flourished in large part due to its astonishingly broad range of application. Transformational techniques have been applied to pitch structures and rhythmic structures, pitch space and pitch-class space, melodic and harmonic constructions, tonal and atonal repertoires; expansive new ground seems to be broken with each passing year. Some of the recent work, however, does not fit entirely within the confines of Lewin's formalism. The uncommon degree of precision with which Lewin delineated his theory resulted in many carefully-presented "rules" for transformational analysis; while Lewin himself (usually) followed his rules scrupulously, much recent work, consciously or not, has violated them in various ways. The aim of this study is not to criticize Lewin for his restrictive formalism or others for failing to respect it, but rather to consider the motivation behind both the rules and the violations, and to expand the structural foundation of transformation theory as needed to reconcile them.[1]

Two main kinds of extensions to transformation theory will be considered here, one involving the definition of *transformations*, the other involving the definition of transformation *graphs* and *networks*. As an illustration of the first, consider the "omnibus" progression shown in Example 1.[2] The omnibus features many hallmarks of the harmonic progressions that have been effectively analyzed using neo-Riemannian

1 Portions of this work were presented at meetings of the Society for Music Theory in 2000 (Toronto) and 2002 (Columbus), and appear also in the author's dissertation (Hook 2002a). I am grateful to David Clampitt, the late John Clough, Richard Cohn, Robert Cook, Jack Douthett, and Eric Isaacson for their comments on various aspects of this work.

2 This analysis appears in a somewhat different form in Hook 2002b, 116–18.



example 1. *Cross-type analysis of an omnibus progression.*

transformational techniques: it is a highly chromatic succession of familiar sonority types, featuring root movement by third (not fifth) and extremely smooth voice leading. But traditional transformation theory is confounded by the mixture of triads and seventh chords appearing here. If the seventh chords are eliminated entirely, the progression is reduced to a series of minor triads, descending by minor thirds (a → f♯ → d♯ → · · ·)—a bare structural skeleton that leaves out two of every three chords and completely disregards the strong voice-leading connections that give coherence to the original progression. If we retain the seventh chords but omit their sevenths, we are confronted with triadic juxtapositions such as F → D, which might be modeled as a transposition T_9 or a neo-Riemannian composite *RP*. (In this article transformations are always combined in left-to-right orthography; thus *RP* means "*R*, then *P*.") The linear connection between the seventh chords, however, is much stronger than either T_9 or *RP* would suggest; indeed, it is smoother than any possible connection between F major and D major triads. This connection between seventh chords is aptly modeled by a transformation that Adrian Childs has called $C_{3(2)}$, but Childs's system, like a similar formalization by Edward Gollin, applies to seventh chords only and offers no way to relate them to the triads (Childs 1998; Gollin 1998).[3]

The network shown beneath the progression in Example 1 offers a more satisfactory analysis of the omnibus than any of the above alternatives. Triads appear in the top row of this network and seventh chords in the bottom row, and the progression zigzags between them, following the solid arrows. The first transformation (a → F^7), is labeled L'; its action is similar to the neo-Riemannian *leittonwechsel* L (a → F) but it yields a seventh chord rather than a triad. More precisely: for any major or minor triad X, $L'(X)$ is by definition the unique major-minor or half-diminished seventh chord that contains all the notes of $L(X)$. Thus L' maps a C major triad to a C-sharp half-diminished seventh chord (which contains the notes of the E minor triad, L of C major), and maps a C minor triad to an A-flat major-minor seventh chord.

3 Childs's notation $C_{3(2)}$ indicates that two of the voices move in contrary motion from one seventh chord to the next, with the stationary voices forming interval class 3 and the moving voices (initially) forming interval class 2.

The next step in the progression ($\mathrm{F}^7 \rightarrow \mathrm{D}^7$) could be modeled by Childs's $C_{3(2)}$ or by the pc-set transposition T_9. The third transformation ($\mathrm{D}^7 \rightarrow$ f♯) is a seventh-chord-to-triad transformation that exactly reverses the action of L' and is therefore labeled $(L')^{-1}$. The resulting network includes all the chords in the progression, shows that the extended omnibus cycles through these three transformations repeatedly, and conveys the long-range T_9 relations between triads as well. It might be objected that the network itself does not explain the progression's smooth voice leading, but a quick examination of the relationship of the pitch classes in a triad X and the seventh chord $L'(X)$ will demonstrate the smooth voice-leading potential of the transformation L' and hence of the omnibus in general.[4]

The transformation L' is the crucial link in the analysis of Example 1. This L' is a *cross-type transformation*, which is to say that it transforms objects of one type (triads) into objects of a different type (seventh chords)—and therein lies the problem with this example in Lewinian transformational terms. According to the most fundamental definition in Lewin's *Generalized Musical Intervals and Transformations*, cross-type transformations are not transformations at all: a transformation is defined as "a function from a family [any set of objects] S into S itself" (Lewin 1987, 3, Definition 1.3.1). The restriction "into S itself" disqualifies mappings from one set S (such as the set of triads) into a different set T (such as the set of seventh chords) from consideration as musical transformations, notwithstanding their mathematical validity or musical suggestiveness. The omnibus example already demonstrates the considerable analytical potential of cross-type transformations, a potential that several subsequent examples will reinforce.

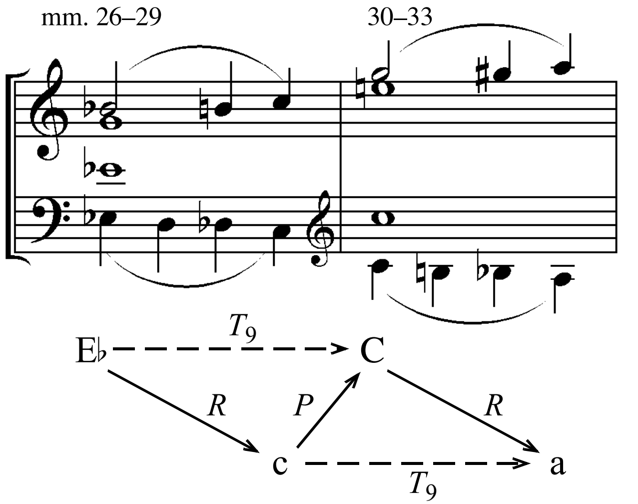
Part I of this paper will develop the general theory of cross-type transformations and situate it in relation to Lewin's theory. A case of particular interest arises when the sets S and T are Generalized Interval Systems in the sense defined by Lewin. In this situation we may study the relationship between a cross-type transformation and the interval structures of the two GISes involved. When these are compatible in a certain way, the mapping may be called a *GIS homomorphism*. Part II presents a theory of GIS homomorphisms, including a general theorem asserting the existence of GIS homomorphisms under a wide variety of circumstances. Transposition and inversion operators are among the most fundamental transformations within one GIS; in Part III we define *cross-type transpositions* and *cross-type inversions* as an application of the results of Part II. Many interesting cross-type transpositions and inversions can be found in examples combining two different scale structures; such examples constitute the bulk of Part IV.

A second main category of extensions to transformational theory arises in conjunction with transformation networks such as the one in Example 2, based on an analysis by Robert Cook of an eight-measure passage from César Franck's Piano Quintet.[5] The first four measures of this excerpt (corresponding to one measure of the reduction) trace a path from E-flat major to C minor, modeled as an R transformation. There follows a mode shift (P) to C major, after which precisely the same music repeats a major sixth higher than before (T_9).

As sensible as this analysis appears, it violates another of Lewin's requirements. The transformations themselves are well-behaved; the problem lies in the construction of the network. Lewin requires that whenever there are two different directed paths ("arrow chains") from one node to another in a graph, the products of the transformation labels along both paths must be identical (Lewin 1987, 195, Definition 9.2.1, condition (D)). Applied to the two paths from E♭ to C in Example 2, this requirement demands that the composite transformation RP ("R, then P") must be identical with the

4 A powerful technique for analyzing the potential voice-leading closeness of any two chords has been devised by Dmitri Tymoczko (2005, section IV).

5 See Cook (forthcoming).



EXAMPLE 2. *Franck, Piano Quintet in F Minor, i, mm. 26–33: score reduction and transformational analysis (after Robert Cook).*

single transformation T_9. At first blush this may appear to be true: certainly when applied to the E-flat major triad, *RP* gives C major, the same result as T_9. But as abstract transformations, *RP* is *not* the same as T_9, as can be seen by applying them to any *minor* triad: for example, *RP* applied to C minor gives E-flat minor, while T_9 applied to C minor gives A minor. *RP* is actually a Riemannian *schritt* rather than an ordinary transposition: it transposes major triads in one direction and minor triads an equal distance in the other direction.[6] Nevertheless, the last four measures of the passage in question are a precise transposition of the first four, and T_9, not *RP*, is the natural description of this relationship. (*RP*, in fact, is not a possible label for the lower dashed arrow in the network of Example 2; if a Riemannian transformation is desired here, it must be *PR*.)

I shall refer to Lewin's requirement on transformational products along paths as the *path consistency condition*, and to transformation graphs satisfying this condition as *path-consistent*. Lewin requires *all* graphs to be path-consistent, but Cook's analysis illustrates that non-path-consistent graphs can sometimes be musically more illuminating than any path-consistent alternative. Part V of this paper develops a theory of path consistency and several other related properties of graphs and networks, and concludes with a cross-type, non-path-consistent, but highly suggestive analysis of two passages from an opera by Rimsky-Korsakov.

I. CROSS-TYPE TRANSFORMATIONS

Mappings from one set to another are everyday fare in mathematics, and are by no means unknown in music theory.[7] The literature offers several examples of mappings between chords of different types, similar to L' discussed above. Elsewhere I have described the *inclusion transformation*, which maps any major or minor triad to the unique dominant or half-diminished seventh chord that contains it (and which combines with the usual *leittonwechsel* to produce L'); Guy Capuzzo has employed this same transformation in his study of pop-rock music. More often, such mappings have been considered only implicitly, by means of chordal *relations* not formalized as mathematical functions; examples include the split/fuse relations studied by Clifton Callender, the relations between chords of different set classes appearing in the graphs called "Cube Dance" and "Power Towers" by Douthett and Steinbach, and the voice-leading relations between sets of different cardinalities studied by Joseph Straus (Hook 2002b, 117–18; Capuzzo 2004,

6 This equal-and-opposite behavior is characteristic of Riemannian transformations; see Hook 2002b, 74–83.

7 In this paper we regard the words *function*, *mapping*, and *transformation* as synonyms.

191–92; Callender 1998, 224; Douthett and Steinbach 1998, 254–56; Straus 2003, 313).

Chords of different types, however, constitute only one of many possible scenarios in which cross-type transformations may prove fruitful. Whenever we invoke octave equivalence, we are implicitly applying a mapping from pitch space to pitch-class space (a mapping that will be defined precisely in Part II below). Stephen Soderberg and Matthew Santa have defined functions from structures in a mod-m pitch-class space to structures in a mod-n pitch-class space; both are somewhat similar to some of the transformations to be considered in Part IV. Studies of multi-parametric serial works sometimes involve (for example) mappings from the pitch domain to the rhythm domain. At a more abstract, metaphorical level, Lawrence Zbikowski has discussed "cross-domain mapping" as a general category of cognitive processes with many implications in music theory (WARP functions in Soderberg 1998, paragraph 2.5; MODTRANS operator in Santa 1999, 202; discussion of duration rows and time point rows in Mead 1994, 38–53; Zbikowski 2002, Chapter 2).

Cross-type mappings make occasional appearances in Lewin's work as well. Examples include the interval function of a Generalized Interval System (which maps pairs of elements of the GIS into the interval group, as discussed further in Part II below), IFUNC (which maps an interval group into the integers), embedding functions (which map pairs of sets of elements into the integers), and, in the discussions of graph and network isomorphism and isography, several explicit cross-type mappings from the objects of one network to the objects of another (Lewin 1987: 26, 88, 105, 198–202). Like most of the other authors cited above, however, Lewin does not regard any such mappings as "transformations" that are themselves suitable for use in transformational analyses and networks—a constraint that seems on occasion to limit the scope of his results. His transformation network for Dallapiccola's "Simbolo," for example, breaks apart into two disconnected components because the two sections of the piece are based on two different kinds of "configuration" that no single-type transformation can relate (Lewin 1993, 13, Example 1.5).

In this section we review some elementary terminology and propose some adaptations to Lewin's formalism in order to accommodate cross-type mappings as transformations in full standing. We consider, in general, a function f defined on a set S (consisting of objects of one "type") and taking values in a set T (objects possibly of a different "type"). This situation is commonly abbreviated $f\colon S \to T$. The set S is called the *domain* of the function, and T its *range*.[8] Such a function may or may not be *one-to-one* (which means that whenever x_1 and x_2 are two different elements of S, then $f(x_1)$ and $f(x_2)$ are two different elements of T) or *onto* (which means that every element of T is $f(x)$ for some x in S). If f is both one-to-one and onto, then it is a *bijection*, or a *one-to-one correspondence*, between the sets S and T.[9] If S is a finite set, we

8 It will have occurred to many readers that one can readily form the set-theoretic union $S \cup T$ of the two sets—or more generally a single set combining objects of all the "types" under consideration—and can then undertake to define appropriate functions on this larger set. Such mappings can then be considered single-type transformations in the Lewinian sense. In some cases this approach can be made to work, effectively circumventing the limitation of Lewin's definition. For instance, the triad-to-seventh-chord mapping L' described above may be taken together with its seventh-chord-to-triad inverse mapping to define a single function on the set of all 48 triads and seventh chords of the appropriate qualities, by means of which the omnibus analysis of Example 1 may be reconstructed. This workaround, however, is somewhat counterintuitive at best (are L' and its inverse really "the same function"?); it often introduces its own formal complications (any other transformations appearing in the same graph must also be defined on $S \cup T$); and in many situations (particularly involving sets of different cardinalities or functions that do not have inverses) it proves completely unworkable. Even when a single-type solution is theoretically possible, the cross-type approach is often simpler and more natural.

9 Lewin's use of the word *operation* for a (single-set) mapping that is both one-to-one and onto (Lewin 1987, 3) is not standard mathematical terminology. We shall avoid this term here in discussing cross-type transformations.

write $|S|$ for the cardinality of S (the number of elements in the set). If S and T are finite sets, there can be no one-to-one function from S to T unless $|S| \leq |T|$, and there can be no function mapping S *onto* T unless $|S| \geq |T|$. Hence there can be no bijection from S to T unless $|S| = |T|$.

Suppose f: $S \rightarrow T$ and g: $T \rightarrow U$ are functions, where the same set T serves as both the range of f and the domain of g. Then the *composite function* $h = fg$ is the mapping h: $S \rightarrow U$ defined by the equation $h(x) = g(f(x))$ for all x in S.[10] Notice that fg is undefined unless the range of f is the same set as (or is a subset of) the domain of g. Even if fg is defined, the composite gf of the same two functions in the opposite order will be undefined unless U happens to be the same as (or a subset of) S; if this is the case, then gf is a function from T to T. These restrictions on the construction of composite functions are obvious ones, but they lead to inevitable complications in cross-type transformation theory. Any number of functions from S to S can be composed in any order, and as Lewin details, such functions form *transformation groups* in many cases. When cross-type mappings are allowed, we must accept the fact that we cannot always compose them and cannot expect to encounter a group structure. Indeed, it seems likely that this complication was one of the primary reasons for Lewin's disinclination to admit cross-type transformations: the algebraic messiness of the theoretical apparatus these transformations require is somewhat at odds with the elegance of the formalism with which Lewin so adeptly modeled his own musical intuitions.

As long as domains and ranges are properly coordinated, composite functions continue to obey many familiar algebraic laws. For example, let E_S: $S \rightarrow S$ denote the *identity mapping* on a set S; that is, $E_S(x) = x$ for every x in S. Likewise let E_T denote the identity mapping on T. Then for any function f: $S \rightarrow T$, it should be clear that $E_S f = f$ and $fE_T = f$. If f: $S \rightarrow T$ is one-to-one and onto, then its *inverse function* f^{-1}: $T \rightarrow S$ is well-defined; for every y in T, $f^{-1}(y)$ is the unique element x of S for which $f(x) = y$. The functions f and f^{-1} satisfy the composite equations $ff^{-1} = E_S$ and $f^{-1}f = E_T$. (An illustration of the latter can be seen in the omnibus analysis of Example 1, where the triangular segment in the center of the network indicates that the composite function $(L')^{-1}(L')$ is equal to T_0, the identity mapping on seventh chords.) Of course, when S and T are the same set, all of this is simpler: there is only one identity mapping, and the composites ff^{-1} and $f^{-1}f$ are the same.

Lewin's axiomatic formulation of transformation graphs and networks requires that there be a single set S from which the contents of all the nodes in a network (the objects upon which the transformations act) are assigned, and that all arrows be labeled with transformations on S (Lewin 1987, 193–197, especially 196, Definition 9.3.1). In the more general cross-type setting, the various nodes N_1, N_2, . . . may have their contents selected from various sets S_1, S_2, ... We shall call S_i the *space* associated with the node N_i. Of course, some of the spaces S_i *may* be the same, but they are no longer required to be. In the omnibus analysis of Example 1, the nodes in the top row are associated with the set of (major and minor) triads, while those in the bottom row are associated with the set of (major-minor and half-diminished) seventh chords.

Lewin requires that each arrow be labeled with a transformation belonging to some semigroup of transformations on S. For the reasons noted above, we cannot necessarily expect a semigroup or any other particular algebraic structure in the cross-type setting; all we can require is that an arrow from node N_i to node N_j be labeled with a mapping from S_i to S_j.

Lewin draws a careful distinction (to be followed here) between transformation *networks* and transformation *graphs*.

10 Note again the "chronological" left-to-right orthography convention, followed throughout this paper. The function f is applied before g, and we write the composite function accordingly as fg, even though the functions appear in the opposite order in the expression $g(f(x))$. Right-to-left orthography, in which this composite would be called gf, is in common use, but leads to ever greater confusion as more than two functions are composed. See further discussion in Lewin 1987, 2 and Hook 2002a, 193.

