C2

CHAPTER 2

MUSIC ANALYSIS AS THE COGNITION OF SIMILARITY RELATIONSHIPS

The Music of Charles Mingus as a Case Study

MORGAN PATRICK AND DANIEL SHANAHAN

C2P1 At the 1960 Antibes Jazz Festival, Charles Mingus bookended his performance with two very similar songs: “Wednesday Night Prayer Meeting” from the 1959 album *Blues and Roots* and “Better Get It in Your Soul” from the 1959 album *Mingus Ah Um*. By beginning and ending the concert with these tunes, Mingus provided the audience with a sort of frame, making explicit the textures, forms, and melodic ideas shared between these two pieces, which are often considered to be siblings—both inspired by, and meant to evoke, participatory church services.

C2P2 Both pieces begin with the same opening bass gesture, providing the audience at Antibes with a cue that this was an intentional refrain. Both pieces are based on an F blues progression, both have six pulses per bar ($\frac{6}{8}$ and $\frac{6}{4}$), and both begin with other instruments gradually entering before the main melody chimes in over the bass gesture. The structure of the arrangement is also quite similar, with the solos following a similar order. Table 2.1 illustrates the arrangement of both pieces in their recorded versions. At Antibes, the opening sax solos were replaced by trumpet solos, and the trombone solos were replaced by sax solos. The fact that they were rearranged for live performance in an almost identical way is itself quite interesting. Additionally, Mingus’ frequent vocal interjections (“Oh Lord!,” etc.) pervade both pieces. These similarities make sense when one considers that they share more than a generic church topos: they are both meant to evoke a prayer meeting and were released only a year apart.

C2P3 At the same time, the pieces are formally quite different. The structure of “Prayer Meeting” is a twelve-bar blues. By contrast, while “In Your Soul” uses the twelve-bar blues’ harmonic progression over the solos, the actual tune of the piece (the “head”) consists of a ten-bar A section (structured like a twelve-bar blues, but with the typical V–IV of the ninth and

C2T1

Table 2.1 Comparing the arrangements and layouts of Charles Mingus' "Better Get It in Your Soul" and "Wednesday Night Prayer Meeting."

"Better Get It in Your Soul"								
Intro (bass and piano)	Head (10-bar A; 10-bar A; 8-bar B; 10-bar A)	Sax solo	Piano ostinato with trombone pedal	Trombone solo with collective improvisation	Sax solo; handclap accompaniment	Sax solo continues with rhythm section	Drum solo: "trading 8s"	Head (10-bar A; 10-bar A; 8-bar B; 10-bar A)
"Wednesday Night Prayer Meeting"								
Intro (bass and piano)	Head (12-bar blues)	Sax solo	Piano ostinato with trombone pedal	Sax solo, with added saxophone layers	Sax solo; handclap accompaniment	Sax solo continues with collective improvisation and rhythm section	Drum solo: "trading 8s"	Head (12-bar blues)

tenth measures omitted), followed by an eight-bar B section. Viewed in terms of the formal organization of their melodic material, one could argue that these pieces are quite different.

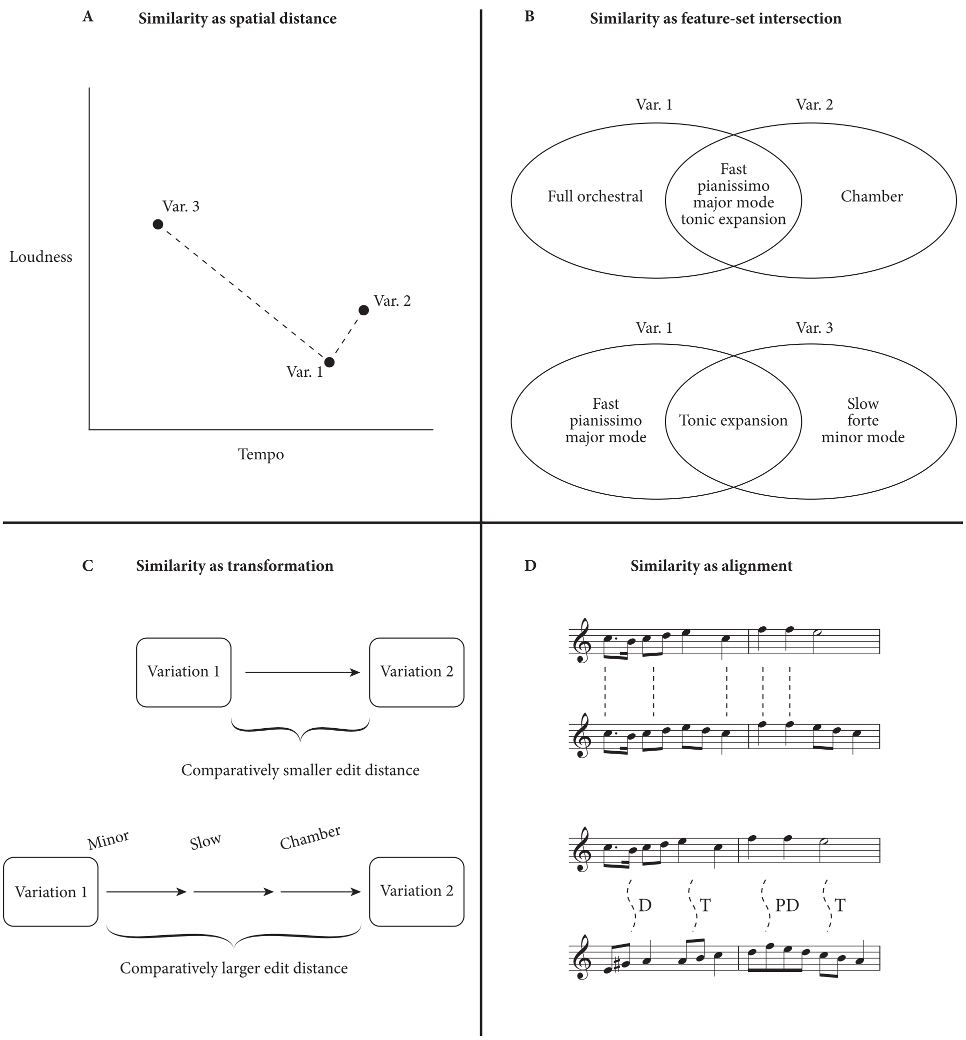
C2P4 How we evaluate musical similarity depends a great deal on how we define the objects under the microscope—that is, how we "represent" them. For example, if we look at the lead sheets for both of these Mingus pieces, they appear fairly different: again, one is a twelve-bar blues, the other a fairly standard A–A–B–A song structure. If our object of study, however, is not the lead sheet but the arrangement, these pieces are quite similar. This points to a well-known complexity in the psychology of similarity: similarity depends upon the format of the objects being compared. Because our minds do not have direct access to music as it sounds, we must convert the sensory signal into mental representations of the musical surface. Those forms, and the features of the music they privilege, vary depending on the task at hand.[1] Note that both Mingus pieces have nine vowels in their title, contain piano accompaniment, and feature three beats of rest in their respective melodies more than once. Obviously, it would be difficult to argue that these are meaningful similarities in any analytical context. A primary task of music theorists is to constrain the universe of possible formats and features by which to gauge similarity, and to construct taxonomies that regulate similarity according to internally consistent principles.[2]

C2P5 Another source of complexity stems from how we *evaluate* similarity. What are we doing when we judge two musical objects to be similar? Are we measuring musical dimensions that are continuous, and observing how "close" objects are to one another in the resulting space? Or are we instead comparing discrete features of the two pieces to understand their overlap? Perhaps instead we are transforming one into the other, which involves a more general cognitive process of mapping corresponding elements of two musical structures. While these might seem to be minor distinctions, the difference between them is conceptually quite meaningful. In fact, psychologists are quick to observe that how we measure similarity greatly constrains what we perceive to be similar (Medin, Goldstone, and Gentner 1993).

C2P6 In sum, both the "what" and the "how" of similarity are intimately tied to context and the act of comparison. For this reason, psychologists often reject the notion that similarity is an objective property shared between two objects, instead recognizing that it depends heavily on context and on the individual measuring it.[3]

C2P7 How do we evaluate similarity? Hahn (2014) identifies four major cognitive approaches: spatial distance, feature set intersection, transformation, and alignment.[4] Figure 2.1 provides an overview of these. To measure similarity as *spatial distance* is akin to looking at two addresses and assessing how close they are to one another on a map (see Figure 2.1A). For example, consider two hypothetical variations of a theme that are nearly matched in tempo (perhaps 120 bpm and 122 bpm) and loudness (80 db and 100 db, respectively). Along comes a third variation that is significantly slower (80 bpm) and softer (50 db). Since the first two variations are closer together in our parameter space, we would judge them to be more like one another than either is to the third, nicely reflecting the contextual nature of similarity: while an absolute distance may be objective, we use a larger context to evaluate "how" similar two things seem to one another.

C2P8 To measure similarity as *feature set intersection* is to consider individual features that are shared and distinct between two musical objects. A feature is defined as a particular musical "attribute," such as tempo, and the particular "value" it takes on. The more attributes with shared values between two objects, the greater the similarity (see Figure 2.1B). For example,



C2F1 **FIGURE 2.1** Four approaches to musical similarity: (A) represents similarity as spatial distance, with three variations being mapped in terms of tempo and loudness. (B) represents similarity as a collection of shared and unique features. (C) represents similarity as transformational distance, or how many edits are required to turn one variation into another. (D) represents similarity as the quality of alignment, with analogous timespans matching in terms of surface structure (straight lines) and/or higher order relations (curved lines).

tonally related variations might differ on the basis of tempo (e.g., fast or slow), mode (e.g., major or minor), and instrumentation (e.g., orchestral or chamber). A fast, major, and orchestral variation (Variation 1) would be more similar to a fast, major, and chamber variation (Variation 2) than it would be to a slow, minor, and chamber variation (Variation 3): Variations 1 and 2 share attributes of two features, whereas Variations 1 and 3 share none, except, perhaps, an underlying harmonic structure.

C2P9 To measure similarity as *transformation* is to ask: "How much effort does it take to turn one variation into another?" The more effort required, the less similar the variations. For instance, in the previous example, it is considerably easier to turn Variation 1 into Variation 2 (only one "edit" is needed: instrumentation) than to turn Variation 1 into Variation 3 (three edits are needed: tempo, mode, and instrumentation; see Figure 2.1C). As we will see, the use of "edit" distance is quite common in music theoretic approaches to similarity, especially in corpus studies.

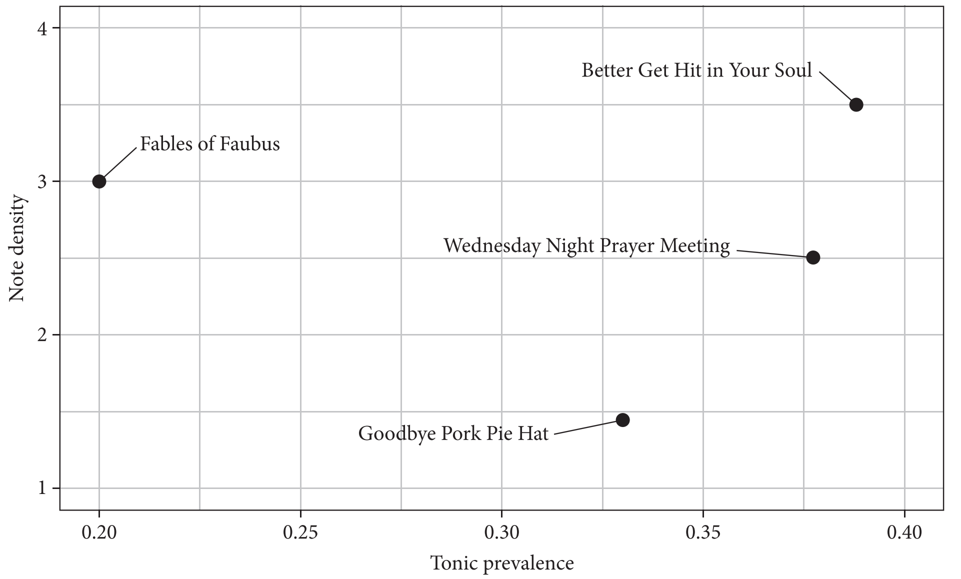
C2P10 Finally, to measure similarity as *structural alignment* is to judge how systematically the parts of one variation map onto the parts of another (see Figure 2.1D).[5] The mapping can involve merely surface features (e.g., absolute pitches and durations), higher-order relational features (e.g., intervallic relationships or tonal functions), both, or neither. Similarity increases when surface similarity is sustained by deeper, more relational connections.

C2P11 The methods of measuring similarity in the cognitive sciences occupy a significant place in music theory and analysis. Whether implicitly or explicitly, music theorists regularly engage with these four modes of comparison when constructing their own analytical frameworks. In this chapter, we provide an overview of these psychological approaches, drawing on much of the work already done on the topic in music cognition, music theory, and music information retrieval. We hope to elucidate how principles of human cognition constrain (and inform) what it means for two musical entities—namely, motives, variations, and themes—to be similar. Here, we have chosen to view these approaches to similarity through the lens of the music of Charles Mingus, but they could easily be generalized to any repertoire.

C2S1 Similarity as Spatial Distance

C2P12 Let us return to the Mingus pieces above and throw two others into the mix, examining two dimensions along which they vary: note density and tonic prevalence (see Figure 2.2). These dimensions are obviously just two of many, and they have been chosen somewhat arbitrarily. Note density is a simple measurement of the average number of onsets per second, and tonic prevalence measures the proportion of notes that are the global tonic (F for "Better Get It in Your Soul," "Wednesday Night Prayer Meeting," and "Fables of Faubus," and E-flat for "Goodbye Pork Pie Hat"). These metrics are based solely on the melodies and do not include any harmonic progressions. The melodies were analyzed using Humdrum software (Huron 1994).

C2P13 Note how this simple plot can demonstrate relative proximity between the pieces in question: "Prayer Meeting" and "In Your Soul" are modestly close to one another, demonstrating that they are similar not only in the categorical respects (key, form) mentioned above, but in terms of more specific, continuous metrics. The placement of the other pieces can also lead to further discussion not grounded in such quantitative metrics: "Fables," a protest song with sections meant to evoke discord, contains less tonic prevalence, while "Pork Pie Hat" has the lowest note density and a tonic relationship more closely aligned with the two aforementioned pieces than to "Fables."[6] Clear from this figure is the idea that although distance is absolute, similarity remains a relative concept, one that depends on the location of other points in the similarity space. The positions of "Fables" and "Pork Pie Hat" inform our evaluation of the degree of similarity between "Better Get It in Your Soul" and "Prayer Meeting," reflecting the earlier point that similarity is a relative construct.



C2F2 **FIGURE 2.2** Four pieces by Charles Mingus plotted according to each's frequency of onsets and the relative duration of the tonic triad—two distinguishing features that provide us with a map of distance between the four.

The similarity between *A* and *B* — equals the inverse of the distance between them

$$sim(A, B) = \frac{1}{1 + \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2}}$$

where distance is totaled across all relevant musical dimensions

C2F3 **FIGURE 2.3** An equation for similarity in terms of the Euclidean distance formula, in which the maximum similarity is 1. The difference between musical objects A and B are computed along each of *i* dimensions, and these differences are squared, added together, and then square rooted (to undo the square). With greater differences across dimensions, the denominator of the equation becomes larger, resulting in a smaller overall similarity value.

C2P14 A *spatial distance* approach equates similarity with how close two objects are along some predefined dimensions. Distance is inversely proportional to similarity: the greater the distance between objects, the less similar they are. This concept can be represented mathematically by the formula in Figure 2.3. The details here are less crucial than the broader concept—that distance is equal to the sum of the differences between two objects along whatever dimensions are relevant for assessment. Spatial measurement assumes that the properties of musical objects under comparison are (a) continuous—they can be measured as "more" or "less" of something, rather than being one thing or another—and (b) independent of one another.

C2P15 Because many musical parameters can be represented continuously, spatial distance measures have been used extensively in stylistic analyses examining distance between two groups of compositions. For example, many studies of authorship attribution employ clustering techniques as a way of collating many musical features (e.g., how often certain pitches occur, how often certain tones go to certain other tones, etc.) and create a multidimensional space within which to measure these pieces (see Brinkman, Shanahan, and Sapp 2016; Rodin, Sapp, and Bokulich 2010). In this approach, the pieces are measured by their distance from one another, and clusters are formed. Those similar enough to each other are conceived as belonging to the same cluster. ▶ Web Example 2.1 demonstrates how pieces known to be composed by Du Fay are placed in different locations than those known to have been composed by Josquin. Note the arrows demonstrating a location in relation to a central point. The circles demonstrate where most pieces by this composer are likely to fall, and the arrows show ranges along several continuous dimensions, such as frequency of parallel, similar, or oblique motion in voices, the rhythmic variability (nPVI) of a segment, the number of 9–8 suspensions in a piece, among others. Put succinctly, spatial distance calculations can be used to classify musical objects into discrete categories, such as works by a particular composer. We will observe that classifying musical objects is a central application of measuring similarity.

C2P16 In the field of Music Information Retrieval (MIR), spatial distance measures have been used to quantify similarity between musical segments such as phrases or entire works. For instance, de Haas, Wiering, and Veltcamp (2013) created a tool to measure the harmonic similarity between any two chord progressions and then used the measure to retrieve similar progressions from a corpus. The measure utilizes Lerdahl's theory of tonal pitch space (2001) to plot, beat by beat within a progression, each chord's distance from tonic. The authors then calculated the similarity between two progressions by overlaying each progression's beat-by-beat distance from tonic. The less space between the plots, the more similar the progressions, reflecting the inverse relationship between distance and similarity.

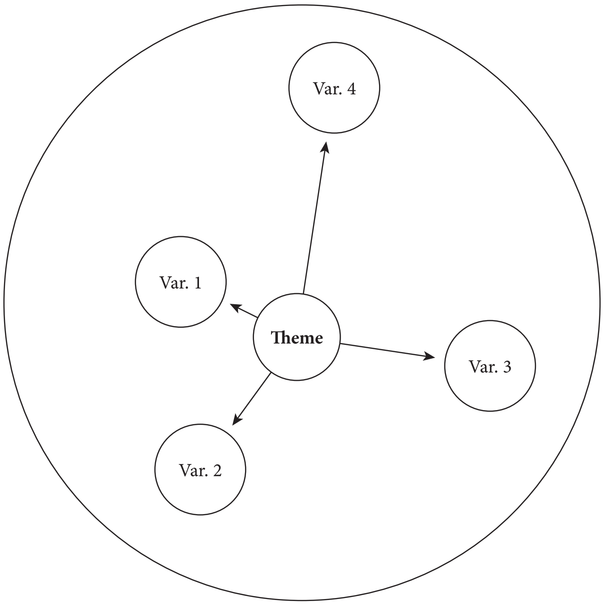
C2P17 At the larger scale, computational approaches to musical similarity are often rooted in the distance mentality for a simple reason: measuring musical distance allows one to collapse many different musical dimensions into a single similarity space. Velardo et al. (2016) terms this common approach in melodic-similarity modeling the "linear combination of metrics." For a case in point, Müllensiefen and Frieler (2004) combined a vast array of distance measures based on pitch, duration, contour, tonality, and accent structure to arrive at the combination that best matches human similarity judgments. Because they rely on *edit distance*, a transformation-based approach to similarity, we will return to their work below. The important point is that their model represented similarity between musical objects in terms of distance, and these distances tended to reflect how they were rated as similar by listeners.

C2P18 Each of the above approaches involved a priori decisions about which musical dimensions warranted measurement. But another advantage of the spatial distance approach is that it allows researchers to move in the other direction, taking listener ratings of similarity as a "given" and uncovering the underlying psychological or music-theoretical dimensions guiding those judgments. Multidimensional scaling (MDS) is a statistical technique that takes the distances between musical objects and finds a spatial arrangement of those objects that best "fits" all of their distances. The dimensions of the space can then be interpreted meaningfully. In a study by Lucy Pollard-Gott (1983), listeners heard an excerpt of Liszt's Piano Sonata in B minor, which contains variations of two distinct themes. She then played pairs of those variations and asked listeners to judge their similarity, analyzing the similarity data using MDS. Listeners who heard the piece only once tended to rely on nonthematic features like pitch height and loudness when determining how similar they were. In contrast, listeners who heard the piece three times recognized that variations of the same theme were similar. By measuring similarity spatially, Pollard-Gott was able to demonstrate how repeated encounters with a piece help us organize and pull apart distinct themes.

C2P19 In another series of applications, music cognition researchers leveraged MDS to uncover how our minds organize various musical parameters based upon similarity. Investigating tonality, Carol Krumhansl and Edward Kessler (1982) asked listeners how well each of the twelve chromatic pitches fit a preceding tonal context. The logic behind this *probe-tone method* is that goodness of fit serves as a proxy for tonal relatedness: a pitch belonging to the tonic triad is judged to fit better with a given key compared to a chromatic note, reflecting that it is more related to the key. Using MDS, the authors demonstrated (perhaps unsurprisingly) that listeners' ratings resulted in a spatial map of keys reflecting circle-of-fifths, parallel, and relative key relationships. By showing how this mental organization of distances mirrors music theorists' spatial maps of keys, the authors were able to demonstrate the "psychological reality" of music-theoretical intuition. Investigating timbre, Stephen McAdams and colleagues (1995) obtained pairwise similarity ratings for eighteen distinct musical timbres. The MDS solution revealed that the perceptual dimensions of timbral similarity were correlated with the acoustic properties of *rise time*, *spectral centroid*, and *spectral variation*, suggesting that these features drive timbre perception more broadly.[7] MDS thus helped uncover the properties of sound that cause certain timbres to sound more similar than others. Finally, investigating melodic similarity, Tuomas Eerola and Micah Bregman (2007) asked which aspects of folk-song melodies predict listener assessments of melodic relatedness. Using MDS, the authors found that pitch direction and pitch range were most influential for listeners in that regard.