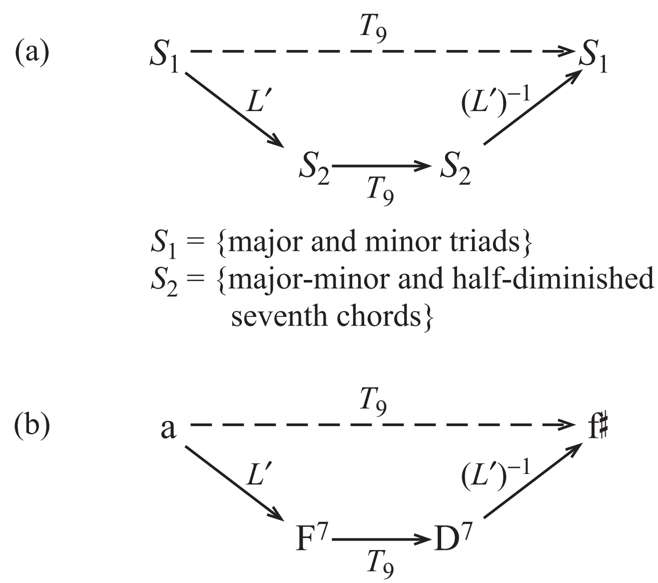
In a transformation *network*, the contents of the nodes have been assigned, while in a transformation *graph* they have not (only the arrows are labeled). The *graph* is thus the more abstract of the two structures; a *network* adds information to a graph, bringing it to life in a real musical setting via the node labels. In the cross-type setting, it is important to remember that the set S_i to which each node N_i "belongs" must be specified, *even in a transformation graph*. That is, even if the contents of the nodes have not been assigned, it is essential to know the space(s) from which they *may* be assigned. This understanding is implicit in Lewin's discussion of single-type transformation graphs, where there is only one space and no confusion is likely; in the cross-type setting the point must be made explicitly.

A natural way to exhibit the spaces graphically in a transformation graph is to label each node with the name of its associated space. In a transformation *network*, of course, each node is labeled with its actual contents, an element of that space. See, for instance, Example 3, derived from Example 1. Two spaces are involved: S_1, the set of triads, and S_2, the set of seventh chords. The figure in part (a) is a transformation graph, while (b) is a transformation network. In such situations we shall say that the network (b) *realizes* the graph (a), and that (a) is the *underlying transformation graph* of network (b).

Lewin introduces two additional constraints in his definitions of transformation graphs and networks. One of these, which applies both to graphs and to networks, is the path consistency condition, to which we shall return in Part V; for the time being we simply note that path consistency is not required of graphs and networks in this paper, although all of those appearing in Parts I–IV happen to be path-consistent. Lewin's other requirement, for networks only, might be called *label consistency*. A separate consideration from path consistency, label consistency requires that the arrow labels in a network be consistent with the node labels. More precisely, if an arrow labeled with a transformation f points from a node labeled with an object x to a node labeled with an object y, then it must actually be the case that $f(x) = y$. The transformations in a network must "tell the truth": false statements such as

$$C \xrightarrow{T_5} A\flat,$$

asserting that A-flat is a 5-semitone transposition of C, are violations of label consistency and are forbidden in transformation networks. The label consistency requirement will be retained here for cross-type networks as well as for single-type ones.



EXAMPLE 3. *A cross-type transformation graph (a) and network (b) derived from Example 1.*

Summarizing our generalized definitions, a *transformation graph* consists of:

- an underlying *directed graph*, comprising a set of *nodes* and a set of *arrows* linking certain pairs of nodes;

- for each node N_i, a set S_i, called the *space associated with* N_i; and
- for each arrow A from node N_i to node N_j, an associated transformation f_A: $S_i \rightarrow S_j$.

A *transformation network* consists of:

- an underlying *transformation graph*, with nodes N_i, arrows A, spaces S_i, and transformations f_A as above; and
- for each node N_i, an element x_i of the associated space S_i, called the *contents* of N_i.

Transformation networks are required to satisfy the *label consistency* property:

- for each arrow A from node N_i to node N_j, the associated transformation f_A and contents x_i and x_j must satisfy $f_A(x_i) = x_j$.

II. GIS HOMOMORPHISMS

Some types of transformations (notably transpositions and inversions) depend on an *interval* structure defined on the space(s) involved. In this section we shall study cross-type mappings from one Generalized Interval System (GIS) to another, especially those of a particular kind called *GIS homomorphisms*. This work may seem forbiddingly abstract at first, but we shall reap the benefits in the next section when the theory of cross-type transpositions and inversions falls out with relative ease.[11]

A GIS, as defined by Lewin, has three components: a set S, called the *space* of the GIS; a group G (called IVLS by Lewin), the *interval group* of the GIS; and a function int: $S \times S \rightarrow G$, the *interval function* of the GIS, which maps ordered pairs of elements of S into the interval group G (26, Definition 2.3.1). The group G need not be commutative in general, but it is commutative in the cases that will concern us here, so we will use additive notation for intervals. A GIS is subject to the following two requirements:

(a) For all x, y, and z in S, $\mathrm{int}(x, y) + \mathrm{int}(y, z) = \mathrm{int}(x, z)$; and

(b) For every x in S and every i in G, there exists a unique element y in S such that $\mathrm{int}(x, y) = i$.

In principle, a GIS is an ordered triple (S, G, int), but informally we will often refer to the set S itself as the GIS when the interval structure is clear from context.

Now suppose that $(S_1, G_1, \mathrm{int}_1)$ and $(S_2, G_2, \mathrm{int}_2)$ are two GISes. The intent behind the definition of a *GIS homomorphism* from S_1 to S_2 is that it be a mapping that preserves all GIS structural relationships among the elements being mapped. (This is in keeping with the accepted use of the word *homomorphism* for structure-preserving mappings among various kinds of mathematical structures: group homomorphisms, graph homomorphisms, and so on.)[12] To be precise, a GIS homomorphism consists of *two* functions, f: $S_1 \rightarrow S_2$ and g: $G_1 \rightarrow G_2$, satisfying the following two conditions:

(a) For all i and j in G_1, $g(i + j) = g(i) + g(j)$; and

(b) For all x and y in S_1, $g(\mathrm{int}_1(x, y)) = \mathrm{int}_2(f(x), f(y))$.

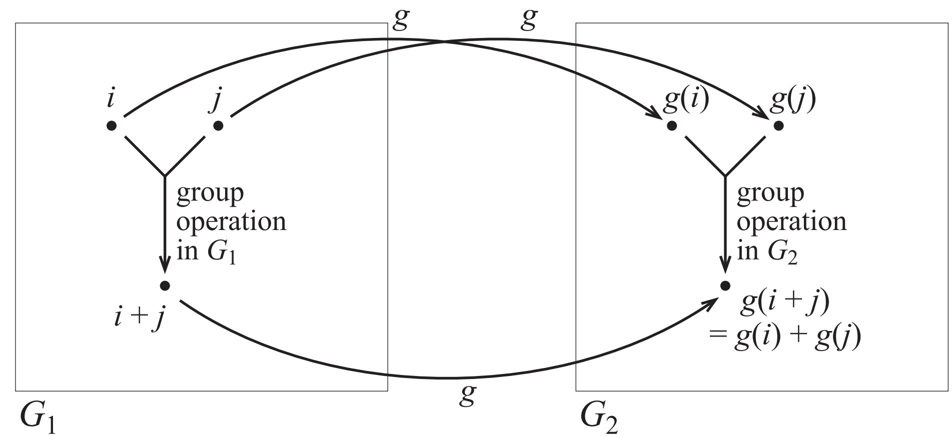
Condition (a) is familiar: it simply states that g is a group homomorphism from G_1 to G_2. That is, the mapping g, from one group to another, preserves the group operation (+), as illustrated in Example 4(a): "g-related elements have g-related sums." Condition (b) may be described in somewhat similar fashion, as shown in Example 4(b): "f-related elements form g-related intervals."

Strictly, a GIS homomorphism always consists of two components (f, g) as just described; informally, however, we may refer to the mapping f itself as the GIS homomorphism if the

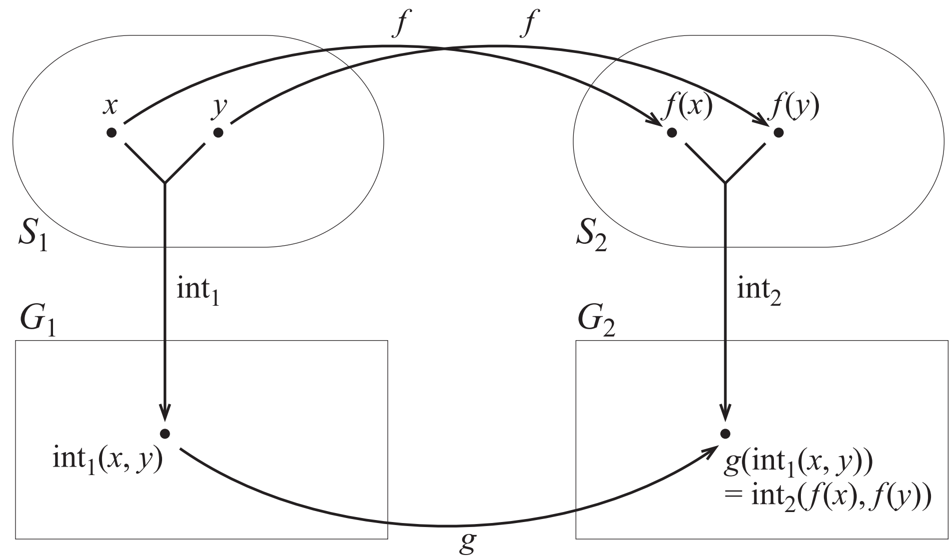
11 Oren Kolman has independently explored a number of ideas related to the concepts discussed in this section, including formulations of GIS homomorphism and isomorphism equivalent to those presented here. The GIS Homomorphism Theorem (Theorem 2.1 below) illustrates an equivalence of group structure and GIS structure that forms the essence of Kolman's "Second Transfer Principle" (Kolman 2004, 158).

12 For mathematical background, the reader may wish to consult an abstract algebra text such as Dummit and Foote 2004.

(a)



(b)



EXAMPLE 4. *The two requirements for a GIS homomorphism (f, g).*

group mapping g is obvious. (In many important applications the groups G_1 and G_2 are the same, and g is the identity mapping on the group.) If both f and g are one-to-one and onto, then the GIS homomorphism (f, g) is called a *GIS isomorphism*; in this case the GISes S_1 and S_2 are *isomorphic*, which is to say that their structures are essentially identical except for the names given to the objects.

Two simple examples of GIS homomorphisms may be helpful at this point. Let $\mathcal{C}$ be chromatic pitch space. That is, $\mathcal{C}$ is the set {. . . , B♭3, B3, C4, C♯4, D4, . . .} (C4 = middle C), theoretically extending infinitely in both directions. This space has a GIS structure in which the interval ($\mathrm{int}_\mathcal{C}$) between two pitches is the ordinary directed pitch interval, measured in semitones; thus $\mathrm{int}_\mathcal{C}(\mathrm{C4}, \mathrm{B3}) = -1$, while $\mathrm{int}_\mathcal{C}(\mathrm{C4}, \mathrm{C5}) = 12$. The interval group is $G_\mathcal{C} = \mathbb{Z}$, the additive (commutative) group of integers.

Let $\mathcal{C}_{12}$ be chromatic pc-space. Thus $\mathcal{C}_{12}$ = {C, C♯, . . . , B} is a finite set of cardinality 12. The interval (int_{12}) between two pitch classes in $\mathcal{C}_{12}$ is the directed pitch-class interval; thus $\mathrm{int}_{12}(\mathrm{C}, \mathrm{B}) = 11$ and $\mathrm{int}_{12}(\mathrm{C}, \mathrm{C}) = 0$. The interval group is $G_{12} = \mathbb{Z}_{12}$, the additive group of integers mod 12.

We define a GIS homomorphism (f, g) from $\mathcal{C}$ to $\mathcal{C}_{12}$ as follows. The function f maps each pitch in $\mathcal{C}$ to its corresponding pitch class in $\mathcal{C}_{12}$; thus $f(\mathrm{C4}) = f(\mathrm{C5}) = \mathrm{C}$. The function g maps every integer in $\mathbb{Z}$ to its mod-12 reduction in $\mathbb{Z}_{12}$; for instance, $g(23) = g(-1) = 11$. Condition (a) in the definition of a GIS homomorphism (g is a group homomorphism) is a standard algebraic result; g is the natural mapping from a group to one of its *quotient groups*. Condition (b) is also straightforward. For example, taking x and y to be the pitches E5 and G3 respectively, we calculate that

$$g(\mathrm{int}_\mathcal{C}(x, y)) = g(\mathrm{int}_\mathcal{C}(\mathrm{E5}, \mathrm{G3})) = g(-21) = 3,$$

while

$$\mathrm{int}_{12}(f(x), f(y)) = \mathrm{int}_{12}(f(\mathrm{E5}), f(\mathrm{G3})) = \mathrm{int}_{12}(\mathrm{E}, \mathrm{G}) = 3.$$

The homomorphism (f, g) is clearly not an isomorphism, as neither f nor g is one-to-one. (In fact, the GISes $\mathcal{C}$ and $\mathcal{C}_{12}$ cannot possibly be isomorphic, as one is an infinite set and the other finite.)

For the second example, let $\mathcal{C}$ be chromatic pitch space as above, and let $\mathcal{D}$ be white-key diatonic pitch space, the infinite set {. . . , A3, B3, C4, D4, E4, . . .}. As a set, $\mathcal{D}$ is a subset of $\mathcal{C}$, but the two GISes measure intervals differently on the pitches they share. The diatonic interval ($\mathrm{int}_\mathcal{D}$) between two pitches in $\mathcal{D}$ is the number of diatonic steps separating them, positive if ascending, negative if descending; thus $\mathrm{int}_\mathcal{D}(\mathrm{C4}, \mathrm{B3}) = -1$ and $\mathrm{int}_\mathcal{D}(\mathrm{C4}, \mathrm{C5}) = 7$. The interval group $G_\mathcal{D}$, consequently, is again the group $\mathbb{Z}$ of integers, the same group as $G_\mathcal{C}$.

Now we define a GIS homomorphism (f', g') from $\mathcal{C}$ to $\mathcal{D}$ as follows. First we define f' so that it fixes middle C: $f'(\mathrm{C4}) = \mathrm{C4}$. For any other pitch x in $\mathcal{C}$, define $f'(x)$ to be the unique white-note pitch y in $\mathcal{D}$ such that the diatonic interval $\mathrm{int}_\mathcal{D}(\mathrm{C4}, y)$ is equal to the chromatic interval $\mathrm{int}_\mathcal{C}(\mathrm{C4}, x)$. For example, $f'(\mathrm{B}\flat 4) = \mathrm{F5}$, because B♭4 is 10 chromatic steps above C4 and the pitch 10 diatonic steps above C4 is F5. To complete the definition, g' is defined to be $E_\mathbb{Z}$, the identity mapping on the additive group of integers. Obviously g' is an isomorphism, so condition (a) in the definition of a GIS homomorphism is automatic. The definition of f' ensures that condition (b) is also satisfied, as demonstrated by the following chain of equalities:

$$\begin{aligned}
&g'(\mathrm{int}_\mathcal{C}(x, y)) \\
&\quad = \mathrm{int}_\mathcal{C}(x, y) && \textit{(definition of } g'\textit{)} \\
&\quad = \mathrm{int}_\mathcal{C}(x, \mathrm{C4}) + \mathrm{int}_\mathcal{C}(\mathrm{C4}, y) && \textit{(property of any GIS)} \\
&\quad = -\mathrm{int}_\mathcal{C}(\mathrm{C4}, x) + \mathrm{int}_\mathcal{C}(\mathrm{C4}, y) && \textit{(property of any GIS)} \\
&\quad = -\mathrm{int}_\mathcal{D}(\mathrm{C4}, f'(x)) + \mathrm{int}_\mathcal{D}(\mathrm{C4}, f'(y)) && \textit{(definition of } f'\textit{)} \\
&\quad = \mathrm{int}_\mathcal{D}(f'(x), \mathrm{C4}) + \mathrm{int}_\mathcal{D}(\mathrm{C4}, f'(y)) && \textit{(property of any GIS)} \\
&\quad = \mathrm{int}_\mathcal{D}(f'(x), f'(y)). && \textit{(property of any GIS)}
\end{aligned}$$

The above observations establish that (f', g') is a GIS homomorphism. Both functions are one-to-one and onto; hence (f', g') is a GIS isomorphism. Although chromatic pitch space and diatonic pitch space may appear quite different, their GIS structures are in fact isomorphic. Intuitively this is because intervals in both spaces are measured simply by counting "steps," indexed by the integers; the group structure is oblivious to the acoustical sizes of the steps.

This isomorphism turns out to be an example of the cross-type transpositions that we will consider in Part III of this paper. It is a useful example to keep in mind as we work through some general properties of GIS homomorphisms. In fact, it is a special case of the construction described in the following theorem.

Theorem 2.1 (GIS Homomorphism Theorem). Suppose that (S_1, G_1, int_1) and (S_2, G_2, int_2) are GISes; u is an element of S_1; v is an element of S_2; and g: $G_1 \to G_2$ is a group homomorphism. Then there exists a unique function f: $S_1 \to S_2$ such that $f(u) = v$ and (f, g) is a GIS homomorphism. Moreover, (f, g) is a GIS isomorphism if and only if g is a group isomorphism. (A proof of this theorem may be found in the Appendix to this paper.)

Once reference points u in S_1 and v in S_2 have been chosen, the GIS Homomorphism Theorem guarantees that any group homomorphism from the interval group of S_1 to the interval group of S_2 gives rise to one and only one GIS homomorphism that maps u to v. The function f is completely determined by the group homomorphism g and the two reference points u and v. We shall designate this f by the notation $\overline{g}^{v}_{u}$ (or, more explicitly, $\overline{g}^{v\in S_2}_{u\in S_1}$) and call it "the GIS homomorphism induced by g that maps x to y." (We will be using similar notation on several later occasions. Notice that u appears as a subscript and v as a superscript; these functions will always map *from subscript to superscript*, a notational convention that some readers will recognize from definite integrals and other standard mathematical constructions. This arrangement is also required for consistency with Lewin's notation for inversions, to be discussed in Part III below.)