C2P20 Music theorists have directly applied spatial distance measures to capture similarity relationships. For instance, post-tonal theory has a longstanding tradition of measuring the aural similarity of pitch-class sets by applying elements of the distance formula above. Richard Teitelbaum (1965), Robert Morris (1979), and David Lewin (1979), among others, created equations to measure the distance between two interval vectors.[8] While these methods spurred advances in modeling set class relatedness, Ian Quinn (2001) countered that their adherence to equations rather than to aural intuition caused them to lose sight of aural perception—ironically, the very thing they sought to measure. He advocated for, among other things, a return to behavioral measurement of similarity, of the kind explored in the approaches above, to map set class similarity.

C2P21 Of course, even without explicitly invoking the distance formula, music theorists frequently use spatial metaphors to represent relationships between musical entities that are more or less similar to one another. Inheriting perspectives from conceptual metaphor theory and cognitive linguistics, music theorists such as Lawrence Zbikowski (2002, 2017) and Michael Spitzer (2004) invoke distance metaphors in our understanding of music's dynamic unfolding. Spitzer and Zbikowski explain that one way we tend to contemplate musical categories is in terms of prototypes, or central tendencies, around which more and less similar tokens are thought to be arranged radially (Figure 2.4). This kind of center-periphery organization is implied in discussions of motivic variation when, for instance, a motive "moves further away from its typical form" (Zbikowski 1999, 31), or is measured in



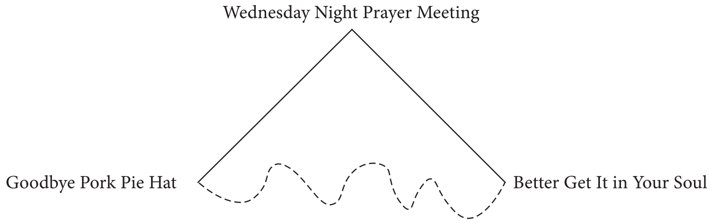
C2F4 **FIGURE 2.4** The radial organization of a thematic category. Variations that are more similar to a prototype appear closer to the center.

terms of "remoteness from the motivic source" (Boss 1992, 136). In fact, this mapping between spatial distance and similarity, or typicality, is fundamental to music-theoretical discourse on a larger scale. Consult any theoretical treatise and you will see instances of scalar predication—descriptions of objects as *more of* or *less than* something else. This language offers a window into the distance orientation of the music-theoretical mind.

C2P22 Spatial approaches, for all their utility, are not without limitations, and as they gained traction in the 1970s, psychologists such as Amos Tversky (1977) identified several such issues. For one, because identical objects have the same "address" in space, spatial accounts imply that they must always be maximally similar; however, this *reflexivity* constraint does not always hold. Krumhansl observed that it is easier to judge two chords to be identical if they are central to the key rather than out-of-key (Krumhansl 1979). That attests to the fact that highly "referential" musical objects are easier to recognize, unevenly affecting their perceived similarity.

C2P23 Another problematic assumption of spatial accounts is that the perceived similarity between A and B must be the same as that between B and A. But this *symmetry* assumption is frequently violated. James Bartlett and Jay Dowling (1988) found that when a diatonic melody followed a highly chromatic one, participants tended to judge the two to be more similar than when the presentation order was reversed. This demonstrates the issues with the symmetry axiom: phenomena differ in terms of our familiarity and fluency with them, such that it is easier to map from an unfamiliar knowledge domain onto a familiar one than the reverse.

C2P24 Finally, similarity assessments frequently violate what is known as the *triangle inequality*, which states that when three objects are being compared, no two objects should be further apart than the sum of the other two distances. In other words, if A is similar to B, and B is similar to C, then A should not be too dissimilar to C. Tversky's famous example compares



C2F5 **FIGURE 2.5** "Goodbye Pork Pie Hat" and "Better Get It in Your Soul" are more dissimilar to one another than a spatial account of similarity would require, considering each of their proximities to "Wednesday Night Prayer Meeting"

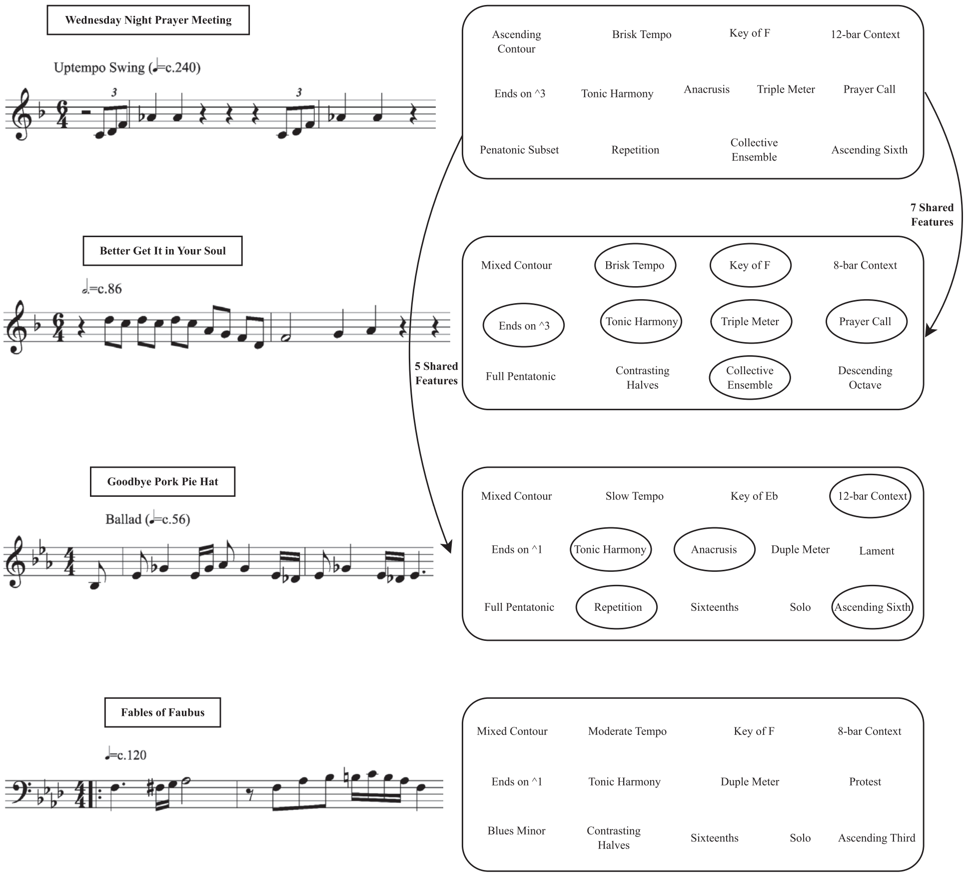
three countries to one another: Jamaica, Cuba, and (then) Soviet Russia. Jamaica and Cuba are similar because they are located in the same geographic region and share a tropical climate, while Cuba and the USSR are similar by their adherence to communism. But Jamaica and the USSR are quite dissimilar. This failure of *transitivity*, inherent to spatial distance accounts, can easily be observed in our Mingus examples (see Figure 2.5). "Pork Pie Hat" and "Prayer Meeting" share a twelve-bar blues form, while "Prayer Meeting" and "In Your Soul" share an arrangement structure (refer back to Table 2.1). But "Pork Pie Hat" and "In Your Soul" are completely different: one is a twelve-bar blues, the other a ten-bar A–A–B–A; one is fast, the other slow; and their arrangements could not be more dissimilar—the repeated blues form consists of only an extended saxophone solo between statements of the "head." What is needed is a system that can recognize that the relevant dimensions of similarity change as one moves through the triangle—in short, a feature-based approach.

C2P25 In sum, spatial accounts assume reflexivity, symmetry, and transitivity, but these assumptions are frequently at odds with perceptual judgments of musical similarity. If, as Quinn suggests, the goal of measuring similarity is to capture listener intuition, then an alternative to spatial distance would be welcome. The three other approaches to measuring similarity reported below overcome some or all of these deficits.

C2S2 Similarity as Feature Set Intersection

C2P26 To overcome these limitations, Tversky (1977) proposed that similarity is based not on continuous metrics but on discrete features. He presented empirical evidence that the perceived similarity between two objects increases when they have more features in common and decreases when they have more unique features. In Figure 2.6 we have presented the opening two measures of each of the four Mingus pieces, along with a collection of features that can be identified for each. Note that these features have been chosen based upon our own analysis and are not meant to form an exhaustive list. For ease of reference, we have also arranged these features in Table 2.2, with relevant properties ("attributes") in rows and pieces and their values in columns. Note that "Prayer Meeting" and "In Your Soul" share seven features and differ along five, while "Prayer Meeting" and "Pork Pie Hat" share five features and differ along seven. By a basic application of Tversky's model of similarity, "Prayer Meeting" is more similar to "In Your Soul" than to "Pork Pie Hat."

C2P27 But Tversky's approach to feature similarity was more nuanced. He intuited that the importance of features in guiding similarity should vary as a function of their *salience*. He argued that features become more salient when they help to distinguish objects from one



C2F6 **FIGURE 2.6** A feature-based approach to similarity represents musical segments as collections of independent features. Each opening is notated with a selection of features that we find to be of analytical interest. The circled features of "Better Get It in Your Soul" and "Goodbye Pork Pie Hat" are shared with "Wednesday Night Prayer Meeting." If similarity is proportional to the number of shared features minus unique features, then "Wednesday Night Prayer Meeting" would be more similar to "It in Your Soul" than to "Pork Pie Hat" in this context.

another—in other words, when they become more analytically useful or "diagnostic." Thus, for example, since all four Mingus openings use tonic prolongation, that feature is not salient; and because it is less informative, we tend to give it less weight than we give other features.

C2P28 Tversky also recognized that a feature's salience depends upon the individual making the judgment. Let us assume that our analytical goal is to assess motivic similarity. Because Tverskian features are mutually independent, we could capture this bias toward motivic structure by selectively increasing the influence of shared features related to motive, such as contour, intervallic structure, repetition, anacrusis, and formal context. For instance, we might choose to double count these features in our measurement. The total similarity between "Pork Pie Hat" and "Prayer Meeting" would increase from five to nine, while only raising the total between "Better Get It in Your Soul" and "Prayer Meeting" from seven to eight. This demonstrates how a feature approach can capture the effect of analytical context on perceived similarity: "Better Get It in Your Soul" becomes more similar to "Wednesday Night Prayer Meeting" when motivic aspects are given greater weight.