Each of the two examples of GIS homomorphisms given previously—indeed, any GIS homomorphism at all—can arise as an induced GIS homomorphism in the manner described by the theorem. Let us briefly reconsider the second example, the mapping from chromatic pitch space $\mathcal{C}$ to diatonic pitch space $\mathcal{D}$. Here both interval groups are the group of integers $\mathbb{Z}$, and the group homomorphism is $g' = E_{\mathbb{Z}}$. The function f' was explicitly defined so as to map C4 (in $\mathcal{C}$) to C4 (in $\mathcal{D}$). Therefore $f' = \overline{(E_{\mathbb{Z}})}^{\mathrm{C4}\in\mathcal{D}}_{\mathrm{C4}\in\mathcal{C}}$. Notice, however, that the same function f' could also be induced by other choices of the pair of reference points; we could just as well write $f' = \overline{(E_{\mathbb{Z}})}^{\mathrm{D4}\in\mathcal{D}}_{\mathrm{C\sharp4}\in\mathcal{C}}$ or $f' = \overline{(E_{\mathbb{Z}})}^{\mathrm{A5}\in\mathcal{D}}_{\mathrm{C5}\in\mathcal{C}}$. The general question of alternate notations for the same induced GIS homomorphism will be addressed in Theorem 2.4 below.

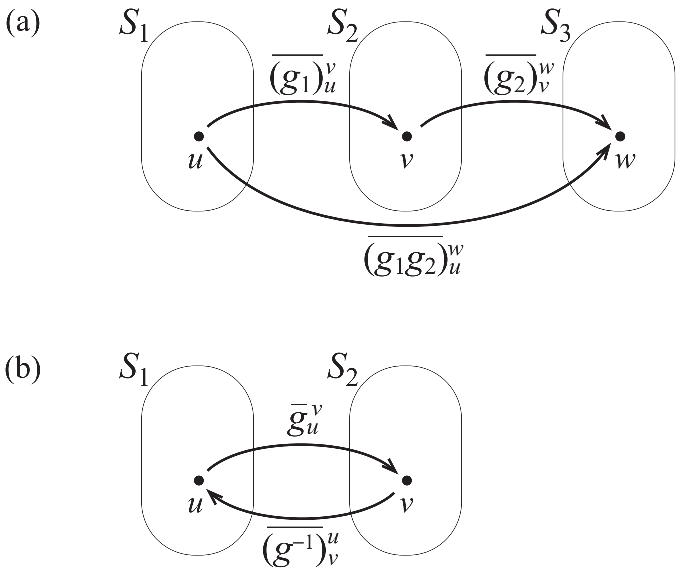
The remainder of this section consists of a series of straightforward but rather abstract propositions about GIS homomorphisms. The first two of these results (Theorems 2.2 and 2.3) show that compositions and inverses of GIS homomorphisms behave as we should expect; the situations described in these two theorems are illustrated by the networks appearing in Example 5. The third result (Theorem 2.4) specifies exactly the conditions under which a different choice of the reference points u and v will result in the same induced GIS homomorphism. Once the reader has assimilated the notational complexity, these properties should be fairly intuitive. Proofs are omitted here, but mathematically inclined readers should be able to supply them readily.

Theorem 2.2 (composition of GIS homomorphisms).

(a) Suppose S_1, S_2, and S_3 are GISes; (f_1, g_1) is a GIS homomorphism from S_1 to S_2; and (f_2, g_2) is a GIS homomorphism from S_2 to S_3. Then ($f_1 f_2$, $g_1 g_2$) is a GIS homomorphism from S_1 to S_3. If both (f_1, g_1) and (f_2, g_2) are GIS isomorphisms, then ($f_1 f_2$, $g_1 g_2$) is a GIS isomorphism.

(b) Suppose (S_1, G_1, int_1), (S_2, G_2, int_2), and (S_3, G_3, int_3) are GISes; u, v, and w are elements of S_1, S_2, and S_3, respectively; and g_1: $G_1 \to G_2$ and g_2: $G_2 \to G_3$ are group homomorphisms. Then

$$\overline{(g_1)}^{v}_{u}\,\overline{(g_2)}^{w}_{v} = \overline{(g_1 g_2)}^{w}_{u}.$$



EXAMPLE 5. *Compositions (a) and inverses (b) of GIS homomorphisms.*

Theorem 2.3 (inverses of GIS isomorphisms).

(a) Suppose S_1 and S_2 are GISes, and (f, g) is a GIS isomorphism from S_1 to S_2. Then (f^{-1}, g^{-1}) is a GIS isomorphism from S_2 to S_1.

(b) Suppose $(S_1, G_1, \mathrm{int}_1)$ and $(S_2, G_2, \mathrm{int}_2)$ are GISes; u is an element of S_1; v is an element of S_2; and g: $G_1 \rightarrow G_2$ is a group isomorphism. Then

$$(\overline{g}^v_u)^{-1} = \overline{(g^{-1})}^u_v .$$

Theorem 2.4 (change of reference points). Suppose $(S_1, G_1, \mathrm{int}_1)$ and $(S_2, G_2, \mathrm{int}_2)$ are GISes; u and x are elements of S_1; v and y are elements of S_2; and g: $G_1 \rightarrow G_2$ is a group homomorphism. Then $\overline{g}^v_u = \overline{g}^y_x$ if and only if $\overline{g}^v_u(x)=y$.

Let (S, G, int) be a GIS, and let a reference point u in S be selected. As Lewin shows, there is then a natural way to label each element of S with an interval in G, namely $\mathrm{lab}_u(x) = \mathrm{int}(u, x)$ (Lewin 1987, 31–32).[13] The reference point u is labeled with the identity element of the interval group, and each other element is labeled with its interval measured from u. The function lab_u: $S \rightarrow G$ is one-to-one and onto because of condition (b) in the definition of a GIS, and the equation $\mathrm{int}(x, y) = \mathrm{lab}_u(y) - \mathrm{lab}_u(x)$ (in additive group notation, assuming commutativity) holds for all x and y in S. The following theorem, proved in the Appendix, show how to calculate the label of the value of any induced GIS homomorphism.

Theorem 2.5 (label functions for induced GIS homomorphisms). Suppose $(S_1, G_1, \mathrm{int}_1)$ and $(S_2, G_2, \mathrm{int}_2)$ are GISes; u and x are elements of S_1; v and y are elements of S_2; lab_u: $S_1 \rightarrow G_1$ and lab_v: $S_2 \rightarrow G_2$ are the label functions associated with u and v; and g: $G_1 \rightarrow G_2$ is a group homomorphism. Then for every a in S_1,

$$\mathrm{lab}_v(\overline{g}^y_x(a)) = \mathrm{lab}_v(y) - g(\mathrm{lab}_u(x)) + g(\mathrm{lab}_u(a)).$$

The preceding theorem may be applied to derive a general formula for the interval between the values of two different induced GIS homomorphisms, as the final result of this section shows.

Theorem 2.6 (intervals formed by induced GIS homomorphisms). Suppose $(S_1, G_1, \mathrm{int}_1)$ and $(S_2, G_2, \mathrm{int}_2)$ are GISes; u and x are elements of S_1; v and y are elements of S_2; and g: $G_1 \rightarrow G_2$ and h: $G_1 \rightarrow G_2$ are group homomorphisms. Then for all elements a and b of S_1,

$$\mathrm{int}_2(\overline{g}^v_u(a), \overline{h}^y_x(b)) = g(\mathrm{int}_1(a, u)) + \mathrm{int}_2(v, y) + h(\mathrm{int}_1(x, b)).$$

III. CROSS-TYPE TRANSPOSITION AND INVERSION

In a GIS (S, G, int), Lewin defines the transposition operator T_i : $S \rightarrow S$ so that, for any x in S, $T_i(x)$ is the unique y

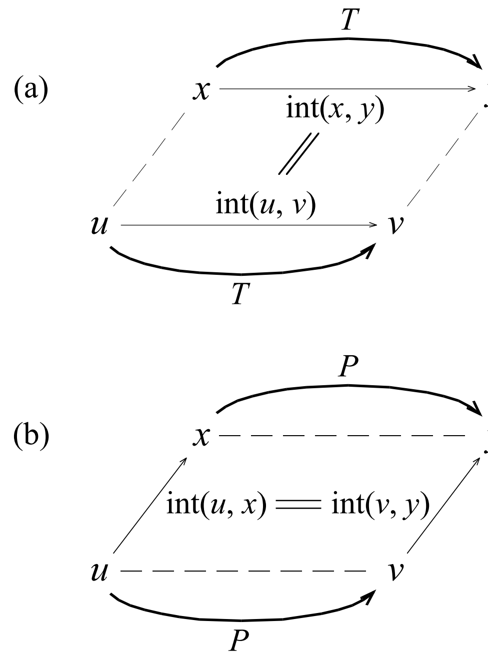
13 Lewin writes "ref" for the reference point u and "LABEL" for the function lab_u.

in S such that $\mathrm{int}(x, y) = i$ (Lewin 1987, 46–50). In a notation that will suit our purposes better here, this function could also be denoted T_u^v, where u and v are any two elements for which the function maps u to v (that is, for which $\mathrm{int}(u, v) = i$). (In general, many different choices of the reference points u and v will give rise to the same transposition. The *interval* from u to v determines the function; if either u or v is chosen arbitrarily, the other is uniquely determined.)

Lewin also discusses *interval-preserving* mappings in the single-GIS situation: a function $P : S \to S$ is interval-preserving if $\mathrm{int}(P(a), P(b)) = \mathrm{int}(a, b)$ for all a and b in S. Let us introduce the notation P_u^v for an interval-preserving function that maps u to v. (It should be checked that this notation is well-defined—that is, that there is only one such function for any u and v. This follows from the interval-preserving property: for any x, $P_u^v(x)$ must be the unique y such that $\mathrm{int}(v, y) = \mathrm{int}(u, x)$.)

Examples 6(a) and 6(b) illustrate the relationship between transpositions and interval-preserving mappings. In each case, a function maps u to v and x to y, and in each case two corresponding intervals are required to match. In the case of transpositions, $\mathrm{int}(u, v) = \mathrm{int}(x, y)$; in the case of interval-preserving mappings, $\mathrm{int}(u, x) = \mathrm{int}(v, y)$. If the interval group is commutative, these two conditions are equivalent, as can be seen by adding $\mathrm{int}(v, x)$ to both sides of the first equation and rearranging the terms on one side. In a commutative GIS, therefore, $T_u^v = P_u^v$; the transpositions and the interval-preserving mappings are precisely the same. In a non-commutative GIS this is not the case, as Lewin shows.

Now consider a cross-type situation in which u and x inhabit one GIS (S_1) and v and y inhabit another (S_2). Here things are more complicated. The first difficulty is that the intervals $\mathrm{int}(u, v)$ and $\mathrm{int}(x, y)$ are not defined, as the elements involved belong to two different spaces. The other two intervals, $\mathrm{int}(u, x)$ and $\mathrm{int}(v, y)$, are defined, but cannot in general be compared because they belong to two different interval groups. Often, however, two different GISes share the same interval group. We saw one such example previ-



EXAMPLE 6. *Matching intervals under transposition (a) and interval-preserving mappings (b) within a single GIS.*

ously: the chromatic and diatonic pitch spaces of Part II share the interval group $\mathbb{Z}$. In such cases $\mathrm{int}(u, x)$ and $\mathrm{int}(v, y)$ belong to the common group, and so the notion of "interval-preserving" is meaningful.

Suppose, then, that (S_1, G, int_1) and (S_2, G, int_2) are two GISes sharing the same interval group G. The two interval functions int_1 and int_2, even though they may be defined on different spaces, have the same range of values. A function P: $S_1 \to S_2$ is then interval-preserving if $\mathrm{int}_2(P(a), P(b)) = \mathrm{int}_1(a, b)$ for all a and b in S_1; again the notation P_u^v (or, more explicitly, $P_{u\in S_1}^{v\in S_2}$) may be introduced for the unique interval-preserving function that maps u to v. We will be working here in the commutative case, in which this function generalizes not only the interval-preserving mappings P but the transpositions T as well; indeed, the transposition-like behavior of these functions is our primary interest. For this reason we shall use the notation T_u^v (or $T_{u\in S_1}^{v\in S_2}$) rather than P_u^v.

As it turns out, the theoretical apparatus developed in Part II tells us much about these generalized transpositions, once we have made one simple observation. Suppose u is an element of S_1 and v is an element of S_2, and let $E_G\colon G \to G$ denote the identity mapping on the group G. The simple observation is that (T_u^v, E_G) is a GIS homomorphism from S_1 to S_2. (Condition (a) in the definition of a GIS homomorphism is immediate, since E_G is certainly a group homomorphism. Condition (b) requires that $E_G(\mathrm{int}_1(a, b)) = \mathrm{int}_2(T_u^v(a), T_u^v(b))$ for all a and b in S_1; because E_G is the identity mapping, this is nothing but the interval-preservation property.) By the GIS Homomorphism Theorem (Theorem 2.1), there is only one function $f\colon S_1 \to S_2$ such that $f(u) = v$ and (f, E_G) is a GIS homomorphism, namely the induced GIS homomorphism $(\overline{E_G})_u^v$. But T_u^v is such a function, so it follows that $T_u^v = (\overline{E_G})_u^v$. Moreover, E_G is a group isomorphism, so (T_u^v, E_G) is a GIS isomorphism; this demonstrates that any two GISes sharing the same interval group are isomorphic.

If the two GISes are actually the same—that is, if $S_1 = S_2 = S$ and $\mathrm{int}_1 = \mathrm{int}_2 = \mathrm{int}$—then presumably T_u^v should be the same as some ordinary transposition T_i. Indeed it is: the chain of equalities

$$
\begin{aligned}
&\mathrm{int}(x, T_u^v(x)) \\
&\quad = \mathrm{int}(x, v) + \mathrm{int}(v, T_u^v(x)) && \textit{(property of any GIS)} \\
&\quad = \mathrm{int}(x, v) + \mathrm{int}(T_u^v(u), T_u^v(x)) && \textit{(because } T_u^v(u) = v\textit{)} \\
&\quad = \mathrm{int}(x, v) + \mathrm{int}(u, x) && \textit{(}T_u^v \textit{ is interval-preserving)} \\
&\quad = \mathrm{int}(u, x) + \mathrm{int}(x, v) && \textit{(G is commutative)} \\
&\quad = \mathrm{int}(u, v) && \textit{(property of any GIS)}
\end{aligned}
$$

demonstrates that $T_u^v(x)$ is the unique y such that $\mathrm{int}(x, y) = i$, where $i = \mathrm{int}(u, v)$, and this is precisely Lewin's definition of the function T_i. In case $u = v$, then $i = \mathrm{int}(u, v) = 0$ (the identity element of G), so $T_u^u = T_0$ is the identity mapping on S_1.

The following theorem enumerates various fundamental properties of the cross-type transpositions T_u^v. Parts (a)–(e) summarize the above discussion; (f)–(j) follow directly from the general results about GIS homomorphisms in Part II; (k) and (l) are discussed further below.

Theorem 3.1 (properties of cross-type transposition). In the following statements, (S_1, G, int_1), (S_2, G, int_2), and (S_3, G, int_3) are GISes sharing the same commutative interval group G; u and x are elements of S_1; v and y are elements of S_2; w is an element of S_3; $\mathrm{lab}_u\colon S_1 \to G$ and $\mathrm{lab}_v\colon S_2 \to G$ are the label functions associated with u and v; and $E_G : G \to G$ is the identity mapping on the group G.

(a) (T_u^v, E_G) is a GIS isomorphism from S_1 to S_2.
(b) $T_u^v(x) = y$ if and only if $\mathrm{int}_2(v, y) = \mathrm{int}_1(u, x)$.
(c) T_u^v is interval-preserving; that is, $\mathrm{int}_2(T_u^v(a), T_u^v(b)) = \mathrm{int}_1(a, b)$ for all a and b in S_1.
(d) If $S_1 = S_2$ and $\mathrm{int}_1 = \mathrm{int}_2$, then $T_u^v = T_i$, where $i = \mathrm{int}_1(u, v)$.
(e) T_u^u is the identity mapping on S_1.
(f) $T_u^v T_v^w = T_u^w$.
(g) $(T_u^v)^{-1} = T_v^u$.
(h) $T_u^v = T_x^y$ if and only if $T_u^v(x) = y$.
(i) For all a in S_1, $\mathrm{lab}_v(T_x^y(a)) = \mathrm{lab}_v(y) - \mathrm{lab}_u(x) + \mathrm{lab}_u(a)$.
(j) For all a and b in S_1, $\mathrm{int}_2(T_u^v(a), T_x^y(b)) = \mathrm{int}_1(a, u) + \mathrm{int}_2(v, y) + \mathrm{int}_1(x, b)$.
(k) For all a in S_1, $\mathrm{int}_2(T_u^v(a), T_x^y(a)) = \mathrm{int}_1(x, u) + \mathrm{int}_2(v, y)$.
(l) If $i = \mathrm{int}_1(u, x)$ and $j = \mathrm{int}_2(v, y)$, then $T_i T_x^y = T_u^v T_j$ and $T_x^y = (T_i)^{-1} T_u^v T_j$.

The result in part (k) follows immediately from (j) by setting $b = a$, rearranging, and combining the two int_1 terms. Notice that a does not appear to the right of the = sign in this formula: for any given u, v, x, and y, the interval from $T_u^v(a)$ to $T_x^y(a)$ is a constant, independent of a.

The last assertion in part (l) offers an explicit formula for changing the reference points in a transposition operator. The situation is illustrated in the transformation graph (not network!) of Example 7. The formula says that, starting in S_1 at the upper-left node of the graph, we can move to S_2 by applying T_x^y directly, or we can apply $(T_i)^{-1}$, T_u^v, and T_j in succession; the result will be the same either way. (Mathematicians say that "the diagram commutes"; in the terminology to be explored in Part V below, the graph is path-consistent.) To prove this formula, note that by part (d), the two single-GIS transpositions are $T_i = T_u^x$ (in S_1) and $T_j = T_v^y$ (in S_2), and therefore by part (f),

$$T_i T_x^y = T_u^x T_x^y = T_u^y = T_u^v\, T_v^y = T_u^v T_j.$$

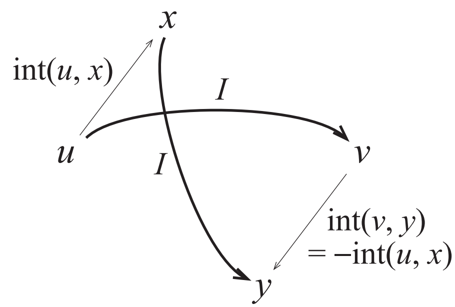
This is the first statement in part (l), and the second follows immediately, multiplying both sides of the above equation on the left by $(T_i)^{-1}$.