C2T2 **Table 2.2 Each motive's features organized in a table, with attributes as rows and their values for each piece in columns.**

C2P139

		Piece			
		"Wednesday Night Prayer Meeting"	"Better Get It in Your Soul"	"Goodbye Pork Pie Hat"	"Fables of Faubus"
Attribute	Tonic	F	F	E♭	F
	Meter	Triple	Triple	Duple	Duple
	Pentatonic set	Partial	Full	Full	Blues
	Anacrusis	Yes	No	Yes	No
	Contour	Ascending	Mixed	Mixed	Mixed
	Intervallic opening	Sixth	Octave	Sixth	Third
	Terminal scale degree	$\hat{3}$	$\hat{3}$	$\hat{1}$	$\hat{1}$
	Formal context	12-bar	8-bar	12-bar	8-bar
	Implied harmony	Tonic	Tonic	Tonic	Tonic
	Sixteenths	No	No	Yes	Yes
	Tempo	Brisk	Brisk	Slow	Moderate
	Orchestration	Collective ensemble	Collective ensemble	Solo	Solo
	Pragmatic context	Prayer call	Prayer call	Lament	Protest

C2P29 Finally, Tversky recognized that entire objects and their feature collections can be more or less salient depending on their familiarity or prototypicality. If "Prayer Meeting" were highly familiar to us, then it would make sense that comparisons involving "Prayer Meeting" would increase the impact of attributes it shares (and, simultaneously, downplay the attributes that are unique to the other piece). For instance, in comparing "Prayer Meeting" to "Pork Pie Hat," ascending contour would tend to be more salient and final scale degree less so, because the former is shared between the two and the latter is unique to "Pork Pie Hat," which is less familiar.

C2P30 Like the spatial account, this feature set approach can be traced to a single equation (Figure 2.7). However, in this case, the terms in the equation are not spatial distances but discrete features, weighted according to their importance. Again, we will focus not on the minutiae of the equation, but rather on the concepts it models. Broadly speaking, similarity is equal to the number and salience of shared features minus the number and salience of features unique to each object. Similarity increases with the number of shared features, and even more so if those shared features have greater salience. Finally, scaling parameters allow one to manipulate the relative importance of shared features and those that are unique to each object, again allowing for one object's unique features to "detract more" from similarity than the others.

The similarity between A and B — equals the influence of shared features — minus the influence of features distinct to A and B

$$sim(A, B) = \alpha f(A \cap B) - [\beta f(A - B) + \gamma f(B - A)]$$

scaled according to the importance of similarities (α) and differences (β and γ) — and weighted according to the importance of each feature (f)

C2F7 **FIGURE 2.7** Tversky's feature-set approach to similarity, arguing that similarity is based upon the number (and importance) of shared and unique features.

C2P31 An obvious question for this approach is: what constitutes a feature? In the field of MIR, similarity may be calculated using features of an audio signal or more abstract representations of audio.[9] Distinct collections of features may afford multiple viewpoints of a given musical surface, allowing researchers to evaluate similarity along distinct criteria (Conklin and Witten 1995). One kind of feature is the "*n*-gram," defined as any continuous string of musical objects of length *n*. Müllensiefen and Pendzich (2009) applied a variant of Tversky's equation to model similarity in terms of melodic interval *n*-grams. Measuring each *n*-gram's salience as a function of its frequency in a corpus of melodies (less frequent *n*-grams were considered more salient), they showed that melodic similarity resulting from Tversky's equation succeeded in predicting over 90% of plagiarism court case outcomes. Importantly, the winning model relied on the asymmetry assumption fundamental to Tverskian logic: features shared with the plaintiff's song mattered more than those unique to the defendant's. Elsewhere in music cognition research, features emerge as a core representational unit in studies of similarity. Lamont and Dibben (2001) presented listeners with pieces by Beethoven and Schoenberg and asked them to rate similarity between pairs of excerpts taken from the same piece. The researchers found that for the Beethoven piece, dynamics, texture, and articulation mattered more for similarity judgments, whereas for the Schoenberg, tempo and dynamics were most important. They concluded that "the pieces set their own similarity criteria," further supporting Tversky's notion that feature importance should vary as a function of context.

C2P32 Just as spatial distance accounts of similarity enable one to classify objects, an important application of feature-based approaches is to categorize musical phenomena. In fact, as Emilios Cambouropoulos (2001) argues, one quite simply cannot speak of similarity without a robust theory of categorization. Though a thorough review of the complex relationship between similarity and categorization is beyond our present scope (see Cambouropoulos 2001 and Ziv and Eitan 2007 for helpful overviews), we note several highlights here. First, psychologists argue that categories may be organized in different ways depending upon how they enlist features. A category on the *classical* or *criterial view* assumes that certain features are necessary and sufficient for a particular musical object to qualify as a category member. For instance, pitch-class sets are categories of necessary and sufficient interval class content, and Classical-style formal functions are categories of necessary and sufficient motivic and tonal criteria. Alternatively, the *prototype view* of categories, which we introduced in the preceding section, holds that features of objects are not required for membership, but rather are more or less typical. The prototype is an abstract composite of the most typical features, even if that composite does not actually occur in the real world (Rosch 1978, 40).

C2P33 While it may seem intuitive that with classical categories, measures of similarity are all-or-none, and with prototype-based approaches, similarity is graded, the relationship between similarity and categorization is more complex. Objects within classical categories may vary in similarity based upon non-criterial features. For instance, three pitch collections belonging to the same set class might be more or less similar to one another based on the number of pitch classes their normal forms share. Moreover, entire classical categories may be more or less similar to one another. We saw in the previous section that set theorists quantified similarity between set classes, while similarity was assumed to be all-or-none within the classes themselves.

C2P34 Many music-theoretical approaches implicitly align with Tversky's view of similarity for two reasons. First, regardless of whether similarity is graded or all-or-none, the feature is the core representational unit of a musical object. Second, the influence of features on similarity relationships critically depends upon their salience. Schoenberg famously held that "features are the marks of the motive" (Schoenberg 1995, 130), a definition that has spurred studies of motivic similarity in feature terms. For instance, Zbikowski's theory of motivic similarity and categorization employs several of Schoenberg's ideas, representing musical motives as collections of features. In Zbikowski's approach, features are analyzed in terms of attributes and corresponding values, much like our features in both Figure 2.6 and Table 2.2. In a compelling analysis of motive in Mozart's String Quartet in C, K. 465, Zbikowski (1999) demonstrated how typicality (and similarity) evolves over the course of the piece, as motive forms lose shared features with a prototype and gain unique ones. The ebb and flow of featural similarity thus becomes a powerful description of formal process in this approach.

C2P35 Many other theorists tacitly endorse Tversky's intuition that motivic similarity is proportional to the number of shared features between musical events (see, for instance, Meyer 1973, Gjerdingen 1988, and Huron 2001). In recent work, Janet Bourne casts Meyer's primary parameters as networks of relationships that variations may share; these relations resemble Tverskian features that are summed to arrive at a similarity score as in the contrast model. Many of those authors endorse a prototype view, where features of a category are more or less typical.

C2P36 Finally, a third group of theorists allows the analyst to operationalize features in the way they see most useful. Dora Hanninen's (2004) associative sets permit the analyst to choose whether prototypes, classical categories, or some combination thereof best captures their analytical goals. Like Hanninen, Ian Quinn (2001) recognizes the utility of having more than one categorization scheme. He advocates for fuzzy sets, which permit both binary and graded category membership. In both cases, musical sets are collections of features. Scott Murphy's (2020) recent innovations to the class system of tonal-triadic progressions enforce criterial categories like traditional set classes, but they empower the analyst to choose which (combinations of) features define a class. With these flexible feature schemes, musical objects may still be classically categorized but have graded similarity relationships to other tokens within a category and to other categories.

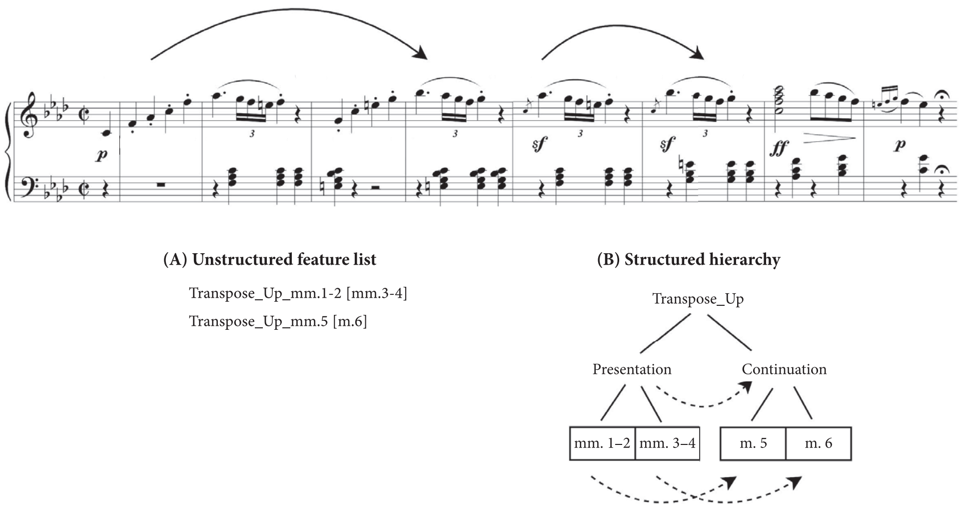
C2P37 In summary, many music theorists implicitly apply principles of Tversky's contrast model when engaging in analysis. Lists of features, often cast as attribute-value pairs, form a core representational unit of musical objects. Theorists consider the similarity between two objects to be proportional to the number and salience of features shared between them. Feature salience is guided by analytical goals, heuristic constructs like primary and secondary parameters, an attribute's diagnosticity (as a function of the other objects or categories in the feature space), and general perceptual principles.

C2S3 Limitations to Feature-Based Approaches

C2P38 As with the spatial approach, there are some notable limitations of feature-based approaches. For one, the contrast model requires that an analyst know in advance all of the relevant musical features and how salient they are before a similarity computation can proceed (Cambouropoulos 2001; Cambouropoulos and Kaliakatsos-Papakostas 2022). Is this a realistic requirement? More importantly, however, feature approaches tend not to encode the *connections* between features. This means that, like the mental coordinates of the spatial distance approach, feature lists lack higher-order structure (they are what psychologists call *unstructured representations*). Hahn (2014) identifies a critical drawback to this lack of structure: to capture relational information, one must burdensomely encode the entire relation as its own feature, limiting the mind's ability to pick up on components that match between different features. Note, for instance, that the opening sentence of Beethoven's Piano Sonata in F minor, op. 2, no. 1 (depicted in Figure 2.8A), contains a similar transpositional relationship between the basic ideas of the presentation and their subsequent fragmentation in measures 5–6. In a feature-based approach, separate features would be needed to encode the transpositional relationship in each section. Because a feature does not separate the relation "transpose" from the rest of the attribute-value system, it cannot pick up on the analogy between mm. 1–4 and mm. 5–6. These are separate features with no higher-order relationships.

C2P39 Musical similarity, however, arguably springs from the connections between parts of the list. The palpable increase in energy at mm. 5–6, characteristic of sentential fragmentation, relies on the recognition that the same transposition function is applied to timespans that are half as long. Recognizing this analogy is proof that the mind separates out the relation from the spans it relates, allowing for higher-order formal connections to emerge in a structured way (Figure 2.8B). Perceiving ordinal relationships therefore requires cognitive structure. However, even when features are not timespans but abstract attributes like musical parameters, it is truly the *interactions* between parameters that undergird mental representations of musical materials. (For a simple thought experiment, conjure up an auditory image of some musical rhythm. It would be exceedingly difficult to hear the rhythm without simulating a timbre or some other piece of musical information to go with it. Rhythm may be isolable in theory but is not independent of other entities in a perceptual representation.)