We proceed now to define the cross-type inversion operator I_u^v (or $I_{u\in S_1}^{v\in S_2}$). The discussion parallels that of cross-type transposition above. The general situation is depicted in Example 8. The intent here is that $I_u^v(x)$ should be the unique y whose interval from v (in S_2) is the same as the interval from u to x (in S_1), but in the opposite direction; that is, $\mathrm{int}_2(v,\ y) = -\mathrm{int}_1(u,\ x) = \mathrm{int}_1(x,\ u)$. As before, this condition is possible only if the two GISes share the same interval group. When the two GISes themselves are the same, this is precisely the same definition that Lewin gives for I_u^v, so our generalized notation is entirely consistent with his.[14]

To reframe this definition in terms of a GIS homomorphism, let $F_G\colon G \to G$ be the *inverse mapping* on the interval group G; that is, $F_G(i) = -i$ for all i in G. This F_G is a group isomorphism under the assumption that G is commutative. (If G is not commutative, then F_G is not a homomorphism but an anti-homomorphism; that is, $F_G(ij) = F_G(j)F_G(i)$. Because we are assuming commutativity, however, homomorphisms and anti-homomorphisms are the same thing.)

$$\begin{tikzcd}
S_1 \arrow[r, "T_x^y"] \arrow[d, dashed, bend left, "T_i^{-1}"] & S_2 \\
S_1 \arrow[u, "T_i = T_u^x"] \arrow[r, "T_u^v"'] & S_2 \arrow[u, "T_j = T_v^y"']
\end{tikzcd}$$

EXAMPLE 7. *Illustration of Theorem 3.1, part (l).*



EXAMPLE 8. *Interval relationships under inversion.*

The inversion I_u^v is then the induced GIS homomorphism $(\overline{F_G})_u^v$. Condition (b) in the definition of a GIS homomorphism is the equation $F_G(\mathrm{int}_1(u,\ x)) = \mathrm{int}_2(v,\ I_u^v(x))$, which is precisely the definition of I_u^v in the preceding paragraph.

The following theorem, parallel in many respects to Theorem 3.1, summarizes the properties of the inversions I_u^v, as well as some relationships with the transpositions T_u^v.

Theorem 3.2 (properties of cross-type inversion). In the following statements, $(S_1,\ G,\ \mathrm{int}_1)$, $(S_2,\ G,\ \mathrm{int}_2)$, and $(S_3,\ G, \mathrm{int}_3)$

14 Lewin 1987, 51. Example 8 is similar to Lewin's Figure 3.7.

are GISes sharing the same commutative interval group G; u and x are elements of S_1; v and y are elements of S_2; w is an element of S_3; lab_u: $S_1 \rightarrow G$ and lab_v: $S_2 \rightarrow G$ are the label functions associated with u and v; and F_G: $G \rightarrow G$ is the inverse mapping ($F_G(i) = -i$) on the interval group G.

(a) (I^v_u, F_G) is a GIS isomorphism from S_1 to S_2.
(b) $I^v_u(x) = y$ if and only if $\mathrm{int}_2(v, y) = -\mathrm{int}_1(u, x)$.
(c) I^v_u is interval-reversing; that is, $\mathrm{int}_2(I^v_u(a), I^v_u(b)) = -\mathrm{int}_1(a, b) = \mathrm{int}_1(b, a)$ for all a and b in S_1.
(d) $I^v_u I^w_v = T^w_u$.
(e) $I^v_u T^w_v = T^v_u I^w_v = I^w_u$.
(f) $(I^v_u)^{-1} = I^u_v$.
(g) $I^v_u = I^y_x$ if and only if $I^v_u(x) = y$.
(h) If $S_1 = S_2$ and $\mathrm{int}_1 = \mathrm{int}_2$, then $I^v_u = I^u_v$, and $(I^v_u)^2$ is the identity mapping on S_1.
(i) For all a in S_1, $\mathrm{lab}_v(I^y_x(a)) = \mathrm{lab}_v(y) + \mathrm{lab}_u(x) - \mathrm{lab}_u(a)$.
(j) For all a and b in S_1, $\mathrm{int}_2(I^v_u(a), I^y_x(b)) = \mathrm{int}_1(u, a) + \mathrm{int}_2(v, y) + \mathrm{int}_1(b, x)$.
(k) For all a in S_1, $\mathrm{int}_2(I^v_u(a), I^y_x(a)) = \mathrm{int}_1(u, x) + \mathrm{int}_2(v, y)$.
(l) If $i = \mathrm{int}_1(u, x)$ and $j = \mathrm{int}_2(v, y)$, then $T_i I^y_x = I^v_u T_j$ and $I^y_x = (T_i)^{-1} I^v_u T_j$.

The result in part (h), the self-inverting property of inversions within a single GIS, is familiar (Lewin 1987, 53, Corollary 3.5.5, recalling that we are assuming commutativity). In the context of the present theorem, (h) follows from (b) and (g), setting $x = v$ and $y = u$. The condition $S_1 = S_2$ is of course essential to this result; if S_1 and S_2 are different, then I^v_u and I^u_v are two very different mappings.

The change-of-reference-point formula in part (l) of Theorem 3.2 is exactly analogous to that for transpositions in part (l) of Theorem 3.1. The "modulations" T_i and T_j required in the formula remain transpositions, not inversions.

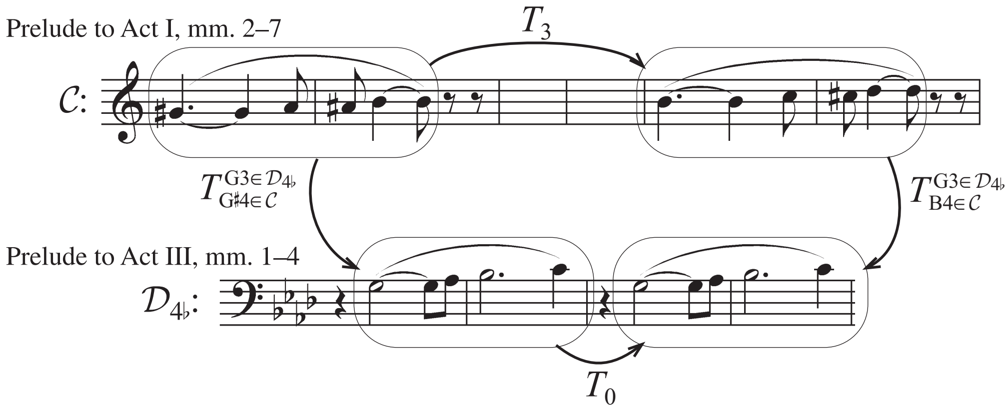
IV. INTERSCALAR TRANSPOSITION AND INVERSION

For present purposes we may define a *scale* as a GIS whose interval group is isomorphic to the group $\mathbb{Z}$ of integers. Two examples have surfaced already, chromatic pitch space $\mathcal{C}$ and diatonic pitch space $\mathcal{D}$, but in principle a scale can be any ordered succession of pitches, extending (theoretically) infinitely in both directions, with intervals measured by counting scale-steps. This construction is general enough to include not just chromatic and diatonic scales but a wide variety of other "modes" and microtonal scales as well.

Since all scales share the same interval group, cross-type transposition and inversion mappings from one scale to another are readily defined by the methods of Part III. In fact, the GIS homomorphism $\overline{(E_{\mathbb{Z}})}^{\mathrm{C4} \in \mathcal{D}}_{\mathrm{C4} \in \mathcal{C}}$ introduced in Part II is nothing more than the cross-type transposition $T^{\mathrm{C4} \in \mathcal{D}}_{\mathrm{C4} \in \mathcal{C}}$; analogous interscalar transpositions $T^{v \in S_2}_{u \in S_1}$ and inversions $I^{v \in S_2}_{u \in S_1}$ can be defined for any choice of scales S_1 and S_2 and reference points u and v. The examples in this section will illustrate a variety of applications of interscalar transpositions and inversions.

It has often been observed that the oboe line at the beginning of the Prelude to Act I of *Tristan und Isolde* is similar to the melody that opens the Prelude to Act III, except that the former is chromatic and the latter diatonic. As shown in Example 9, the precise relationship between the two may be modeled as a cross-type transposition. The Act I melody inhabits the GIS $\mathcal{C}$, chromatic pitch space, while the Act III melody inhabits a GIS we shall call $\mathcal{D}_{4\flat}$, four-flat diatonic pitch space.[15] The interscalar transposition $T^{\mathrm{G3} \in \mathcal{D}_{4\flat}}_{\mathrm{G\sharp4} \in \mathcal{C}}$ maps the initial pitch G♯4 of the chromatic motive to the initial G3 of the F-minor version, and maps other pitches, any number of chromatic scale-steps above or below G♯4, onto the pitches a corresponding number of F-minor scale-steps above or

15 The GIS structure does not recognize tonal centers, so $\mathcal{D}_{4\flat}$ could represent a scale that we are accustomed to calling A-flat major, F minor, or in principle any of five other modal alternatives. As analysts we readily identify the tonality of the present example as F minor, and the following discussion freely refers to the scale as such—but it should be acknowledged that such distinctions are an additional layer of interpretation, beyond the reach of GIS structure.



EXAMPLE 9. *Interscalar transposition in Wagner, Tristan und Isolde.*

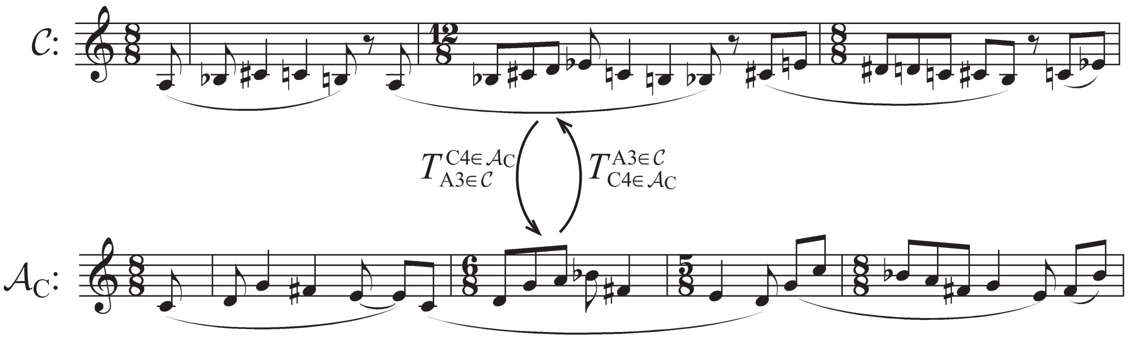
below G3.[16] This mapping completely describes the pitch relationship between the two four-note motives. Because the notes of the diatonic scale are more widely spaced than those of the chromatic, a chromatic-to-diatonic transposition may be called an *expansion*, while its inverse, a diatonic-to-chromatic mapping, is a *contraction*.

The second statement of the melodic fragment in the Act I Prelude is a chromatic T_3 transposition of the first statement. The second statement in Act III is identical to the first, indicated here by T_0 (strictly speaking, transposition by zero scale-steps in the F minor scale—but of course T_0 has no effect in any scale). Completing the network is the $T^{\mathrm{G3}\in\mathcal{D}_{4\flat}}_{\mathrm{B4}\in\mathcal{C}}$ relationship between the two second statements. The network, if it is to be path-consistent, must satisfy $T_3 T^{\mathrm{G3}\in\mathcal{D}_{4\flat}}_{\mathrm{B4}\in\mathcal{C}} = T^{\mathrm{G3}\in\mathcal{D}_{4\flat}}_{\mathrm{G\sharp4}\in\mathcal{C}} T_0$. Indeed, this equality follows from Theorem 3.1(l), setting u = G♯4 and x = B4 in $\mathcal{C}$, $v = y$ = G3 in $\mathcal{D}_{4\flat}$.

Example 10(a) models the well-known expansion of the fugue subject from Bartók's *Music for String Instruments, Percussion, and Celesta* as a transposition from chromatic pitch space to $\mathcal{A}_{\mathrm{C}}$, the so-called C acoustic scale (C-D-E-F♯-G-A-B♭).[17] This example illustrates the important point that the GISes involved in interscalar transposition and inversion are generally *pitch* spaces, not pitch-*class* spaces. The transformations do not respect octave equivalence: the inverse mapping shown here, the acoustic-to-chromatic contraction $T^{\mathrm{A3}\in\mathcal{C}}_{\mathrm{C4}\in\mathcal{A}_{\mathrm{C}}}$, maps the initial C4 to an A but later in the excerpt maps C5 to an E. Consequently such a transformation does not correspond to any mapping from 12-note chromatic pc-space to 7-note acoustic pc-space. These two pc-spaces have interval groups $\mathbb{Z}_{12}$ and $\mathbb{Z}_7$, respectively (every diatonic pc-space also has interval group $\mathbb{Z}_7$). There is no group homomorphism from $\mathbb{Z}_{12}$ to $\mathbb{Z}_7$ except for the trivial

16 In this and several subsequent figures in which cross-type transformational arrows point downward, the reader must be alert to the inverted relationship between the indices in the notation $T^{\mathrm{G3}\in\mathcal{D}_{4\flat}}_{\mathrm{G\sharp4}\in\mathcal{C}}$ and the layout of the network: the *lower* index (G♯4∈ $\mathcal{C}$) refers to a note in the *upper* portion of the diagram, and vice versa.

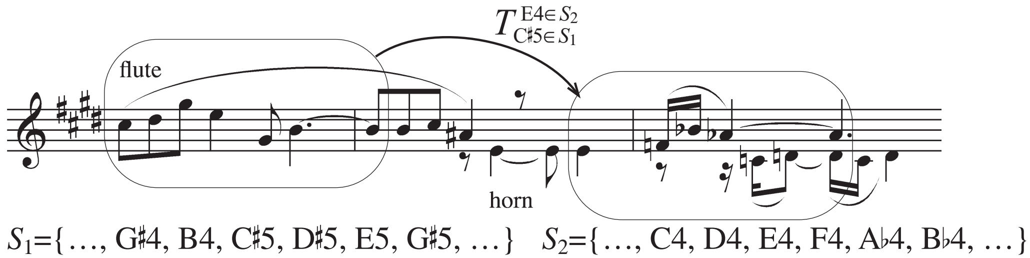
17 This transformation is described somewhat differently in Santa 1999, 206–207. Alpern 1998 considered several related examples.



(a) Bartók, Music for String Instruments, Percussion, and Celesta, i, mm. 1–3, and iv, mm. 204–07. (Copyright © 1937 by Boosey & Hawkes, Inc. for the USA. Copyright renewed. Reprinted by permission.)

(b) Dvořák, Piano Concerto in G Minor, Op. 33, i, mm. 41–47.

EXAMPLE 10. *Three examples of interscalar transposition.*



(c) Debussy, Prélude à "L'Après-midi d'un faune," mm. 3–5.

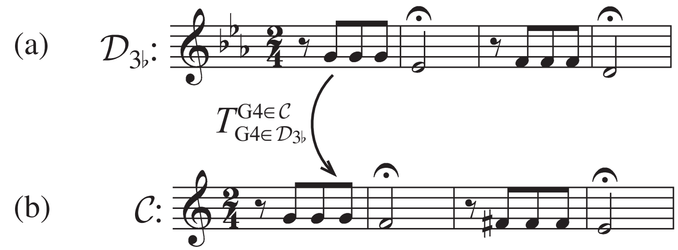
EXAMPLE 10. [*continued*]

one that maps every element to 0; it follows that there can be no nontrivial GIS homomorphism from chromatic pc-space to any diatonic pc-space. (Interscalar transposition and inversion can be defined as mappings on pc-spaces if the two scales have the same number of notes per octave, as in mappings from one diatonic scale to another, and in certain other special cases. Examples 12 and 13 below will illustrate.)

In the orchestral exposition of Dvořák's Piano Concerto (Example 10(b)), a simple diatonic motive is immediately followed by an octatonic transformation of itself—a very slight contraction—beginning on the same pitch. The GISes here are $\mathcal{D}_{4\flat}$ (here the A-flat major scale) and $\mathcal{O}_{C\sharp/D}$ (the octatonic scale containing the notes C-sharp and D). At the upcoming solo entrance, the soloist develops this motive into a pure octatonic scale spanning more than two octaves.

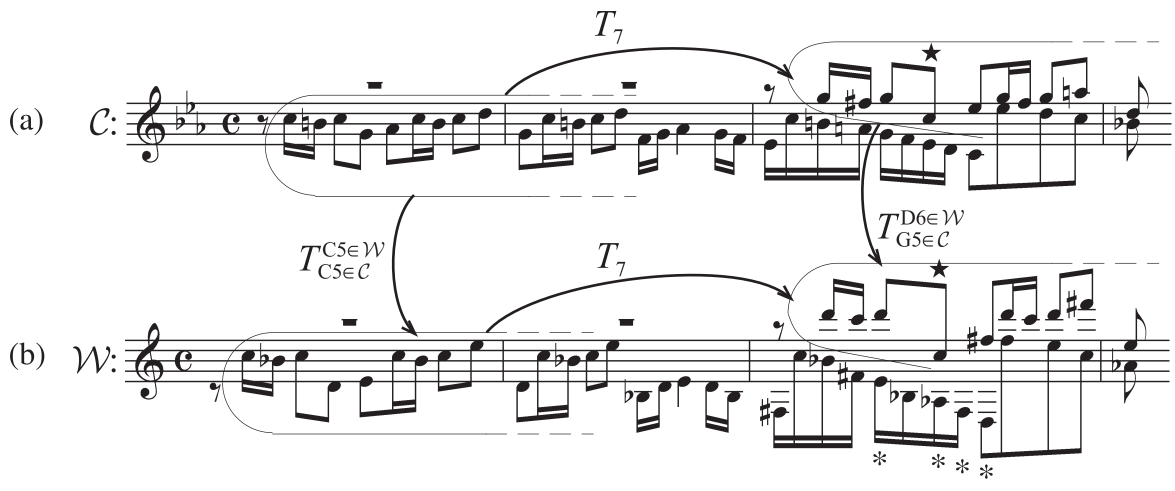
Any two contour-equivalent lines are related by transposition between suitably defined GISes. In Example 10(c), the flute line in m. 3 and the composite of the two horn parts in mm. 4–5 both have contours representable as ⟨235401⟩. The GISes S_1 and S_2 can then be any scales containing the appropriate segments of six pitches.

Interscalar transpositions are used to whimsical effect in two of Nicolas Slonimsky's *51 Minitudes* for piano. In the



EXAMPLE 11. *(a) Beethoven, Symphony No. 5; (b) Slonimsky, Minitude No. 1, "$\sqrt{B^5}$." (51 Minitudes for Piano by Nicolas Slonimsky, copyright © 1979 by G. Schirmer Inc. (ASCAP). International copyright secured. All rights reserved. Reprinted by permission.)*

first Minitude, subtitled "$\sqrt{B^5}$" ("the square root of Beethoven's Fifth"), shown in Example 11(b), Slonimsky subjects Beethoven's "fate" motive to a diatonic-to-chromatic contraction, so that interval ratios are approximately square-rooted. ("After a while," promises Slonimsky in his Preface, "you will like the square root better" [Slonimsky 1979].) Minitude No. 48 ("Bach × 2 = Debussy") treats Bach's C Minor Fugue to a chromatic-to-whole-tone expansion, resulting in the curiously impressionistic counterpoint shown



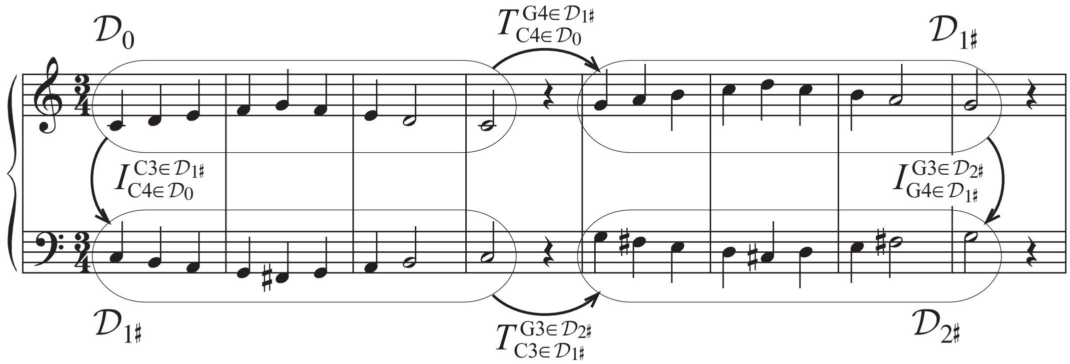
EXAMPLE 12. *(a) Bach, Well-Tempered Clavier, Book I, Fugue in C Minor; (b) Slonimsky, Minitude No. 48, "Bach × 2 = Debussy." (51 Minitudes for Piano by Nicolas Slonimsky, copyright © 1979 by G. Schirmer Inc. (ASCAP). International copyright secured. All rights reserved. Reprinted by permission.)*

in Example 12(b). (The descending passage in the left hand in m. 3 includes a few "wrong notes," asterisked in the example; this descent "should" arrive at C3, not D3.)[18]

The fugal subjects and answers of Example 12 provide several illustrations of the properties of GIS transpositions enumerated in Part III. Slonimsky's fugue, like Bach's, begins on C5, so the basic transformation is $T^{C5\in\mathcal{W}}_{C5\in\mathcal{C}}$ (here $\mathcal{W} = \mathcal{W}_C$ is the whole-tone scale containing C). Bach's answer in m. 3 is, as usual, the subject transposed by (chromatic) T_7, except for one altered note in the tonal answer (marked with a black star in the example). Slonimsky's answer is also T_7 of his subject, with the corresponding one-note alteration —but this is T_7 as measured in the whole-tone GIS $\mathcal{W}$. Thus while Bach's answer begins on G5, Slonimsky's begins on D6. The two answers are therefore related by $T^{D6\in\mathcal{W}}_{G5\in\mathcal{C}}$. One can verify by part (l) of Theorem 3.1 that $T^{C5\in\mathcal{W}}_{C5\in\mathcal{C}}T_7 = T_7T^{D6\in\mathcal{W}}_{G5\in\mathcal{C}}$, as the (path-consistent) network suggests. The equivalent change-of-reference-point formula is $T^{D6\in\mathcal{W}}_{G5\in\mathcal{C}} = (T_7)^{-1}T^{C5\in\mathcal{W}}_{C5\in\mathcal{C}}T_7$. But Slonimsky's entire fugue (excepting the "wrong notes") is $T^{C5\in\mathcal{W}}_{C5\in\mathcal{C}}$ of Bach's fugue, so it appears that the two cross-type transpositions $T^{C5\in\mathcal{W}}_{C5\in\mathcal{C}}$ and $T^{D6\in\mathcal{W}}_{G5\in\mathcal{C}}$ must be the same. Inasmuch as $T^{C5\in\mathcal{W}}_{C5\in\mathcal{C}}$ maps G5 to D6, Theorem 3.1(h) confirms that this is indeed the case.

This chromatic-to-whole-tone transformation, unlike the previous examples, could easily be defined as an ordinary single-type transformation within the chromatic scale. Slonimsky simply doubles Bach's intervals; cross-type analysis is not needed to describe this mapping. Also, uniquely among

18 Alpern observed that chromatic-to-whole-tone transposition, under the name of "geometric expansion," was advocated by Joseph Schillinger as a useful compositional technique (Schillinger [1941] 1978, 216). Interestingly, Schillinger chose several Bach fugue subjects for his illustrations, including the one later reshaped by Slonimsky.

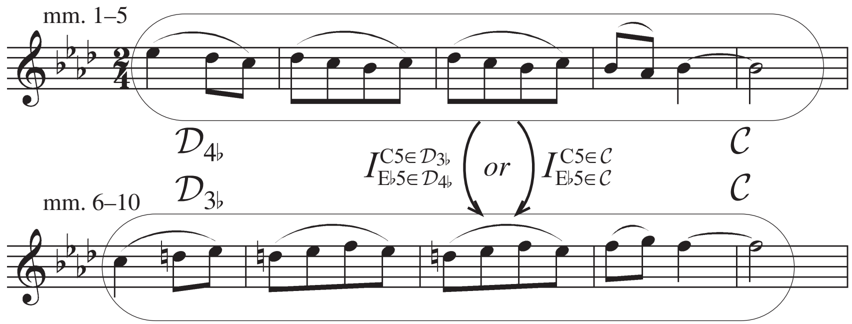


EXAMPLE 13. *Cross-type transpositions and inversions in Bartók, Mikrokosmos I, No. 17, "Contrary Motion," mm. 1–8. (Mikrokosmos copyright © 1940 by Hawkes & Sons (London) Ltd. Copyright renewed. Definitive corrected edition copyright © 1987 by Hawkes & Sons (London) Ltd. Reprinted by permission of Boosey & Hawkes, Inc.)*

the examples so far, this transformation respects octave equivalence and could therefore be defined as a mapping on pitch-*class* space. (Allowing for Slonimsky's "wrong notes," one can check that every G in Bach's fugue, regardless of register, maps to a D in Slonimsky's.) It is the nature of the relationship between the chromatic and whole-tone scales that makes these special properties possible. Because 12 is a multiple of 6, there is a natural homomorphism $g\colon \mathbb{Z}_{12} \to \mathbb{Z}_6$ that maps every number mod 12 to its congruence class mod 6. The mapping from chromatic pc-space $\mathcal{C}_{12}$ to whole-tone pc-space $\mathcal{W}_6$ exemplified by Slonimsky's fugue is then the induced GIS homomorphism $\overline{g}^{C\in\mathcal{W}_6}_{C\in\mathcal{C}_{12}}$. This mapping is not, strictly speaking, a cross-type transposition as defined in Part III (the two interval groups are different), nor is it an isomorphism (it is not one-to-one).

Example 13 provides a simple illustration of the interaction of interscalar transpositions and inversions. The right-hand and left-hand lines in mm. 1–4 are not strict inversions of each other in the chromatic universe, but they are related by interscalar inversion from $\mathcal{D}_0$ (C major) to $\mathcal{D}_{1\sharp}$ (C Lydian); in mm. 5–8 the same materials are transposed to $\mathcal{D}_{1\sharp}$ (G major) and $\mathcal{D}_{2\sharp}$ (G Lydian) respectively. All four statements are related by interscalar transpositions and inversions as shown. By Theorem 3.2(e), both of the composite mappings $T^{G4\in\mathcal{D}_{1\sharp}}_{C4\in\mathcal{D}_0} I^{G3\in\mathcal{D}_{2\sharp}}_{G4\in\mathcal{D}_{1\sharp}}$ and $I^{C3\in\mathcal{D}_{1\sharp}}_{C4\in\mathcal{D}_0} T^{G3\in\mathcal{D}_{2\sharp}}_{C3\in\mathcal{D}_{1\sharp}}$ are equal to $I^{G3\in\mathcal{D}_{2\sharp}}_{C4\in\mathcal{D}_0}$. Because all of the GISes here are diatonic, these transformations can be defined on pitch-class space.

Many familiar examples of melodic inversion in tonal music can be described in interscalar terms; networks similar to that in Example 13 can be constructed using original and inverted forms of some Bach fugue subjects, for example. In many such cases the relationships cannot adequately be described by single-GIS transformations, either chromatic or diatonic; this is because melodies are frequently inverted into a different scale, and inversions may start on virtually any scale degree. Occasionally, however, alternate single-GIS descriptions are possible. The second phrase in Example 14 is a C-minor diatonic inversion of the A-flat-major first phrase, but this relationship may also be described as a strict chromatic inversion. To see why this particular interscalar inversion should also operate chromatically, recall that a diatonic scale is symmetric about only one of its notes, the note



EXAMPLE 14. *Two interpretations of inversion in Brahms, Symphony No. 1, iii, mm. 1–10.*

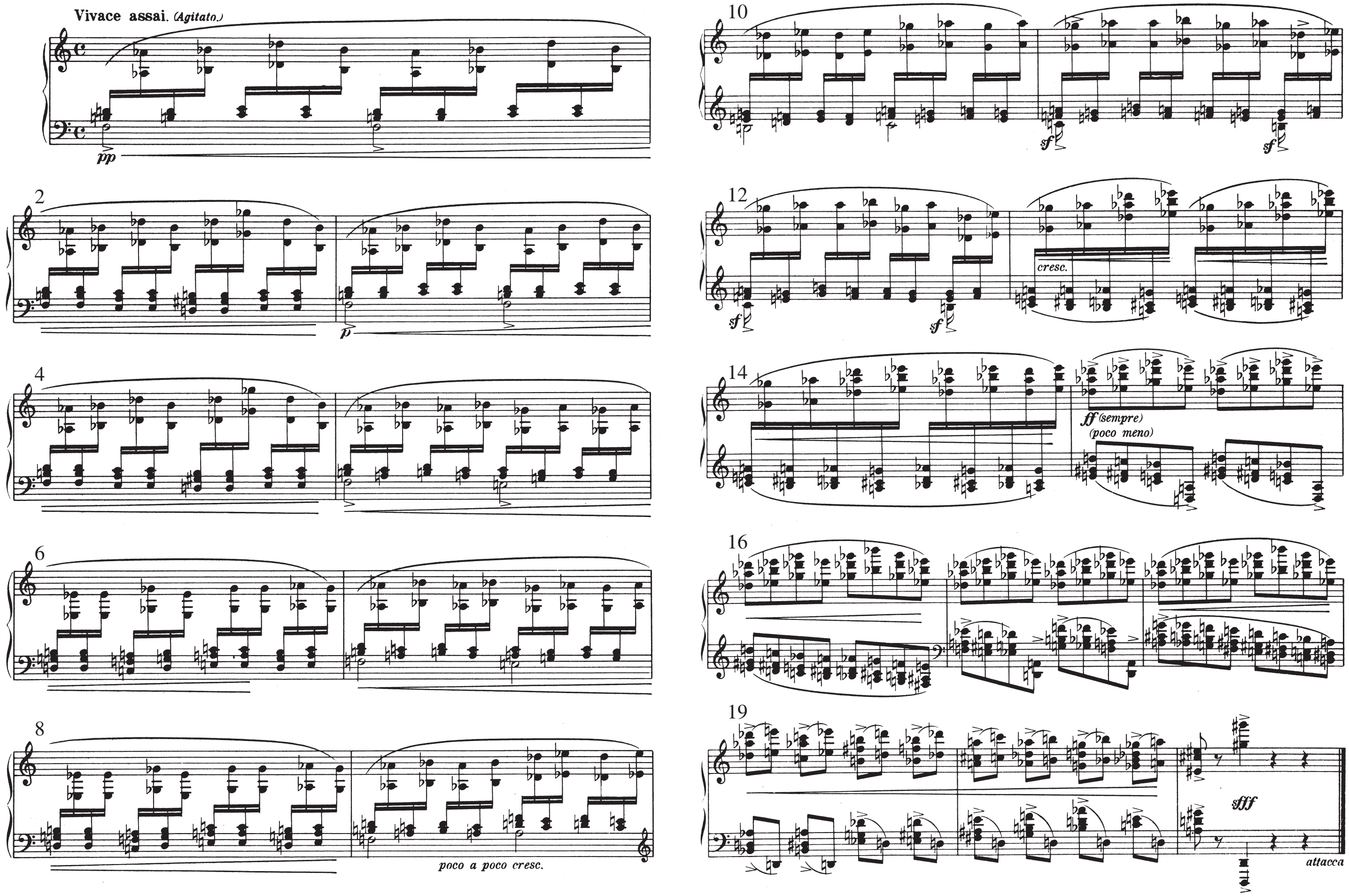
corresponding to scale degree $\hat{2}$ of a major scale. The inversion in question, $I^{C5\in\mathcal{D}_{3\flat}}_{E\flat5\in\mathcal{D}_{4\flat}}$, maps any B-flat ($\hat{2}$ of A-flat major) to an F ($\hat{2}$ of E-flat major). The chromatic coherence of this inversion therefore stems from the unique symmetry of the scales.

Example 15 presents the score of the third of Karol Szymanowski's Twelve Etudes, Op. 33, a collection written in 1916 and dedicated to Alfred Cortot. The transformation network in Example 16 covers the first twelve measures, in which the right hand plays exclusively on black keys and the left hand exclusively on white with the exception of the G-sharps in measures 2 and 4. In the network, then, the lower staff inhabits the white-key diatonic GIS $\mathcal{D}$, the upper staff inhabits the black-key pentatonic GIS $\mathcal{P}$, every vertical arrow is an interscalar transformation from $\mathcal{D}$ to $\mathcal{P}$, and the closed circuits formed by the arrows represent transformational identities, most of them described by Theorem 3.2(1). Because the GISes $\mathcal{D}$ and $\mathcal{P}$ are constant throughout, explicit reference to the GISes is omitted in the transformational labels for convenience.[19]

The transformational activity is most systematic in mm. 5–10, depicted in the second system of the network. The musical material here is represented by a series of beamed four-note cells. The notes in these cells are the top note of each octave in the right hand and the top note of each three-note chord in the left; the omitted voices follow along in predictable parallel motion except when the bass sustains longer notes. Every right-hand cell in this passage is an exact diatonic-to-pentatonic inversion of the corresponding left-hand cell (indicated by the vertical *I* arrows). Moreover, most of the cells are transpositions of each other within their respective scales (horizontal *T* arrows). The exceptions are in m. 6 and its repetition in m. 8, where the melodic contour changes. The dashed arrows here represent a slightly weaker sort of transposition, a transposition of pitch sets but not of melodic cells.[20] These relationships are fairly easy to hear if

19 Black note–white note oppositions of this kind appear in a number of piano works by Debussy. For two examples see the analyses of "Brouillards" in Parks 1980 and of "Feux d'artifice" in Lewin 1993, 97–159. Somewhat similar examples can be found in music by Ravel, Bartók, Ginastera, and other composers. A later example that responds well to analysis via interscalar transpositions is "Désordre," the first of György Ligeti's Etudes for piano.

20 The two right-hand cells in m. 6 are related by pentatonic T_1 but not by any transposition in the chromatic universe. These two different chromatic species of a single two-note pentatonic genus are a simple example



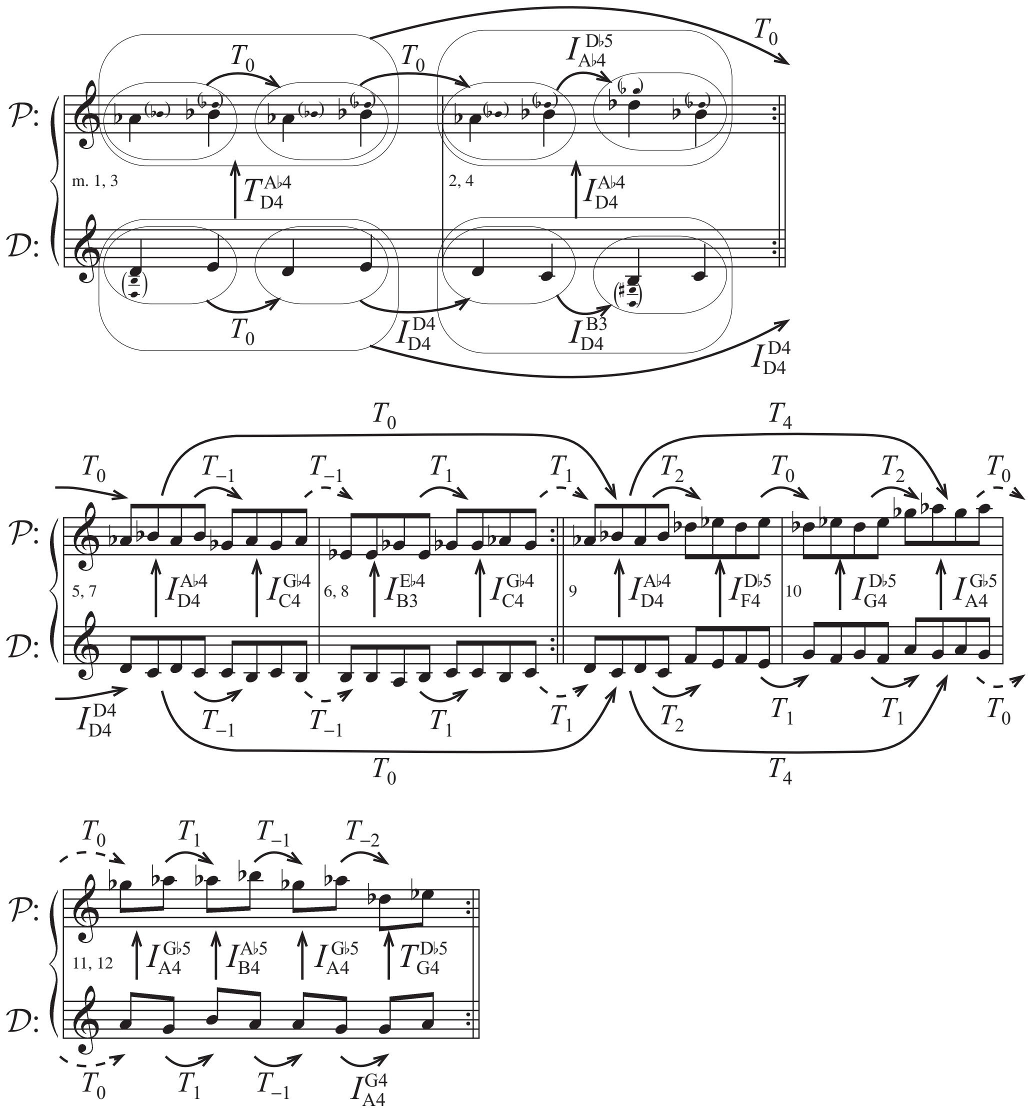
EXAMPLE 15. *Szymanowski, Etude, Op. 33, No. 3.*

one plays the second system of the network at a moderate tempo.