C2P40 MIR researchers likewise rely on structured representations in modeling similarity. Cambouropoulos and Kaliakatsos-Papakostas (2022) suggest that spatial approaches lack cognitive relevance because they are unstructured, advocating instead for a stream-based representation of the musical signal. He reviews how this sort of structural encoding enhances researchers' ability to seek out and match musical patterns of a corpus. Thus, MIR greatly benefits from mimicking the structure of mental representations of musical materials.



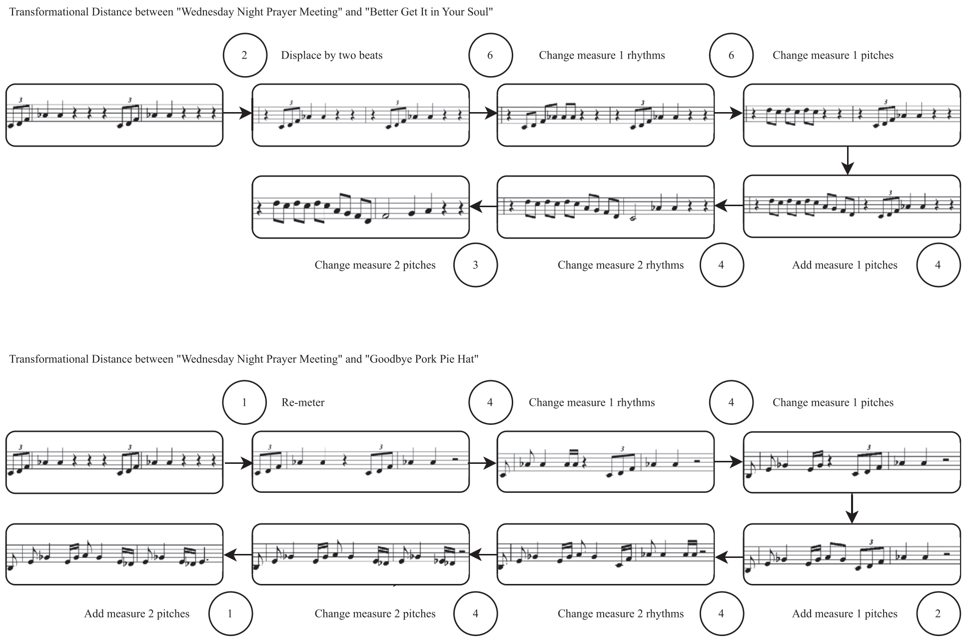
C2F8 **FIGURE 2.8** Two alternative mental representations of motivic structure in Beethoven's Piano Sonata in F Minor, op. 2 no. 1: (A) An unstructured feature list must burdensomely encode parallel relationships as separate features; and (B) A structured symbolic representation separates functions and their arguments and encodes the relationships between them, capturing parallel structure that is a quintessential aspect of musical experience.

C2P41 The next sections therefore turn away from spatial and feature approaches and turn toward those that assume structured representations: similarity as transformational distance and similarity as alignment.

C2S4

Similarity as Transformation

C2P42 Conceiving of the Mingus openings not as lists of features but as structured representations allows us to measure their similarity in terms of part-whole relationships. Returning to the two feature comparisons in Figure 2.6, we can capture similarity between "Wednesday Night Prayer Meeting" and "Better Get It in Your Soul" and also between "Prayer Meeting" and "Pork Pie Hat" as a series of edits by which to turn the incipit of one song into that of another. In Figure 2.9, we demonstrate how information theorists might compute similarity according to the fewest number of additions, deletions, or substitutions that must be made, on a note-by-note basis, to turn the structure of one into the other. Notice that, under our present scheme, it takes twenty-five edits to turn "Prayer Meeting" into "In Your Soul," but only twenty edits to turn "Prayer Meeting" into "Pork Pie Hat." Unlike with the feature approach above, "Prayer Meeting" is now more similar to "Pork Pie Hat," demonstrating how assessments of similarity vary according to the measurement system used. Note that this simplistic scheme is merely one possible *kind* of edit distance; like feature approaches which can vary salience and feature weights, edit distances are versatile by allowing one to manipulate the salience of particular edits and the weights given to particular parts of



C2F9 **FIGURE 2.9** An illustration of similarity as transformational distance between "Wednesday Night Prayer Meeting" and both "Better Get It in Your Soul" (twenty-five total edits) and "Goodbye Pork Pie Hat" (twenty total edits). By this scheme, "Wednesday Night Prayer Meeting" might be considered to be "more similar" to the latter than the former.

a sequence undergoing change. In studies of edit distance, the amount of mental effort required for a particular kind of transformation is termed "cost."

C2P43 While we have treated them as separate approaches to similarity measurement, spatial distance and edit distance stem from a common principle. Both measure the mental proximity of two objects on a continuous scale, and in both cases, this proximity is inversely proportional to similarity. Here, the easier it is to turn one object into another, the closer their mental proximity and the more similar they are judged.

C2P44 Measuring distance based on general edits—such as insertions, deletions, and substitutions—has succeeded in capturing human ratings of similarity across many areas of cognition with a high degree of accuracy, reinforcing the validity of the approach.[10] But to make measurements more precise, music cognition researchers have also sought to replace these general edits with music-specific operations, an approach that is ubiquitous in MIR. Mongeau and Sankoff (1990) replaced deletion and insertion with consolidation and fragmentation, allowing them to calculate similarity between a theme and nine variations in Mozart's Twelve Variations on "Ah vous dirai-je, Maman," K. 265/300^{e}. These edits enabled them to overcome differences in key, tempo, or mode and focus on melodic transformation, and the resulting edit distances largely matched their subjective intuition about variation relatedness. Müllensiefen and Frieler (2004) found that in addition to *n*-gram measures similar to the ones discussed above, edit distance measures between excerpts on the basis of pitches, intervals, contour, and rhythm, optimally predicted listeners' similarity ratings.

For a third application of edit distance, Marcus Pearce and Daniel Müllensiefen (2017) enlisted a computational model of auditory expectation, termed the Information Dynamics of Music (IDyOM), to calculate the transformational distance between two melodies based upon how easily the model can predict each melody from the other (termed "compression distance"). This use of IDyOM succeeded in modeling a body of behavioral similarity data; indeed, expectation-based approaches to musical structure are an exciting avenue for future transformational accounts of musical similarity (see Cambouropoulos and Kaliakatsos-Papakostas 2022 for an overview).

C2P45 Do music theorists understand similarity as a (perhaps implicit) process of turning one musical phenomenon into the next? The approach most aligned with this account is transformation theory, which usually involves measurements of change between diachronic musical spans unfolding in real time. Though it is rarely cast as a theory of similarity because transformations focus our attention on musical change, one nevertheless finds evidence that music theorists implicitly understand transformations as ratios of difference to similarity. For instance, Joseph Straus (2016) defines the standard neo-Riemannian voice-leading transformations in terms of scale degrees that change relative to scale degrees that do not. The basic operators preserve more than they change, while the secondary operators, which derive from combinations of the basic ones, change more than they preserve. Underlying Schoenberg's (1995) approach to variation is an intuitive assumption that musical variation is an agential process (whether on the part of composer or listener) of changing a gestalt motive into new forms. He enumerates various transformational primitives, such as *variation*, *reduction*, and *extension*, which analogize respectively to substitution, deletion, and insertion of the basic edits above. That Schoenberg conceived of change on a continuous scale from "only the most primitive coherence-producing repetition" to "slightly varied repetitions" to "somewhat more richly varied" repetitions (1995, 157) attests to his view that the degree of variation between musical spans is to some degree continuous, and that certain kinds of changes are more drastic (that is, more "costly") according to how much they degrade recognizability with respect to the original. While we have been speaking of edits *resulting* in varying degrees of similarity, operations of repetition and contrast can themselves be cast as transformational primitives. By positioning formal relationships in transformational terms, Schoenberg inspired subsequent music theorists to regard form in this way, with exact and varied repetition being sequenced with contrast in a diachronic chain to produce theme types and higher-order structure (e.g., Caplin 1998 and Yust 2020).

C2P46 In sum, while musical form is rarely cast explicitly as resulting from a transformational approach to similarity, theories of form seem to be concerned with codifying transformational grammars: What are the ways one can turn X into Y? What stays the same, and what changes?

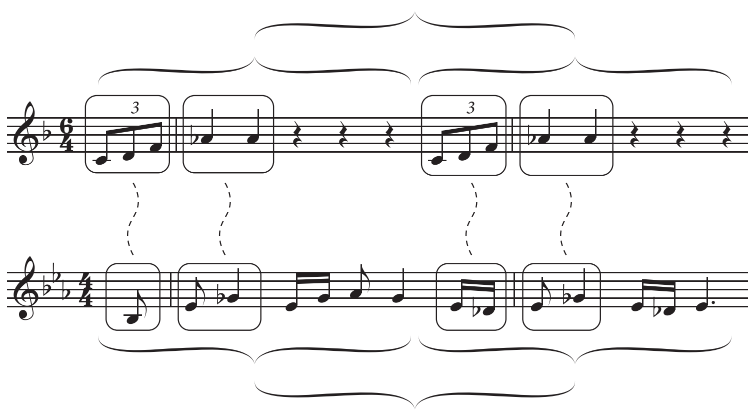
C2P47 Like each of the above approaches, transformational similarity is not without its drawbacks. First, transformational approaches likewise require one to commit to a repertoire of edit types that are known in advance (Hahn 2014). Second, as in cognition studies, when one defines a transformational scheme in music theory, one risks missing a simpler transformation from one object to the next. Third, whereas comparing individual features of two musical objects seems at least somewhat plausible during real-time listening, does the mind really encode distance as a series of edits on the fly? Transformational theorists are indeed burdened by the same question: does the "transformational word length" (Lehman 2018) between two passages meaningfully index perceptual similarity? Finally, by transforming objects step-by-step, one risks losing sight of higher-order correspondences between objects, correspondences that make similarity relationships more straightforwardly apparent.

C2S5

Similarity as Alignment

C2P48 Unlike a transformational approach, which would measure the effort of getting from X to Y, structural alignment describes a more general procedure of bringing the parts of X and the parts of Y into contact with one another to compare their larger wholes, with the result that commonalities and corresponding differences become foregrounded. If the transformational approach to similarity risks losing the forest for the trees, alignment-based approaches ultimately celebrate higher-order structural connections. They do this by offloading the "burden" of finding the aspects of similarity between two objects onto the very act of comparing. Recall that it took twenty edits to turn the surface of "Prayer Meeting" into "Pork Pie Hat." But if we juxtapose these two openings, which encourages us to look beyond their differing surface content (pitches, pitch classes, rhythms, durations, etc.), we uncover higher-order connections (Figure 2.10). Both motives decompose into two halves with exact and varied repetition, respectively. Within each half, anacruses precede a downbeat with two chord-tone onsets. These onsets are exactly repeated across the two halves of each motive. Aligning these analogous spans also foregrounds a critical difference between the two, namely the rests segmenting the first motive's halves versus the varied continuation of the second.