The pentatonic transposition intervals in the right hand of mm. 5–10 are usually, but not always, numerically equal to

---

of the principle "cardinality equals variety" in the 5-in-12 pentatonic setting; the concurrent left-hand cells illustrate the same principle in the more familiar diatonic case. For a general formulation of this property see Clough and Myerson 1985.



EXAMPLE 16. *Diatonic-to-pentatonic transpositions and inversions in Op. 33, No. 3, mm. 1–12.*

the corresponding diatonic transposition intervals in the left hand. The T_4 that spans from the first cell of m. 9 to the second cell of m. 10 appears in both hands, but with different factorizations: $T_2T_1T_1$ in the left hand, $T_2T_0T_2$ in the right. Of course, "T_4" means quite different things in the two scales: diatonic T_4 is transposition by fifth, while pentatonic T_4 is one step short of an octave.

In mm. 11–12, the last system of the network, four-note cells are no longer inversionally related, but two-note cells are. On the last beat, however, the contour of the left-hand cell is inverted to match that of the right; hence an inversion appears among the left-hand transformations and a transposition appears vertically.

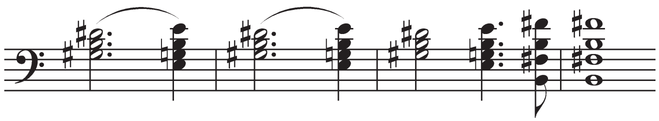
Vertical T relationships appear also at the beginning of the piece, but to hear them we must hear the music at the quarter-note level. That is, we collapse every two right-hand notes into a dyad, effectively slowing down the right-hand motion to match the harmonic rhythm of the left hand in the same passage. The large notes in the top staff in mm. 1–4 of Example 16 represent the lower voice in these dyads, and the vertical transformations relate this voice with the top voice in the left hand. This segmentation reveals the four-quarter-note cell in the right hand of m. 1 to be an augmentation of the four-eighth-note cell at the beginning of m. 5, and the four-quarter-note cell in the right hand of m. 2 to be an augmentation of the surface eighth-note contour of m. 1.

The upper voice in the right-hand dyads in mm. 1–4 generally remains at pentatonic T_1 above the lower, but in measures 2 and 4 it skips over the expected E♭5 to G♭5. This anomaly coincides exactly with the anomalous G-sharps in the left hand. Since neither E-flat nor G-natural has appeared in the piece yet, both anomalies could be accommodated by asserting that the governing scales in mm. 1–4 are not pentatonic and diatonic after all; rather, the right hand plays a "tetratonic" scale (pentatonic without E-flat) while the left plays a modified diatonic scale with G-sharp instead of G ("A harmonic minor"). At m. 5 the pure black- and white-key scales take over, and the arrows joining the first system of the network to the second then represent interscalar transformations in this interpretation.

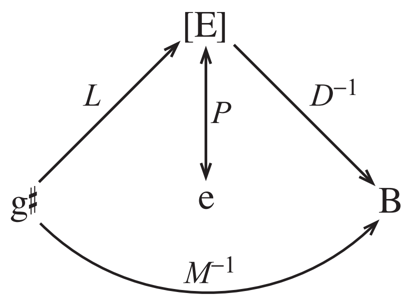
Also at m. 5, the basic unit of transformational action changes from the quarter note to the eighth note. As noted previously, there is another sort of level change at m. 11, from four-note groups to two-note groups. Both of these shifts are from larger units to smaller, and both contribute to the remarkable buildup of intensity in this half-minute composition. Other factors in this mounting tension include the steadily rising dynamic level, the expanding range, and the left hand's venture at m. 13 into chromatic territory, joined by the right six measures later. Several simultaneous developments at m. 15 should also be noted. There is a change in texture, with the hands playing together for the first time. There is a recapitulation of the opening right-hand material; the top line in mm. 15–18 is an exact pentatonic transposition of that in mm. 1–4, but not a chromatic transposition. The lowest notes in m. 15 initiate a pedal point on D that continues to the end of the piece and provides the only plausible suggestion of a tonal center. The concluding tritone in m. 21 marks the axis of symmetry of both the pentatonic and diatonic collections, and relates also to the endings of several of the other Etudes in the set.

V. THE PATH CONSISTENCY CONDITION

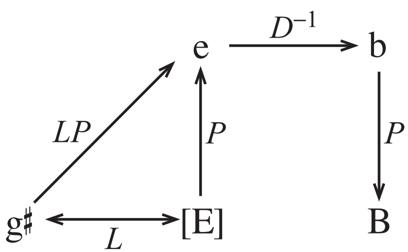
A 1992 article by David Lewin begins with a memorable opening gambit. Lewin reprints a transformation network from *Generalized Musical Intervals and Transformations*, then proceeds to describe everything that is wrong with it (Lewin 1987, 179, Figure 8.2a; revised in Lewin 1992, 52, Figure 3a; reprinted in Lewin 2006, 203, Example 11.4a). The network is based on the "Tarnhelm" motive from *Das Rheingold*, given in Example 17(a). The transformational notation in the network of Example 17(b) has been revised slightly in accord with current preferences and for consistency throughout this paper, but the analysis is Lewin's. The network uses the usual neo-Riemannian transformations L and P, and the "dominant"



(a) Wagner, Das Rheingold, "Tarnhelm" motive.



(b) Transformational analysis by David Lewin, 1987.



(c) Revised analysis by Lewin, 1992.

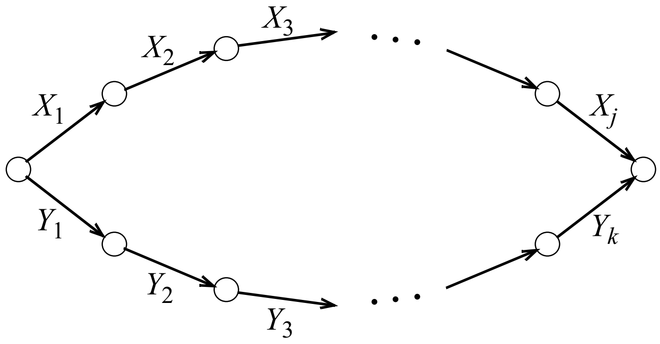
EXAMPLE 17.

and "diatonic mediant" transformations D and M (called DOM and MED by Lewin), whose behavior is illustrated by the actions C $\xrightarrow{D}$ F, c $\xrightarrow{D}$ f, C $\xrightarrow{M}$ a, c $\xrightarrow{M}$ A♭. The specific way in which D and M are defined necessitates that they appear here in inverse form.

Lewin's 1992 revision of this network appears in Example 17(c). Of the several differences between the two networks, the most significant are one addition and one omission: a B minor triad is added, while the mediant relation between the G-sharp minor and B major triads has vanished. The addition is appropriate, as the final sonority of the motive lacks a third; it may be heard as B minor in relation to the preceding E minor, or as B major via another application of P. The omission, however, entails that there is no longer any relationship shown between the initial and final chords of the passage. Lewin's principal justification for omitting this transformation is that its inclusion violates path consistency, which he regards as a requirement for a well-formed transformation network.[21] The path consistency condition is the subject of this final section, along with the question of its appropriateness as a universal requirement for graphs and networks. We shall return to Example 17 shortly, after a more detailed examination of the path consistency condition itself.

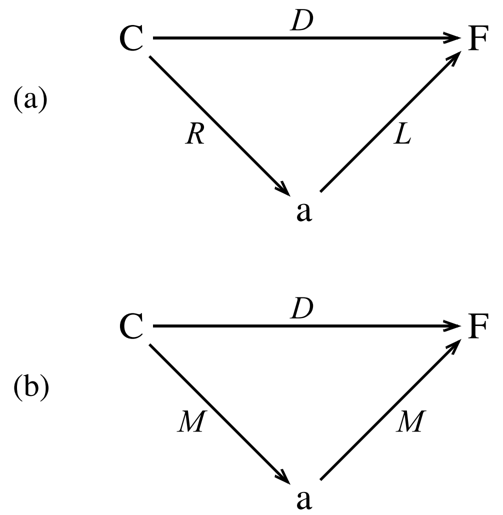
Lewin does not use the term *path consistency*; in his formulation this condition is part (D) in the definition of a transformation graph, where it is said to be needed "to ensure that no contradiction can possibly arise," and where it is illustrated with a diagram much like Example 18 (Lewin 1987, 195, Figure 9.2). The requirement is that when a transformation graph offers two alternative paths from one node to another, the products of the transformations along both paths must be identical. In the situation depicted in Example 18, the product $X_1X_2 \cdots X_j$ must equal $Y_1Y_2 \cdots Y_k$. (The numbers j and k, which enumerate the transformations

21 Path consistency is not the only impetus behind the revision. Lewin also sought to clarify a resemblance between the "Tarnhelm" graph and a graph depicting the "Valhalla" theme (not considered here).



$$X_1X_2 \cdots X_j = Y_1Y_2 \cdots Y_k$$

EXAMPLE 18. *The path consistency condition.*



EXAMPLE 19. *Two transformational analyses of the progression C–a–F.*

along the two paths, need not be the same.) In this paper, as noted in Part I above, we are extracting path consistency from the definition of a transformation graph, thereby obtaining a more lenient definition; we regard path consistency as an independent property that a graph may or may not satisfy.

The simple networks in Example 19 offer an instructive illustration. These networks depict two analyses of a descending-thirds progression C major–A minor–F major (possibly I-vi-IV in C major, or possibly a part of a longer pattern such as the extended sequence beginning at m. 143 of the Scherzo of Beethoven's Ninth Symphony).[22] In the partially Riemann-inspired reading of Example 19(a), C major and A minor are related by *R*, A minor and F major by *L*, and C is the dominant of F. Example 19(b) recasts both the *R* and *L* relations as diatonic mediants.

The network of Example 19(a) is not path-consistent. There are two paths from C to F, one labeled with the composite transformation *RL* and the other with the single transformation *D*, and the product *RL* is not equal to *D*. The situation is similar to that considered in regard to the Franck network of Example 2 in the introduction to this paper: when *RL* is applied to C major or to any other major triad, it yields the same result as *D*, but applied to any minor triad it does not. Applying the mediant transformation *M* twice in succession, on the other hand, *always* gives the same result as *D*, regardless of the mode of the starting triad; hence the transformations *MM* and *D* are identical as mathematical functions, and the network of Example 19(b) is path-consistent.

For Lewin, therefore, Example 19(b) is a valid transformation network, but 19(a) is not. In the rather different perspective adopted here, both networks, path-consistent or not, represent valid—if somewhat contrasting—analyses of the progression. When we hear a C major triad moving to A minor, we are surely entitled to describe the relationship as a

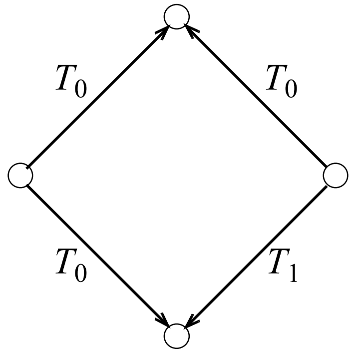
22 For a neo-Riemannian analysis of Beethoven's sequence as an *RL*-chain, see Cohn 1997, 36–37.

relative one if we wish, and likewise we may say that F major is the *leittonwechsel* of A minor. Then, considering the entire passage, we may observe that it starts and ends with dominant-related triads. Path consistency effectively forbids us from making these three observations simultaneously, not because of an actual contradiction in this network but because of a potential contradiction in another network with the same transformations, namely a mode-reversed one that does not actually reflect the music in question. Such a requirement imposes an unnecessary obstacle to the modeling of meaningful musical relationships. The limitation is felt especially severely in attempts to organize transformational structure in a hierarchy of "local" and "global," or "foreground" and "background" transformations: if transformational labels have been assigned on a local, foreground level in some musical passage under consideration (such as C $\xrightarrow{R}$ a $\xrightarrow{L}$ F in Example 19(a)), then path consistency makes it impossible to employ transformations of larger scope in the same network (such as C $\xrightarrow{D}$ F in the same example), except when those transformations are already logically implied by the local ones (with C $\xrightarrow{R}$ a $\xrightarrow{L}$ F in place, the only permissible transformation linking C to F would be RL).

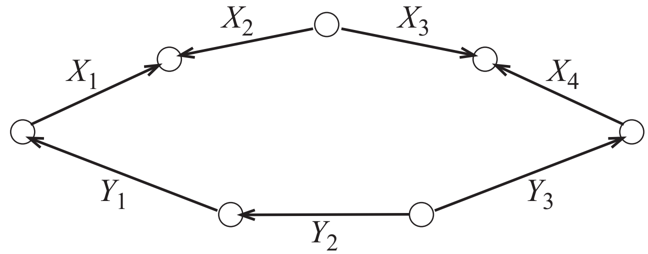
Precisely such a conundrum confronted Lewin in his "Tarnhelm" analysis. The original network (Example 17(b)) accurately depicts triadic relationships, but M^{-1} is not the same transformation as LD^{-1}, as one can check by reversing modes. Upon realizing this, Lewin concluded that the network was invalid and revised it as in Example 17(c). From our perspective, however, he perhaps grabbed himself too roughly by the shoulders: the omission of any mention of a relationship between the initial and final sonorities of the passage suggests that a significant part of our understanding of the music has been sacrificed in the name of path consistency. Example 17(c) is path-consistent, but in a quite trivial way: only one closed circuit is formed by the arrows, and the transformations involved are L and P on one path, LP on the other. Under the constraint of path consistency, the only permissible label for an arrow from G-sharp minor to B major would be $LPD^{-1}P$, which, even if reduced to the equivalent LD^{-1}, is a much less intuitive relationship than the mediant transformation that was there originally.

Apart from the "Tarnhelm" revision, Lewin gives no examples of the possible "contradictions" that path consistency is designed to avoid; indeed, he says very little about why he thinks the requirement is needed. Interestingly, if the intent is to avoid impossible situations of all kinds, path consistency arguably does not go far enough. The situation described in the graph of Example 20, for instance, is clearly impossible: the transformations are the identity T_0 and the semitone transposition T_1, and the three T_0 arrows demand that the same object must occupy all four nodes, a circumstance contradicted by the T_1 arrow. It may come as a surprise, therefore, to discover that this graph is path-consistent. The only way a graph can *fail* to be path-consistent is if it includes two directed paths from one node to another with different transformational products; this graph has no two directed paths from one node to another at all.

The above considerations suggest that for some purposes it might be appropriate to formulate a stronger version of the path consistency condition. *Strong path consistency* applies in situations like the one shown in Example 21: the arrows here form a closed circuit but are not oriented in two well-behaved directed paths from one node to another. As shown in the example, *strong path consistency* requires the transformational products along the two paths forming the circuit to be identical, with the understanding that when an arrow is traversed backward, the corresponding transformation in the product is replaced with its inverse. An equivalent, perhaps more elegant, way of formulating this definition is to require that the product of all the transformations (or their inverses where appropriate) along the closed path *around* such a circuit must equal the identity transformation on the space associated with the node at which the circuit begins and ends. In either formulation, this definition accounts for the difficulty posed by Example 20: although the graph is path-consistent, strong path consistency fails because $T_0T_0^{-1} \neq T_0T_1^{-1}$.



EXAMPLE 20. *An unrealizable but path-consistent transformation graph.*



$$X_1X_2^{-1}X_3X_4^{-1} = Y_1^{-1}Y_2^{-1}Y_3$$
$$\text{or } X_1X_2^{-1}X_3X_4^{-1}Y_3^{-1}Y_2Y_1 = \text{identity}$$

EXAMPLE 21. *The strong path consistency condition.*

A hidden assumption lurks in the above definition: replacing a transformation with its inverse requires that the transformation *have* an inverse. Strong path consistency, therefore, may be defined for a graph only if all of its transformations are invertible—that is, if every transformation is a one-to-one and onto mapping from the space of its starting node to the space of its ending node. We shall call such graphs *reversible* transformation graphs; in a reversible graph the direction of any arrow may be reversed, without loss of information, by inverting its transformational label. Graphs whose transformations belong to a group are always reversible. In fact, most graphs in the current transformational literature are reversible, even cross-type graphs such as those presented earlier in this paper. A cross-type graph cannot be reversible, however, if the spaces assigned to the nodes are of different cardinalities—for instance, in a harmonic analysis combining major and minor triads (a space of cardinality 24), major-minor and half-diminished seventh chords (cardinality 24), and fully-diminished seventh chords (cardinality 3). In light of this restriction, reversibility must be assumed in some parts of the following discussion, notably Theorem 5.1.

Two further definitions relate to the "impossible" nature of graphs such as Example 20. A transformation graph is *realizable* if some transformation network realizes it—that is, if some assignment of object labels to the nodes is consistent with the transformation labels. A transformation graph is *universally realizable* if, for any assignment of an object label to any one node, some assignment of object labels to the remaining nodes is consistent with the transformation labels. The latter definition, stronger than ordinary realizability, is formulated to account for situations such as those encountered previously in which a graph is well-behaved as long as a major triad occupies a certain node but fails if a minor triad is placed there instead.

We shall refer collectively to the four properties introduced above—path consistency, strong path consistency, realizability, and universal realizability—as *consistency attributes*, each of which a transformation graph may or may not satisfy. Example 22 presents five graphs exhibiting five different combinations of consistency attributes. The schizophrenic graph (a) clearly has none of the attributes,[23] while

23 Astute readers may observe that the graph of Example 22(a), like that of Example 20, could be realized by assigning the complete twelve-note

graph (e) (which underlies the network of Example 19(b)) has all of them. The other three graphs exhibit the attributes in various intermediate combinations. Graph (b) (from Example 20), as already noted, is path-consistent but not strongly path-consistent, and is unrealizable. Graph (c) is realizable (Example 19(a) realizes it) but not universally realizable and, as noted previously, not path-consistent. Finally, graph (d) is both path-consistent (for the same reason as (b)) and realizable (assign C major triads to the left and right nodes and A minor to the top and bottom), but does not satisfy the stronger version of either attribute.[24]

The four consistency attributes are of two distinct types. Path consistency and strong path consistency are *syntactic* properties, constraints on the structure of a transformation graph. Realizability and universal realizability, in contrast, are *semantic* properties, as they involve not the *structure* of a graph but its *meaning*: the question of whether it does or does not model any real musical situation. Remarkably, the stronger of the two syntactic attributes and the stronger of the two semantic attributes turn out to be logically equivalent, as the following theorem shows. (A proof is outlined in the Appendix.)

Theorem 5.1 (Completeness Theorem for transformation graphs).[25] A reversible transformation graph is strongly path-consistent if and only if it is universally realizable.

Clearly strong path consistency implies path consistency and universal realizability implies realizability, so the logical relationships among the four consistency attributes may be summarized in the following diagram:

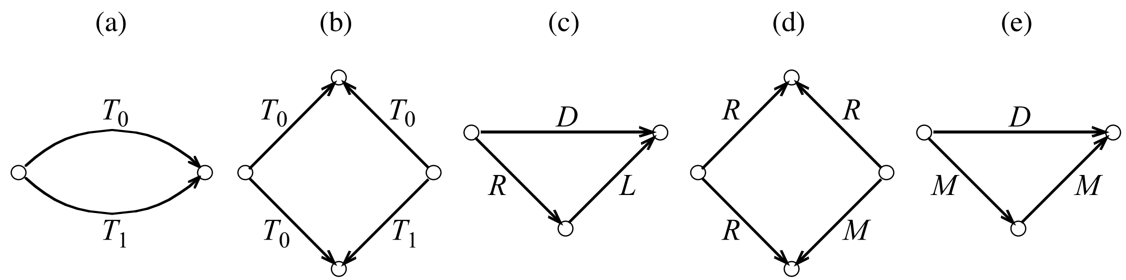
$$\begin{array}{ccc} \text{strongly} & & \text{universally} \\ \text{path-consistent} & \Leftrightarrow & \text{realizable} \\ \Downarrow & & \Downarrow \\ \text{path-consistent} & & \text{realizable} \end{array}$$

From the implications shown, it can be deduced that the five combinations of attributes displayed in Example 22 are the only combinations possible. Every (reversible) transformation graph, in other words, is of one of five *consistency types*: (a) no attributes; (b) path-consistent only; (c) realizable only; (d) path-consistent and realizable only; (e) all four attributes.

aggregate to each of its nodes, for that set is its own T_1-transform as well as its own T_0-transform. It may be recalled from the definitions in Part I that the *space* assigned to each node of a transformation graph, which describes the allowable contents of the node, must be specified in defining the graph. The consistency attributes satisfied by a graph may depend on the space(s) over which the graph is defined. In Example 22, the properties are tabulated accurately assuming that the graphs are defined over the space of major and minor triads (which of course does not contain the aggregate among its objects).

24 John Rahn (2004, 142) has pointed out another limitation hidden in Lewin's formulation of transformation graphs. Lewin's definition of directed graphs, which he calls "node/arrow systems" (Lewin 1987, 193, Definition 9.1.1), identifies arrows with ordered pairs of nodes, thus precluding the possibility of two or more arrows from one node to another. In the graph theory literature, such configurations are sometimes permitted and sometimes forbidden. Harary (1969, 10), for example, distinguishes *graphs* (which have no multiple edges or arrows) from *multigraphs* (which may have some); in this terminology Example 22(a) is a multigraph. It is not difficult to devise an unrealizable graph even if multiple arrows are disallowed, but neither is it difficult to formalize the theory in a way that accommodates them. Multiple arrows are not required in any of the subsequent examples in this paper, but it is worth noting that in conjunction with the relaxation of the path consistency requirement, multiple arrows can sometimes offer a musically suggestive way to model alternative interpretations simultaneously.

25 Theorem 5.1 is called the "Completeness Theorem" by analogy with certain results in mathematical logic asserting the equivalence of syntactic and semantic concepts. For instance, Gödel's Completeness Theorem for First-Order Languages asserts that a set of statements (*axioms*) within a certain kind of formal system is *consistent* (does not give rise to logical contradictions) if and only if it has a *model* (a mathematical structure in which interpretations are assigned to the symbols of the system and all the axioms are true). See, for example, Mendelson 1997, 84–92.



	(a)	(b)	(c)	(d)	(e)
path-consistent?	No	Yes	No	Yes	Yes
strongly path-consistent?	No	No	No	No	Yes
realizable?	No	No	Yes	Yes	Yes
universally realizable?	No	No	No	No	Yes

EXAMPLE 22. *Five transformation graphs and their consistency attributes.*

Some other characteristics of consistency attributes are more straightforward than the equivalence given in the Completeness Theorem. For example, all four attributes are *preserved under isography*; that is, if two transformation graphs are isomorphic they must be of the same consistency type. All four attributes are *hereditary*; that is, if a graph is path-consistent, then all of its subgraphs are path-consistent, and likewise for the other three attributes. (It is possible, however, that a subgraph may be of a stronger consistency type than a larger graph of which it is a part. For instance, a type (b) graph may have a type (e) subgraph.)

In many transformational analyses in the literature, the objects all inhabit the same space S and the transformations are selected from a *simply transitive* transformation group G acting on S. Simple transitivity means that for any two objects u and v in S, there exists one and only one transformation X in G such that $X(u) = v$. Familiar examples of simply transitive groups include the Riemannian *PLR* group (acting on triads) and the *TI* group of pitch-class set theory (acting on any set class containing 24 different pc-sets). In the case of simply transitive groups the assortment of consistency types is simplified:

Theorem 5.2 (realizable graphs with simply transitive groups). If a transformation graph is realizable and if all its transformations belong to a group that acts on a single space in simply transitive fashion, then the graph satisfies all four consistency attributes.

For graphs with simply transitive groups, it follows that consistency types (a), (b), and (e) are the only possibilities. Indeed, in Example 22, it can be seen that the transformations appearing in graphs (c) and (d) are not drawn from simply transitive groups: R and L belong to the Riemannian group, but neither D nor M does. The mediant transformation M generates its own simply transitive group, and inasmuch as $D = M^2$ the transformations in graph (e) are drawn from this group.[26]

It is tempting to assume that an unrealizable graph such as graph (a) in Example 22 could not possibly be musically useful. Example 23 may suggest otherwise. This figure, to be read like a flowchart, depicts the key areas in a Classical sonata form. The first time the piece moves from the primary theme to the secondary theme, there is a modulation by T_7; the last time there is not. This diagram is not strictly a transformation graph, as its spaces and transformations are not precisely defined, but it perhaps suggests a way to incorporate graphical elements that are technically unrealizable into musically plausible structures.

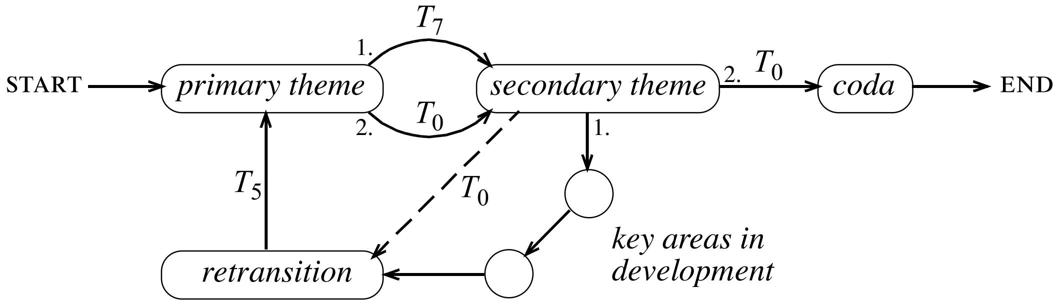
As a concluding illustration, we offer a cross-type, non-path-consistent analysis of two passages from Rimsky-Korsakov's *Christmas Eve*. The Prelude to the opera opens with the series of root-position major triads shown at the top of Example 24. What the example does not show is the remarkable orchestration: every chord is scored for a different combination of winds. The concept of "voice leading" is foreign to this music, for only rarely does one player play a note in as many as two consecutive chords. The chords are projected as a series of isolated, frozen jewels, an apt evocation of the moonlit, snow-covered landscape on which the curtain is about to rise.

The analysis of this passage, shown in the upper portion of the transformation network in the example, begins with the observation that if every other triad were minor instead of major, the result would be an RL-chain (like the Beethoven sequence mentioned in conjunction with Example 19). This hypothetical RL-chain works its way across the top of the network, where the non-path-consistent RLD triangle from Example 19(a) repeats over and over. Rimsky's music, however, follows not the familiar, locally diatonic pattern of this RL-chain but the zigzag path of solid arrows, in which all chords are major. As in the RL-chain, chord roots move alternately by minor and major thirds, and alternate chords are dominant-related, but the exclusive use of major triads implies that each diatonic fifth relation is broken into two *chromatic* third relations. The major triads on C-sharp, F-sharp, and B may therefore be heard either as parallel (P) substitutions for the corresponding minor triads in the RL-chain or as pitch-class transpositions (T_9) of the immediately preceding E, A, and D chords.[27] The network shows both interpretations; the resulting RPT_9 and PLT_8 triangles, it may be noted, are also non-path-consistent.

Christmas Eve, composed in 1894–95, is a charming blend of comedy and fantasy, an adaptation of a story by Gogol in which the village blacksmith Vakula wins his bride by flying to Petersburg on the Devil's back in order to procure for her the Empress Catherine's slippers. The scene of Vakula's flight is divided into several fantastic episodes, among them the Procession of Comets, which opens as shown at the bottom of Example 24. Apart from an identical tempo indication, this excerpt bears little resemblance to the opening of the Prelude. The sonorities are seventh chords, played throughout by a quartet of horns doubled by other winds; the parsimonious voice leading has the result that the chords cycle systematically through all four possible inversions. The restlessness imparted by these unstable sonorities and inversions is intensified by the violin trills, doubled in flute and piccolo, with highlights in harp and celesta. The pattern continues for another nine measures beyond the

26 Lewin introduces simple transitivity and the group generated by M (Lewin 1987, 157, 179–80). For a detailed study of simply transitive groups of triadic transformations see Hook 2002b, 84–93.

27 In Rimsky-Korsakov's own harmony text, written almost a decade before *Christmas Eve*, direct juxtaposition of third-related triads of the same mode falls into the category called "false progressions"; no examples are given. See Rimsky-Korsakov [1886] 1930, 108.



EXAMPLE 23. *Flowchart-style graph depicting key areas in a Classical sonata form.*

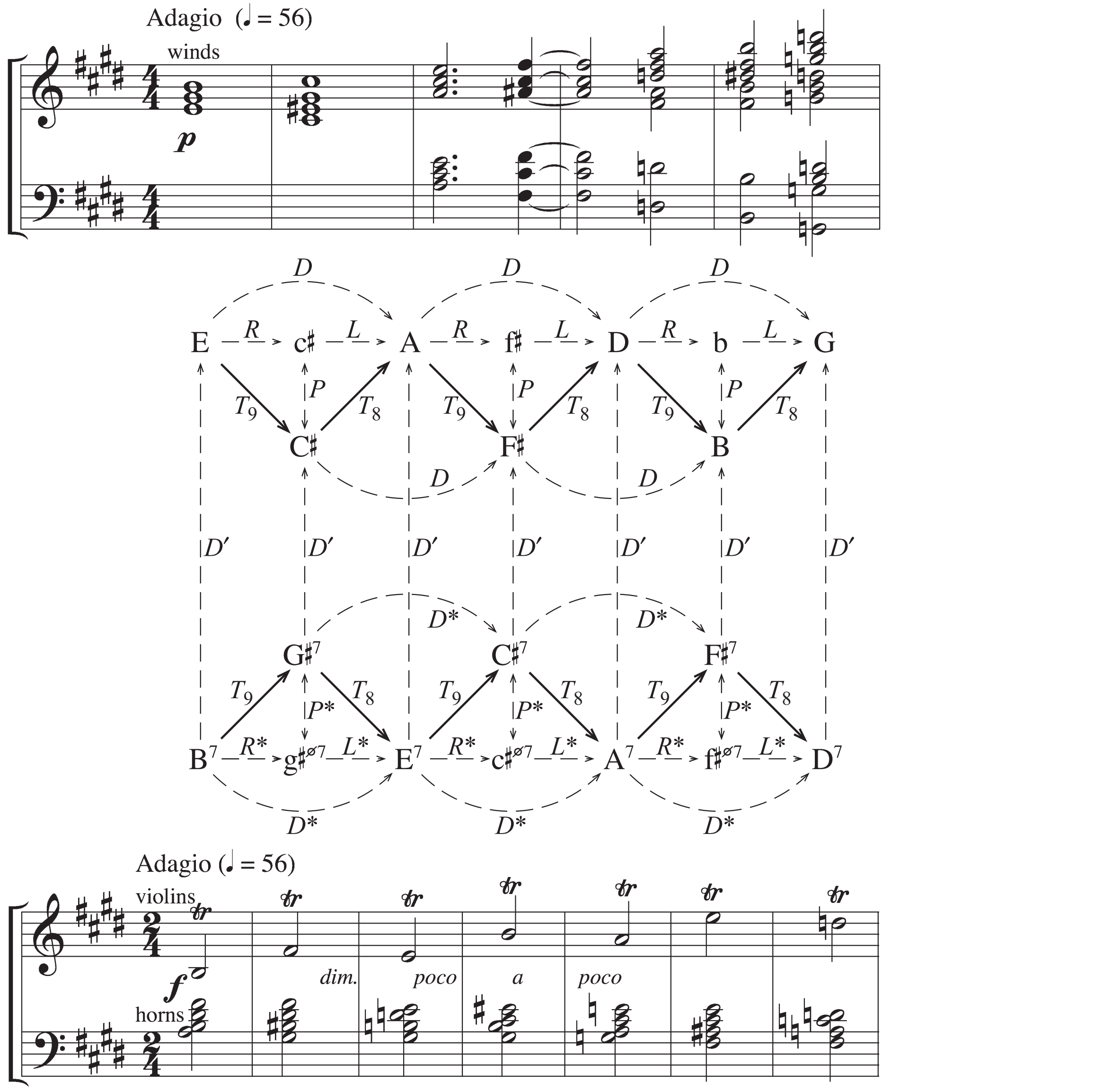
seven shown in the example, giving the trills ample time to continue their inexorable ascent into the realm of the comets. Indeed, the musical depiction of flight is as unmistakable as was the crystalline stillness of the Prelude.

It is remarkable, then, to find precisely the same transformational processes at work in both excerpts, for the root-interval relations among the major-minor seventh chords in the Procession of Comets are identical to those among the major triads in the Prelude. These relations are depicted in the lower portion of the network, using a system of transformations (P^*, L^*, R^*, D^*) for major-minor and half-diminished seventh chords that is algebraically isomorphic to the more familiar system of triadic transformations (P, L, R, D) in the upper portion. For example, just as R relates B major and G-sharp minor triads, R^* relates B^7 and $G\sharp^{ø7}$ chords.[28] Of course, the transpositions T_9 and T_8 may be applied to seventh chords as well.

Regarding the upper and lower portions of the network in Example 24 as distinct networks, we may observe that the two are isographic. The isography is a particularly appealing one, as every chord in the Procession of Comets is the dominant seventh of the corresponding triad in the Prelude. The transformation D' appearing on the vertical arrows joining the two parts of the network is a cross-type counterpart to the dominant transformation D; that is, D' combines the action of D with a change of type, in this case from seventh chords to triads. The relation between D and D' is thus similar to the relation between L and L' in the omnibus analysis with which this study began.

There are limits to the musical inferences that should be drawn from this formal isomorphism. The transformations R and R^* relate corresponding chords in the two analyses and play identical structural roles in their respective transformation groups, but it does not follow that they share other, non-algebraic properties. Triads related by R, for instance, belong to a common diatonic scale, while seventh chords related by R^* do not. The two transformations differ also in their voice-leading behavior: R moves one voice by whole step while R^* moves two voices by semitone. (Notwithstanding the close association of algebraic formalism and voice-leading parsimony in the recent neo-Riemannian literature, this comparative analysis demonstrates that voice-leading properties of transformations are *not* intrinsic to their algebraic structure.) In part for these reasons, it is arguably less natural to hear the seventh-chord progression as

28 See Hook 2002b, 111–13, for remarks on the direct application of triadic transformations to the set class of major-minor and half-diminished seventh chords, and for connections with the systems of Childs and Gollin.



EXAMPLE 24. *Transformational relationships in Rimsky-Korsakov, Christmas Eve: Act I, Prelude (top); Act III, Procession of Comets (bottom).*

an altered R^*L^*-chain than it is to hear the triadic one as an altered RL-chain.

Rimsky-Korsakov seems to have been quite taken with these transformational patterns. The opening chords of the Prelude recur several times during the opera. Once the seventh-chord version has been introduced, it quickly becomes a motto for Vakula's flight, and forms the basis for one section of the *khorovod* that follows the Procession of Comets. The *csárdás* that comes next begins with a newly animated statement of the triadic progression in which the subtle rhythmic acceleration of the Prelude chords, all but imperceptible originally, gains new urgency. This is followed immediately by the seventh-chord variant, then by a hybrid version in which the top voice follows the ascending pattern of the Prelude chords but the harmonies are all seventh chords in third inversion.[29]

29 Writing late in his life about the composition of *Christmas Eve*, Rimsky-Korsakov regretted that his free incorporation of mythological elements into Gogol's simple tale had been a mistake—"but a mistake which offered the opportunity of writing a wealth of interesting music" (Rimsky-Korsakov [1909] 1935, 291). Both the Prelude and the Procession of Comets appear among the examples in Rimsky's posthumously published treatise on orchestration. They are widely separated in the book, and there is no discussion of a connection between them or of the harmonic progressions involved (Rimsky-Korsakov [1913] 1964, Examples 106 and 143).