C2P49 Observe that there is nothing "predetermined" or "pre-specified" about the connections between these two forms. In fact, the analogous timespans are quite different on the surface, occupying different rhythmic durations, for instance. Rather, it is by the act of juxtaposing them that their analogous structure emerges. Once aligned, additional



C2F10 **FIGURE 2.10** A structural alignment of the two opening melodic lines of "Wednesday Night Prayer Meeting" and "Goodbye Pork Pie Hat." The process of alignment reveals analogous halves with distinct kinds of repetition, despite quite a bit of dissimilarity on the surface.

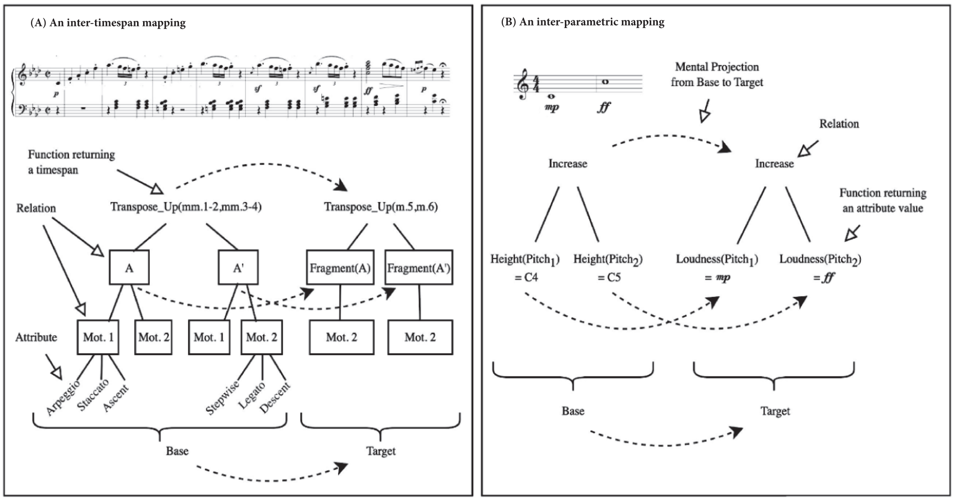
relations become clear: anacruses emphasize the songs' respective $\hat{5}$s, the heads accent respective $\hat{3}$s, and the exact repetition of "Prayer Meeting" foregrounds the modified repetition of "Pork Pie Hat." It becomes possible to transform one into the next with only several higher-order edits, reiterated over corresponding repeats. Moreover, the alignment also foregrounds *how* their repetition schemes differ. The repeated basic idea of "Prayer Meeting" is open-ended, whereas the modified repeat of "Pork Pie Hat" makes possible an antecedent-consequent relation that is considerably more closed. This difference in repetition, as well as the differences between the motives' continuations, are what similarity researchers term *alignable differences*. They become salient as a result of the correspondences motivating them.

C2P50 The process of alignment is formalized according to structure-mapping theory (Gentner 1983), and while structure-mapping theory originally sought to explain analogical reasoning, it is now recognized to be a general cognitive mechanism involved in many different aspects of cognition, such as visual perception (Forbus et al. 2017), discourse comprehension (Day and Gentner 2007), and category learning (Goldwater 2017). Structure-mapping theory assumes that the mental representations of musical objects we compare are made up of attributes, relations, and functions (see Figure 2.11), and these elements constitute the things we align when we compare. Alignment encompasses at least four cognitive stages—*retrieval*, *mapping*, *evaluation*, and *abstraction*. Janet Bourne (2015) advances a theoretical framework for understanding retrieval, mapping, and evaluation in terms of musical sense-making (and provides a detailed account of the stages discussed below).

C2P51 During *retrieval*, when we encounter a musical object that reminds us of something familiar, we *retrieve* the similar object from memory. Alternatively, when we encounter two similar musical objects in succession, no retrieval is necessary. In either case, and in line with structure-mapping theory, music theorists intuitively understand musical events as *invitations to retrieve and align* similar objects.

C2P52 During *mapping*, we mentally project structure from one object onto the next in search of parallel correspondences. This search for correspondences between musical objects is governed by several constraints, which turn out to be near-universal laws of musical analysis: (1) elements across objects must map in one-to-one fashion, (2) higher-order relations between matches should be maintained across objects, and (3) mappings are "better" if relations and attributes match. Functions can, but need not, match. This capacity to align non-identical functions results in cross-domain metaphor (e.g., Zbikowski 2002). For instance, Eitan and Granot (2007) suggest that changes in musical intensity is a parameter-general principle. They argue that the same "intensity contour" can be mapped across different parameters (as depicted in Figure 2.11B, where two different functions—pitch height and loudness—are mapped to one another to create an inter-parametric analogy. (We will return to Figure 2.11A shortly.)

C2P53 During *evaluation*, we qualitatively assess the results of the mapping process. How well do the objects correspond? When some parallel aspects of one object are missing from the other, a listener can project correspondences forward as "hypotheses" that are evaluated at this stage. Additionally, corresponding differences (termed "alignable differences") between the two objects become foregrounded for interpretation. Finally, the result of evaluation is that shared relational structure is reinforced or *abstracted* in memory. Caplin, following



C2F11 **FIGURE 2.11** Two types of structural alignment with music according to structure-mapping theory. (A) An inter-timespan mapping between Beethoven's op. 2, no. 1 sentential presentation and continuation. The return of motive 2 in measure 5 invites comparison between the presentation and continuation, resulting in the projection of mental correspondences depicted with dotted lines. Evaluation of this relationship results in the perception of fragmentation. (B) An inter-parametric analogy between pitch ascent and increasing loudness. Inter-parametric analogies (Eitan and Granot 2007) come about when relations ("Increase") match while functions ("pitch height" and "loudness") do not.

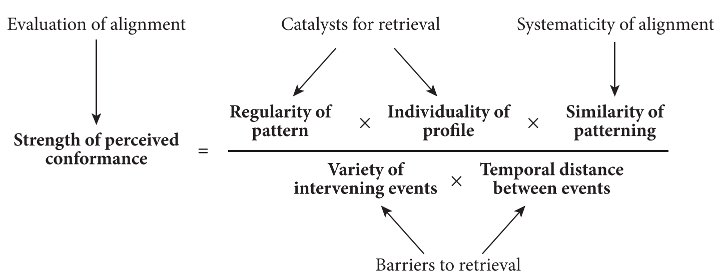
Schoenberg, intuited that repetition cements the motive in memory; structure-mapping theory provides an account of why.[11]

C2P54 Let us examine the mapping process using the Beethoven example in Figure 2.11A. Detecting the return of the opening motive in m. 3 invites a listener to align the two basic ideas, foregrounding the transposition up and inviting a listener to expect motive 2 in a parallel position. This alignment process facilitates the abstraction of the common motivic relationship of the presentation. When the continuation opens with motive 2 without the full motive 1, the match between presentation and continuation highlights the continuation's absence of the four-note arpeggio, resulting in the qualitative experience of fragmentation.

C2P55 Structural alignment has important advantages for theories of musical similarity. First, it does not require prior knowledge of relevant structural properties between objects ahead of time—rather, these may be *bootstrapped* from the comparison process itself. Second, structural alignment of two novel objects provides a mechanism by which musical categories may be acquired or abstracted (Gentner and Calhoun 2010). When two musical objects, such as a theme and variation, are encountered for the first time, mapping serves to reify their common relational structure, rendering the commonalities salient. This, in turn, results in abstraction, without any explicit instruction on the part of the composer. In contrast to accounts of musical development that emphasize the role of passive statistical learning, structural alignment provides an active mechanism for inducing musical knowledge structures through repeated exposure to stylistic norms.

C2P56 A key insight of the structural alignment approach is that any kind of musical feature may be cast as an "invitation to align" corresponding structures based upon how easily the feature's similarity can be detected. According to this outlook, elements of the musical surface may be understood as technologies for facilitating comparison. The music critic Hans Keller seemed intuitively aware of this fact: he designed wordless functional analysis, a method of manipulating motivic relationships to highlight the structural correspondences latent in the motivic structure of concert works (Swett 2022).[12] Moreover, Leonard Meyer's theory of conformant relationships (1973) bears the signature of this alignment-invitational perspective. Examining his equation for similarity (see Figure 2.12), the terms in the denominator of Meyer's equation—variety and length of intervening events—make the detection of similarity more difficult. Elements of feature salience described above work in the opposite direction, facilitating alignment. Finally, the core "similarity" term in the numerator—similarity of patterning—implies structural alignment by invoking pattern similarity. Patterns are by their nature structured, and cannot be described in terms of disembodied features.

C2P57 More than any of the other approaches, structural alignment demonstrates how the act of comparison controls which aspects of two musical objects are relevant for similarity, a reality to which music theorists are sensitive. In her theory of associative sets, Hanninen (2004) suggests that relational properties of a musical object change depending upon analytical context. In other words, the attributes, relations, or functions of a musical object that are matches with another object *depend upon* the objects being compared, reflecting a core principle of structure-mapping theory. Many feature-based approaches, including Hanninen's and Zbikowski's, implicitly endorse an alignment view of similarity: higher-order connections between features, rather than the features themselves, turn out to be essential in understanding, analyzing, and evaluating musical relationships; these connections only arise because we systematically align corresponding parts when we compare.



C2F12 **FIGURE 2.12** Leonard Meyer's formula for the degree of perceived similarity between two motivic structures (taken from Meyer 1973). The terms of his formula can be understood in terms of the stages of structural alignment. Pattern regularity and individuality, two aspects of Tverskyan salience, catalyze the retrieval of a motive from memory, while the variety of intervening events and temporal distance hinder retrieval. Similarity of patterning expresses the systematicity (or hierarchical depth) of correspondences between base and target, and the output strength of perceived conformance corresponds to an evaluation of the alignment.

C2S6 Music Theory's "Alignability Attitude": Structured Comparison in Music Analysis on a Larger Scale

C2P58 Although Schoenberg's theory of musical coherence predates the theory of structural alignment found in psychology, it is difficult not to compare the two. Recognizing that music perception is fundamentally an act of real-time comparison, Schoenberg (1995) defined coherence and contrast not in terms of simple feature overlap, but as the result of forging *structured connections* between two objects' corresponding parts. For Schoenberg, perceptions of coherence, variation, and contrast arose from the very act of mapping correspondences, thereby reflecting the core psychological principles of structure-mapping theory: the act of comparison renders alignable differences salient (as when we perceive variation or contrast) and reifies a common structure (as when we perceive coherence or variation). As Áine Heneghan (2019) has observed, Schoenberg's (1994) definitions of coherence, variation, and contrast presume a continuum of similarity. For Schoenberg, this continuum was based upon the amount of attention paid to each kind of correspondence within a structural alignment: coherence focuses attention on matches, contrast on alignable differences, and variation a little of both.