APPENDIX: PROOFS

Proof of Theorem 2.1 (GIS Homomorphism Theorem). Suppose S_1, S_2, u, v, and g are as in the statement of the theorem, and take any x in S_1. We show first that there is exactly one possible definition of $f(x)$ in S_2 if (f, g) is to be a GIS homomorphism.

Let $\mathrm{int}_1(u, x) = i$ (an element of G_1), and let $g(i) = j$ (an element of G_2). If f satisfies condition (b) in the definition of a GIS homomorphism, then $j = g(\mathrm{int}_1(u, x)) = \mathrm{int}_2(f(u), f(x)) = \mathrm{int}_2(v, f(x))$. But by condition (b) in the definition of a GIS, there is exactly one element y of S_2 such that $\mathrm{int}_2(v, y) = j$; therefore $f(x)$ *must* be this y.

We have shown that only one function f can possibly fulfill the requirements of the theorem; we must still show that it actually does—that is, that if the function f is defined as in the preceding paragraph, then (f, g) truly is a GIS homomorphism. Condition (a) is automatic, since g is a group homomorphism by hypothesis. Condition (b) must be verified for arbitrary x and y in S_1:

$$\begin{aligned}
&g(\mathrm{int}_1(x, y)) \\
&\quad = g(\mathrm{int}_1(x, u) + \mathrm{int}_1(u, y)) && \textit{(property of any GIS)} \\
&\quad = g(-\mathrm{int}_1(u, x) + \mathrm{int}_1(u, y)) && \textit{(property of any GIS)} \\
&\quad = -g(\mathrm{int}_1(u, x)) + g(\mathrm{int}_1(u, y)) && \textit{(g is a group homomorphism)} \\
&\quad = -\mathrm{int}_2(v, f(x)) + \mathrm{int}_2(v, f(y)) && \textit{(definition of f)} \\
&\quad = \mathrm{int}_2(f(x), v) + \mathrm{int}_2(v, f(y)) && \textit{(property of any GIS)} \\
&\quad = \mathrm{int}_2(f(x), f(y)). && \textit{(property of any GIS)}
\end{aligned}$$

For the last sentence of the theorem, if (f, g) is a GIS isomorphism, then g is a group isomorphism by definition. Conversely, suppose g is a group isomorphism; we must show that the function f defined above is one-to-one and onto. Take y in S_2; we must show that there is exactly one x in S_1 such that $f(x) = y$. Let $\mathrm{int}_2(v, y) = j$. Because g is a group isomorphism, there is exactly one i in G_1 such that $g(i) = j$. Because S_1 is a GIS, it then follows that there is exactly one x in S_1 such that $\mathrm{int}_1(u, x) = i$. This is the unique x satisfying $f(x) = y$.

Proof of Theorems 2.5 (label functions for induced GIS homomorphisms) and 2.6 (intervals formed by induced GIS homomorphisms). In the situation of Theorem 2.5,

$$\begin{aligned}
&\mathrm{lab}_v(\bar{g}^y_x(a)) \\
&\quad = \mathrm{int}_2(v, \bar{g}^y_x(a)) && \textit{(definition of } \mathrm{lab}_v\textit{)} \\
&\quad = \mathrm{int}_2(v, y) + \mathrm{int}_2(y, \bar{g}^y_x(a)) && \textit{(property of any GIS)} \\
&\quad = \mathrm{int}_2(v, y) && \textit{(definition of } \bar{g}^y_x\textit{)} \\
&\qquad + \mathrm{int}_2(\bar{g}^y_x(x), \bar{g}^y_x(a)) \\
&\quad = \mathrm{int}_2(v, y) + g(\mathrm{int}_1(x, a)) && \textit{((}\bar{g}^y_x, g\textit{) is a GIS homomorphism)} \\
&\quad = \mathrm{int}_2(v, y) + g(\mathrm{int}_1(x, u) && \textit{(property of any GIS)} \\
&\qquad + \mathrm{int}_1(u, a)) \\
&\quad = \mathrm{int}_2(v, y) + g(-\mathrm{int}_1(u, x) && \textit{(property of any GIS)} \\
&\qquad + \mathrm{int}_1(u, a)) \\
&\quad = \mathrm{int}_2(v, y) - g(\mathrm{int}_1(u, x)) && \textit{(g is a group homomorphism)} \\
&\qquad + g(\mathrm{int}_1(u, a)) \\
&\quad = \mathrm{lab}_v(y) - g(\mathrm{lab}_u(x)) && \textit{(definitions of } \mathrm{lab}_v \textit{ and } \mathrm{lab}_u\textit{)} \\
&\qquad + g(\mathrm{lab}_u(a)),
\end{aligned}$$

as asserted. Then, for Theorem 2.6,

$$\mathrm{int}_2(\bar{g}^v_u(a), \bar{h}^y_x(b)) = \mathrm{lab}_v(\bar{h}^y_x(b)) - \mathrm{lab}_v(\bar{g}^v_u(a))$$

by a general property of the label function lab_v. Applying the result of Theorem 2.5 to both terms separately, this is

$$\begin{gathered}
[\mathrm{lab}_v(y) - h(\mathrm{lab}_u(x)) + h(\mathrm{lab}_u(b))] - [\mathrm{lab}_v(v) - g(\mathrm{lab}_u(u)) + g(\mathrm{lab}_u(a))] \\
= \mathrm{lab}_v(y) - h(\mathrm{lab}_u(x)) + h(\mathrm{lab}_u(b)) - \mathrm{lab}_v(v) + g(\mathrm{lab}_u(u)) - g(\mathrm{lab}_u(a)).
\end{gathered}$$

Of these six terms, $-g(\mathrm{lab}_u(a))$ is $-g(\mathrm{int}_1(u, a))$ or equivalently $g(\mathrm{int}_1(a, u))$, and $\mathrm{lab}_v(y)$ is $\mathrm{int}_2(v, y)$. The two terms $-h(\mathrm{lab}_u(x))$ and $h(\mathrm{lab}_u(b))$ are $h(\mathrm{int}_1(x, u))$ and $h(\mathrm{int}_1(u, b))$, which combine (because h is a group homomorphism) to give $h(\mathrm{int}_1(x, u) + \mathrm{int}_1(u, b))$, or $h(\mathrm{int}_1(x, b))$. The two remaining terms contribute nothing because $\mathrm{lab}_v(v)$ and $\mathrm{lab}_u(u)$ are the identity elements of their respective groups and g is a group homomorphism. The sum therefore reduces to

$$g(\mathrm{int}_1(a, u)) + \mathrm{int}_2(v, y) + h(\mathrm{int}_1(x, b)),$$

as the theorem states. (The rearrangement of the terms here appears to depend on commutativity of the groups, but the careful reader may check that the theorem gives the correct order of terms in the general case, though of course multiplicative notation is preferable in the absence of commutativity.)

Proof of Theorem 5.1 (Completeness Theorem for transformation graphs). Suppose first that a graph is universally realizable. To prove strong path consistency, suppose that a circuit like that in Example 21 occurs within the graph. Let the products of the transformations along the two paths forming the circuit (substituting inverses as appropriate) be denoted X and Y. We must show that $X = Y$; that is, we must show that $X(u) = Y(u)$ for every u in the space associated with the starting node of the two paths (corresponding to the leftmost node in Example 21). For every u, by universal realizability, some realization of the graph assigns u to the starting node. Because the node labels in that realization are necessarily consistent with the transformational labels, the contents of the ending node of the two paths must simultaneously be both $X(u)$ and $Y(u)$; hence $X(u) = Y(u)$, as desired.

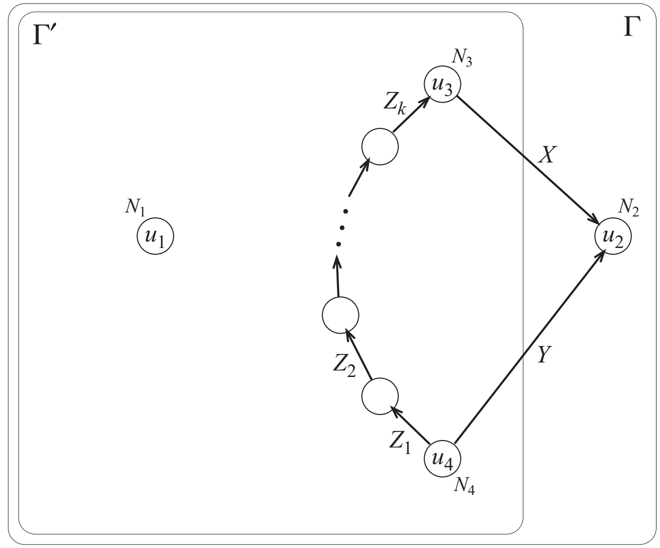
The converse—that strong path consistency implies universal realizability—is more difficult. Even for graphs with a single node, this implication hinges on a subtlety: if a graph is strongly path-consistent, every loop (arrow from a node to itself) must be labeled with the appropriate identity transformation.[30] (This follows immediately from the second for-

30 Some related considerations are explored briefly in Lewin 1987, 195–96, section 9.2.2. An idiosyncratic feature of Lewin's formalization of node/arrow systems (193, definition 9.1.1) is that every node is *assumed* to be in the arrow relation with itself; that is, though not drawn explicitly, loops are understood to be present on every node of every graph. The reasons for this unusual stipulation are not explained, but it may have been intended as a way of strengthening the path consistency condition: the implicit presence of loops, together with path consistency, ensures that the product of the transformational labels along a

mulation of the definition of strong path consistency, because a loop is the simplest kind of circuit. It follows also from the first formulation, by noting that a loop has the same starting and ending nodes as the path obtained by traversing the same loop twice in succession.) In a strongly path-consistent graph with only one node, therefore, the only possible arrows are loops labeled with the identity; clearly any labeling of the node of such a graph automatically realizes the graph, and thus every such graph is universally realizable.

We proceed by contradiction, beginning with the assumption that some strongly path-consistent graph is *not* universally realizable. Choose such a graph Γ with the smallest possible number of nodes. If there are m nodes in this graph, then $m \geq 2$ (by the preceding paragraph), and every strongly path-consistent graph with $m - 1$ or fewer nodes is universally realizable. The graph Γ must be connected, because otherwise its connected components, being smaller graphs, would be universally realizable and it would follow that Γ was universally realizable as well.

Because Γ is not universally realizable, there must exist some node N_1 and some element u_1 in the associated space S_1 with the property that no realization of Γ assigns the label u_1 to the node N_1. Choose another node N_2 of Γ whose removal does not disconnect the graph (every connected graph with two or more nodes has at least two nodes with this property), and let Γ′ be the connected graph obtained from Γ by deleting node N_2 and its associated arrows. (The schematic diagram in Example 25 may be helpful.) Because Γ′ has only $m - 1$ nodes, it is universally realizable, so some network Δ′ realizes Γ′ and assigns the label u_1 to N_1. Because Γ is connected, N_2 must be joined by an arrow to some node N_3 of Γ′, and this arrow has some transformational label X in Γ. (It is possible that N_3 and N_1 are the same node; this is of no consequence. Also, by reversibility we may assume that



EXAMPLE 25. *Illustration of the proof of the Completeness Theorem for transformation graphs.*

the arrow is directed from N_3 to N_2 as shown in the diagram; if not we replace X with X^{-1} in the following argument.)

The network Δ′ includes labels for all the nodes of the network Γ′; we extend this node labeling to a labeling Δ of all the nodes in Γ by assigning the label u_2 to N_2, where $u_2 = X(u_3)$. By hypothesis, Δ cannot be a proper realization of Γ—that is, it cannot be a valid transformation network. How can this be? The graph Γ is well-formed, so the problem must involve node labels. The network Δ′ realizes Γ′, so the node labels within Γ′ are consistent with the transformational labels; hence the problem must involve the label u_2. This label was chosen to be consistent with u_3 and X. Because Γ is strongly path-consistent, a loop on the node N_2, if there is one, must be labeled with the identity transformation. The only way Δ can fail, therefore, is if N_2 is joined by another arrow (in Γ) to some other node N_4 of Γ′, and $Y(u_4) \neq u_2$,

closed circuit of arrows, starting and ending at the same node, must equal the label on the loop at that node.

where u_4 is the label assigned to N_4 and Y is the transformational label on this arrow (replacing Y with Y^{-1} if necessary).

Because Γ' is connected, there exists within Γ' some chain of arrows linking N_4 to N_3. Let the transformational labels on these arrows, in order from N_4 to N_3, be $Z_1, \ldots, Z_k$. (We assume the arrows are all oriented in the same direction, but once more, by reversibility, any or all of the transformations may be inverted if necessary.) By label consistency within the network Δ', the product $Z_1 \cdots Z_k$ of these transformations, applied to the element u_4, yields u_3. Because $u_2 = X(u_3)$, it follows that the composite transformation $Z_1 \cdots Z_k X$, applied to u_4, yields u_2. But the large graph Γ is strongly path-consistent, so $Z_1 \cdots Z_k X$ must be identical to Y. It follows that $Y(u_4) = u_2$ after all, contrary to the conclusion of the preceding paragraph. This contradiction completes the proof of the theorem.

Proof of Theorem 5.2 (realizable graphs with simply transitive groups). Because the transformations belong to a group G, the graph is reversible and Theorem 5.1 applies; it therefore suffices to prove strong path consistency. Suppose a circuit occurs in the graph. Let the transformational products along the two paths forming the circuit (with inverses as appropriate) be X and Y; we wish to show that $X = Y$. Because the graph is realizable, some assignment of node labels is consistent with all the transformational labels in the circuit. Let the node label on the starting node of the two paths be u, and let the node label on the ending node be v. Then $X(u) = Y(u) = v$. Because X and Y are products of elements of the group G, X and Y themselves belong to G. Because G is simply transitive, it follows that $X = Y$.

REFERENCES

Alpern, Wayne. 1998. "Bartók's Compositional Process: 'Extension in Range' as a Progressive Contour Transformation." Presented at the annual meeting of Music Theory Midwest, Louisville.

Callender, Clifton. 1998. "Voice-Leading Parsimony in the Music of Alexander Scriabin." *Journal of Music Theory* 42.2: 219–33.

Capuzzo, Guy. 2004. "Neo-Riemannian Theory and the Analysis of Pop-Rock Music." *Music Theory Spectrum* 26.2: 177–99.

Childs, Adrian P. 1998. "Moving Beyond Neo-Riemannian Triads: Exploring a Transformational Model for Seventh Chords." *Journal of Music Theory* 42.2: 181–93.

Clough, John and Gerald Myerson. 1985. "Variety and Multiplicity in Diatonic Systems." *Journal of Music Theory* 29.2: 249–70.

Cohn, Richard. 1997. "Neo-Riemannian Operations, Parsimonious Trichords, and Their *Tonnetz* Representations." *Journal of Music Theory* 41.1: 1–66.

Cook, Robert. Forthcoming. "Parsimony and Extravagance." *Journal of Music Theory*.

Douthett, Jack and Peter Steinbach. 1998. "Parsimonious Graphs: A Study in Parsimony, Contextual Transformations, and Modes of Limited Transposition." *Journal of Music Theory* 42.2: 241–63.

Dummit, David S. and Richard M. Foote. 2004. *Abstract Algebra*, 3rd ed. Hoboken, NJ: John Wiley & Sons.

Gollin, Edward. 1998. "Some Aspects of Three-Dimensional *Tonnetze*." *Journal of Music Theory* 42.2: 195–206.

Harary, Frank. 1969. *Graph Theory*. Reading, Mass.: Addison-Wesley.

Hook, Julian. 2002a. "Uniform Triadic Transformations." Ph.D. dissertation, Indiana University.

———. 2002b. "Uniform Triadic Transformations." *Journal of Music Theory* 46.1–2: 57–126.

Kolman, Oren. 2004. "Transfer Principles for Generalized Interval Systems." *Perspectives of New Music* 42.1: 150–190.

Lewin, David. 1987. *Generalized Musical Intervals and Transformations.* New Haven: Yale University Press.

———. 1992. "Some Notes on Analyzing Wagner: *The Ring* and *Parsifal*." *19th-Century Music* 16.1: 49–58.

———. 1993. *Musical Form and Transformation: Four Analytic Essays.* New Haven: Yale University Press.

———. 2006. *Studies in Music with Text*. Oxford: Oxford University Press.

Mead, Andrew. 1994. *An Introduction to the Music of Milton Babbitt*. Princeton: Princeton University Press.

Mendelson, Elliott. 1997. *Introduction to Mathematical Logic*, 4th ed. London: Chapman and Hall.

Parks, Richard. 1980. "Pitch Organization in Debussy: Unordered Sets in 'Brouillards.' " *Music Theory Spectrum* 2: 119–34.

Rahn, John. 2004. "The Swerve and the Flow: Music's Relationship to Mathematics." *Perspectives of New Music* 42.1: 130–48.

Rimsky-Korsakov, Nikolai. [1886] 1930. *Practical Manual of Harmony*. Trans. Joseph Achron. New York: Carl Fisher.

———. [1909] 1935. *My Musical Life*. Trans. Judith A. Joffe. New York: Tudor.

———. [1913] 1964. *Principles of Orchestration*. Trans. Edward Agate. New York: Dover.

Santa, Matthew. 1999. "Defining Modular Transformations." *Music Theory Spectrum* 21.2: 200–29.

Schillinger, Joseph. [1941] 1978. *The Schillinger System of Musical Composition.* New York: Da Capo Press.

Slonimsky, Nicolas. 1979. *51 Minitudes for piano: 1972–76*. New York: Schirmer.

Soderberg, Stephen. 1998. "White Note Fantasy." *Music Theory Online* 4.3.

Straus, Joseph. 2003. "Uniformity, Balance, and Smoothness in Atonal Voice Leading." *Music Theory Spectrum* 25.2: 305–52.

Tymoczko, Dmitri. 2005. "Voice Leadings as Generalized Key Signatures." *Music Theory Online* 11.4.

Zbikowski, Lawrence. 2002. *Conceptualizing Music: Cognitive Structure, Theory, and Analysis.* Oxford: Oxford University Press.

Music Theory Spectrum, Vol. 29, Issue 1, pp. 1–40, ISSN 0195-6167, electronic ISSN 1533-8339. DOI: 10.1525/mts.2007.29.1.1