C2P59 That Schoenberg was implicitly aware of these psychological principles is hardly surprising: his was a cognitively oriented theory of form (Zbikowski 1999), one that sought an understanding of musical structure through an understanding of the mind. His insights invite us to consider the perception of variation structure as an act of structural alignment, one in which comparing a variation to its thematic prototype foregrounds both the common structure and salient changes. Of course, Schoenberg would not be alone in endorsing structured comparison as the cognitive process that gives rise to the perception of variation. As Roman Ivanovitch puts it, "the basic act of construing variation is a comparative one: the task of a listener is to relate two stretches of music, to hear one passage 'in terms of' another. To hear something 'as a variation' is perforce to be engaged in such an activity of comparison" (2010, 3).

C2P60 In fact, these insights reflect a larger disciplinary orientation toward structural alignment as a fundamental act of musical analysis. In the same way that David Lewin articulated a "transformational attitude" in music theory—that a musical transformation is an effort of turning X into Y—we can speak of a general "alignability attitude" underpinning the ways in which music theorists approach musical relationships. We define the central thesis behind this alignability attitude as follows:

> C2P61 Musical representations S and T are structured such that configurations within S systematically correspond to and differ from configurations within T. A central goal of musical analysis is to uncover these correspondences by aligning structured representations.

C2P62 Just as structural alignment represents a ubiquitous cognitive process, it represents a ubiquitous act in musical analysis. In this final section, we identify five fundamental principles guiding music-analytical acts that reflect the tenets of structural alignment and our deep disciplinary adherence to alignment-based similarity measurement.

C2S7 Principle 1: Musical Representations Are Inherently Structured

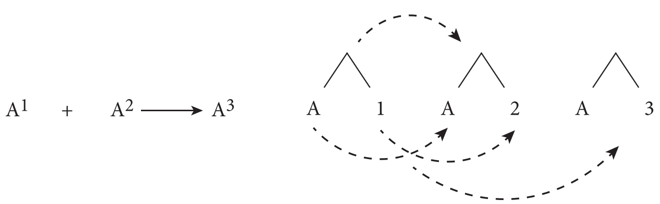
C2P63 In schema theory, it is not uncommon to recognize a schema's gist without its typical or defining features. (A Fenaroli that lacks the requisite scale degrees but nevertheless feels "Fenarolish"—perhaps by virtue of its prolongation function or cyclicity—achieves relational similarity without literal similarity; that is, it *analogizes* to a Fenaroli.) That we recognize this higher-order relational similarity absent essential features provides evidence that music analysts align relational structures when comparing tokens to exemplars. This extends to all manner of musical schemata, from phrase types to prolongational schemes to recurrent metric patterns.

C2S8 Principle 2: Formal Relationships Can Be Cast as Analogies

C2P64 In Figures 2.10 and 2.11 above, we demonstrated how structural alignment is enacted over musical timespans. Beneath that musical surface, deeper, schematic knowledge of formal tendencies is explained via structural alignment as well. For instance, Narmour's (2000) cognitive rules generate specific implications for musical continuation, and these implications are shaped by the abstraction of a relational structure between two generative events. (See Figure 2.13.)

C2S9 Principle 3: Structural Alignments Provide Compelling Evidence for Analytical Argument

C2P65 The ubiquity of structured comparison in music analysis cannot be overstated. Consider the opening paragraphs of Steven Rings' (2011) analysis of Brahms's Intermezzo, op. 118, no. 2, which carries the reader through no fewer than six structural alignments in the span of a few paragraphs (Table 2.3). Parallel correspondences provide the meat and potatoes of an analytical argument, whether by comparing musical objects, hearings, schematic tokens, or what have you.



C2F13 **FIGURE 2.13** One of Narmour's many cognitive rules for musical implication. Implicative structure can be cast as a product of analogical comparison between adjacent pattern elements (Narmour, 2000).

C2T3

Table 2.3 Like countless musical analyses, Steven Rings' account of the opening bars of Brahms's Intermezzo in A, op. 118, no. 2, relies upon numerous structured comparisons to construct analytical evidence. (See Rings 2011, 185–187.)

Excerpt	Base domain	Target domain
"the section is in lyric binary form . . . "	Schematic knowledge of binary form	The form of the section in question
"Schenker places [the onset of A] at the modified thematic restatement . . . as does Allen Cadwallader. A strong case can also be made for placing the onset of a′ at m. 34. Brahms provides signals that support both hearings."	Hearing 1	Hearing 2
"There is a subtle dialectical energy in these motives . . . both leap up melodically across the bar line, gently contradicting the metric up-DOWN of beat 3-to-1 with a melodic down-UP. Both leap away from B4, which proves unable to descend to A4."	Attributes of motive A Metric directionality	Attributes of motive B Melodic directionality
"Schenker hears the B4 motive α as a passing tone in an implied *Terzzug*, a hearing that is reinforced by the realized *Terzzug* A–G#–F# in the alto."	*Terzzug* 1	*Terzzug* 2
"rather than proceeding to its expected goal, A4, the gesture 'leaps off ' of the *Zug* (*springt ab*) to the D5 on the downbeat of bar 1. D-instead-of-A is a motivic substitution in the intermezzo: the music yearns for A but consistently gets D instead."	Schematic knowledge of implicative tendency	Recurrent thwarting of the expected descent

C2S10

Principle 4: Alignable Differences Are Analytically Privileged

C2P66 In Huron's theory of musical features (2001), salience is derived from a musical object's deviation from the norms of a corpus. This principle presumes an alignment between a type and a token. At a larger scale, the music-theoretic orientation toward dialogic form, by which a work's unique aspects are marked against schematic norms, is rooted in the analytical act of mapping a base to a target. As Bourne observes, it underscores Hatten's (1994) theory of markedness, according to which structured correspondences between elements yield relationships of correlation. In Sonata Theory (Hepokoski and Darcy 2006), the very notion of correspondence bars between an exposition and recapitulation presumes structured alignment within parallel connectivity.

C2S11

Principle 5: Structured Comparison Is a Pedagogical Tool

C2P67 As Gjerdingen (2020) observes, composition teachers of the old conservatories carefully crafted learning environments to foster schema learning. Training exercises were deliberately juxtaposed to foster abstraction of a common rule or pattern, reflecting an intentional deployment of structure-mapping principles centuries before the theory was formalized. Gjerdingen (2007) mirrors this principle in *Music in the Galant Style* via his deliberate sequencing of schema exemplars within chapters.[13] Theorists such as Ruwet (1966) and Reti (1951) followed Claude Levi-Strauss' practice of positioning analogous segments vertically on the page, presumably to facilitate comparison. This was essential for Reti, especially, who was after subsurface musical connections lacking obvious surface similarity. In his textbook, *The Complete Musician*, Laitz (2016) directs pupils to compare two examples of a musical period in order to foster abstraction of the common schema, highlighting the most important structural elements of periods—the weak-to-strong cadential relationship, two-phrase composition, and overarching harmonic plan—while foregrounding their motivic alignable difference (one being "parallel" and the other "contrasting"), leveraging principles of structure-mapping theory to teach form.

C2S12 Conclusion

C2P68 Returning to the opening figure of this chapter—a formal comparison of the layouts of "Better Get It in Your Soul" and "Wednesday Night Prayer Meeting"—we find in our own analytical act this broader "alignability attitude" that we argue governs the paradigms and processes of music analysis on a larger scale. We juxtaposed their layouts to highlight a shared relational structure, a process that also calls attention to their differences. Structural consistency and systematic connections between musical elements, though not inviolable laws, are two tenets of alignment that carry much weight in deciding the strength and success of a given analytical act of comparison. However, regardless of whether one adopts a spatial, featural, transformational, or alignment-based view of similarity, the central act of comparison in music theory is constrained and informed by an understanding of how we—as listeners and theorists subject to the laws of human cognition—measure similarity.

C2P69 The reader will note that many music theorists have reappeared throughout this chapter, perhaps most notably Schoenberg. His theories are an example of how these ways of measuring similarity are not necessarily mutually exclusive, but rather provide varied (and in some cases complementary) approaches to understanding musical relationships. Having reviewed the prevalence of spatial, featural, transformational, and alignment-based approaches to musical similarity in MIR and music cognition, and having observed these same approaches in music-theoretical treatises, we might take a renewed interest in the questions they continue to pose to us: What mental acts give rise to similarity? What is the influence of musical context on perceived similarity? How does an analytical apparatus constrain what entities can be similar? How do we represent the categories of musical objects being compared?

C2P70 The two Mingus pieces in Table 2.1 can serve as an interesting point of departure for these questions. To some extent, one is a variation of the other; but by questioning what is the object that is being transformed, and how we can hear those objects as similar or different, we can gain a deeper understanding of our own perceptual processes about musical similarity more generally.

Notes

1. When psychologists discuss the objects or features being evaluated, they use the term *respects* for similarity.
2. In his critique of Rudolph Reti's analysis of motivic transformation in Brahms's second symphony, Leonard Meyer observes that "if one can pick and choose—selecting those voices or pitches which support one's hypothesis and disregarding those which do not (the small notes in Reti's analyses)—then almost any melody can be related to any other whether within or between works" (1973, 62). Thus, Meyer's critique boils down to Reti's failure to articulate *respects* for similarity.
3. Emilios Cambouropoulos (2009) underscores this point in a theoretical discussion of similarity, in which he argues that musical similarity is always context-dependent.
4. Pearce and Müllensiefen (2017) provide a brief overview of these distinctions within the field of music information retrieval.
5. Janet Bourne (2015) advances an analytical framework for applying alignment-based approaches to music analysis, which we summarize below.
6. "Pork Pie Hat" is a ballad with a melody derived from an E-flat-minor blues reportedly improvised immediately after Mingus hearing of the passing of Lester Young, providing one interpretation for the relative reduction of tonic.
7. Rise time is defined as the time it takes for the sound amplitude envelope to reach its maximum amplitude from a threshold of within 2% of its maximum. Spectral centroid is a weighted mean of the frequencies present in the sound signal. Spectral variation is a measure of how different the sound spectrum is from moment to moment.
8. For a review and response to these distance approaches, see Isaacson (1990).
9. Cambouropoulos and Kaliakatsos-Papakostas (2022) offer a nice review of features used to compare audio representations.
10. Examples include Hodgetts et al. (2009) and Hahn et al. (2003).
11. The reader is referred to Bourne (2015) for a more thorough accounting of these three stages.
12. Swett (2022) views Keller's theories through the lens of structure-mapping theory, providing detailed evidence that Keller was intuitively aware of the constraints and possibilities underlying spontaneous comparison of musical materials in real time.
13. In fact, Gjerdingen is following the principle of "progressive alignment" (Gentner and Calhoun 2010), which states that learning is maximized when early and easy comparisons (that is, of entities with high surface similarity) scaffold more relationally complex ones presented later.

C2S13 ## Works Cited

C2P71 Bartlett, James C., and W. Jay Dowling. 1988. "Scale Structure and Similarity of Melodies." *Music Perception: An Interdisciplinary Journal* 5/3: 285–314. https://doi.org/10.2307/40285401

C2P72 Boss, Jack. 1992. "Schoenberg's Op. 22 Radio Talk and Developing Variation in Atonal Music." *Music Theory Spectrum* 14/2: 125–149.

C2P73 Bourne, Janet. 2015. "A Theory of Analogy for Musical Sense-Making and Categorization: Understanding Musical Jabberwocky." PhD diss., Northwestern University.

C2P74 Brinkman, Andrew, Daniel Shanahan, and Craig Sapp. 2016. "Musical Stylometry, Machine Learning, and Attribution Studies: A Semi-Supervised Approach to the Works of Josquin." *Proceedings of the 14th Biennial International Conference on Music Perception and Cognition*: 91–97.

C2P75 Cambouropoulos, Emilios. 2001. "Melodic Cue Abstraction, Similarity, and Category Formation: A Formal Model." *Music Perception: An Interdisciplinary Journal* 18/3: 347–370. https://doi.org/10.1525/mp.2001.18.3.347

C2P76 Cambouropoulos, Emilios. 2009. "How Similar Is Similar?" *Musicae Scientiae* 13/1: 7–24. https://doi.org/10.1177/102986490901300102

C2P77 Cambouropoulos, Emilios, and Maximos Kaliakatsos-Papakostas. 2022. "Symbolic Approaches and Methods for Analyzing Musical Similarity: Representation and Pattern Processing in Harmony." In *The Oxford Handbook of Music and Corpus Studies*. Edited by Daniel Shanahan, John Ashley Burgoyne, and Ian Quinn. Oxford University Press. Online Edition. https://doi.org/10.1093/oxfordhb/9780190945442.013.9

C2P78 Caplin, William E. 1998. *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. New York: Oxford University Press.

C2P79 Conklin, Darrell, and Ian H. Witten. 1995. "Multiple Viewpoint Systems for Music Prediction." *Journal of New Music Research* 24/1: 51–73. https://doi.org/10.1080/09298219508570672

C2P80 Day, Samuel B., and Dedre Gentner. 2007. "Nonintentional Analogical Inference in Text Comprehension." *Memory and Cognition* 35/1: 39–49. https://doi.org/10.3758/BF03195940

C2P81 Eerola, Tuomas, and Micah Bregman. 2007. "Melodic and Contextual Similarity of Folk Song Phrases." *Musicae Scientiae* 11/1_suppl: 211–233. https://doi.org/10.1177/102986490701100109

C2P82 Eitan, Zohar, and Roni Y. Granot. 2007. "Intensity Changes and Perceived Similarity: Inter-Parametric Analogies." *Musicae Scientiae* 11/1_suppl: 39–75. https://doi.org/10.1177/1029864907011001031

C2P83 Forbus, Kenneth D., Ronald W. Ferguson, Andrew Lovett, and Dedre Gentner. 2017. "Extending SME to Handle Large-Scale Cognitive Modeling." *Cognitive Science* 41/5: 1152–1201. https://doi.org/10.1111/cogs.12377

C2P84 Gentner, Dedre. 1983. "Structure-Mapping: A Theoretical Framework for Analogy." *Cognitive Science* 7/2: 155–170. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1207/s15516709cog0702_3

C2P85 Gentner, Dedre, and Julie Colhoun. 2010. "Analogical Processes in Human Thinking and Learning." In *Towards a Theory of Thinking: Building Blocks for a Conceptual Framework*. Edited by Britt Glatzeder, Vinod Goel, and Albrecht Müller, 35–48. Cham, Switzerland: Springer.

C2P86 Gjerdingen, Robert O. 1988. "Shape and Motion in the Microstructure of Song." *Music Perception: An Interdisciplinary Journal* 6/1: 35–64. https://doi.org/10.2307/40285415

C2P87 Gjerdingen, Robert O. 2007. *Music in the Galant Style*. New York: Oxford University Press.

C2P88 Gjerdingen, Robert O. 2020. *Child Composers in the Old Conservatories: How Orphans Became Elite Musicians*. New York: Oxford University Press.

C2P89 Goldwater, Micah B. 2017. "Grammatical Constructions as Relational Categories." *Topics in Cognitive Science* 9/3: 776–799. https://doi.org/10.1111/tops.12272

C2P90 de Haas, W. Bas, Frans Wiering, and Remco C. Veltkamp. 2013. "A Geometrical Distance Measure for Determining the Similarity of Musical Harmony." *International Journal of Multimedia Information Retrieval* 2/3: 189–202.

C2P91 Hahn, Ulrike. 2014. "Similarity." *Wiley Interdisciplinary Reviews: Cognitive Science* 5/3: 271–280. https://doi.org/10.1002/wcs.1282

C2P92 Hahn, Ulrike, Nick Chater, and Lucy B. Richardson. 2003. "Similarity as Transformation." *Cognition* 87/1: 1–32. https://doi.org/10.1016/S0010-0277(02)00184-1

C2P93 Hanninen, Dora A. 2004. "Associative Sets, Categories, and Music Analysis." *Journal of Music Theory* 48/2: 147–218.

C2P94 Hatten, Robert S. 1994. *Musical Meaning in Beethoven: Markedness, Correlation, and Interpretation*. Bloomington: Indiana University Press.

C2P95 Heneghan, Áine. 2019. "Rethinking Repetition: Interrogating Schoenberg's Writings." *Perspectives of New Music* 57/1: 25–74. https://doi.org/10.1353/pnm.2019.0006

C2P96 Hepokoski, James, and Warren Darcy. 2006. *Elements of Sonata Theory: Norms, Types, and Deformations in the Late Eighteenth-Century Sonata*. New York and London: Oxford University Press.

C2P97 Hodgetts, Carl J., Ulrike Hahn, and Nick Chater. 2009. "Transformation and Alignment in Similarity." *Cognition* 113/1: 62–79.

C2P98 Huron, David. 1994. "The Humdrum Toolkit: Reference Manual." Center for Computer Assisted Research in the Humanities. https://www.humdrum.org/

C2P99 Huron, David. 2001. "What Is a Musical Feature? Forte's Analysis of Brahms's Opus 51, No. 1, Revisited." *Music Theory Online* 7/4.

C2P100 Isaacson, Eric J. 1990. "Similarity of Interval-Class Content between Pitch-Class Sets: The IcVSIM Relation." *Journal of Music Theory* 34/1: 1–28. https://doi.org/10.2307/843860

C2P101 Ivanovitch, Roman. 2010. "What's in a Theme? On the Nature of Variation." *Gamut: Online Journal of the Music Theory Society of the Mid-Atlantic* 3/1, Article 3. https://trace.tennessee.edu/gamut/vol3/iss

C2P102 Krumhansl, Carol L. 1979. "The Psychological Representation of Musical Pitch in a Tonal Context." *Cognitive Psychology* 11/3: 346–374.

C2P103 Krumhansl, Carol L., and Edward J. Kessler. 1982. "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys." *Psychological Review* 89/4: 334–368.

C2P104 Laitz, Steven G. 2016. *The Complete Musician: An Integrated Approach to Theory, Analysis, and Listening*. 4th ed. New York and Oxford: Oxford University Press.

C2P105 Lamont, Alexandra, and Nicola Dibben. 2001. "Motivic Structure and the Perception of Similarity." *Music Perception: An Interdisciplinary Journal* 18/3: 245–274. https://doi.org/10.1525/mp.2001.18.3.245

C2P106 Lehman, Frank. 2018. *Hollywood Harmony: Musical Wonder and the Sound of Cinema*. New York: Oxford University Press.

C2P107 Lerdahl, Fred. 2001. *Tonal Pitch Space*. Oxford: Oxford University Press.

C2P108 Lewin, David. 1979. "A Response to a Response: On PCSet Relatedness." *Perspectives of New Music* 18/1-2: 498–502. https://doi.org/10.2307/832999

C2P109 McAdams, Stephen, Suzanne Winsberg, Sophie Donnadieu, Geert De Soete, and Jochen Krimphoff. 1995. "Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities, and Latent Subject Classes." *Psychological Research* 58: 177–192.

C2P110 Medin, Douglas L., Robert L. Goldstone, and Dedre Gentner. 1993. "Respects for Similarity." *Psychological Review* 100/2: 254–278.

C2P111 Meyer, Leonard B. 1973. *Explaining Music: Essays and Explorations*. Berkeley: University of California Press.

C2P112 Mongeau, Marcel, and David Sankoff. 1990. "Comparison of Musical Sequences." *Computers and the Humanities* 24/3: 161–175.

C2P113 Morris, Robert. 1979. "A Similarity Index for Pitch-Class Sets." *Perspectives of New Music* 18/1-2: 445–460. https://doi.org/10.2307/832996

C2P114 Müllensiefen, Daniel, and Klaus Frieler. 2004. "Cognitive Adequacy in the Measurement of Melodic Similarity: Algorithmic vs. Human Judgments." *Computing in Musicology* 13: 147–176.

C2P115 Müllensiefen, Daniel, and Marc Pendzich. 2009. "Court Decisions on Music Plagiarism and the Predictive Value of Similarity Algorithms." *Musicae Scientiae* 13/1_suppl: 257–295. https://doi.org/10.1177/102986490901300111

C2P116 Murphy, Scott. 2020. "Abundant Novelty of Antitonic Harmony in the Music of Nikolai Myaskovsky." In *Analytical Approaches to 20th-Century Russian Music*: *Tonality, Modernism, Serialism*. Edited by Inessa Bazayev and Christopher Segall, 32–53. New York: Routledge.

C2P117 Narmour, Eugene. 2000. "Music Expectation by Cognitive Rule-Mapping." *Music Perception: An Interdisciplinary Journal* 17/3: 329–398.

C2P118 Pearce, Marcus, and Daniel Müllensiefen. 2017. "Compression-Based Modelling of Musical Similarity Perception." *Journal of New Music Research* 46/2: 135–155.

C2P119 Pollard-Gott, Lucy. 1983. "Emergence of Thematic Concepts in Repeated Listening to Music." *Cognitive Psychology* 15/1: 66–94.

C2P120 Quinn, Ian. 2001. "Listening to Similarity Relations." *Perspectives of New Music* 39/2: 108–158.

C2P121 Reti, Rudolph. 1951. *The Thematic Process in Music*. New York: Macmillan.

C2P122 Rings, Steven. 2011. *Tonality and Transformation*. New York: Oxford University Press.

C2P123 Rodin, Jesse, Craig Sapp, and Clare Bokulich. 2010. "Josquin Research Project." https://josquin.stanford.edu/

C2P124 Rosch, Eleanor. 1978. "Principles of Categorization." In *Cognition and Categorization*. Edited by Eleanor Rosch and Barbara B. Lloyd, 27–48. Hillsdale, NJ: Lawrence Erlbaum.

C2P125 Ruwet, Nicolas. 1966. "Méthodes d'analyse en Musicologie." *Revue Belge de Musicologie/Belgisch Tijdschrift voor Muziekwetenschap* 20/1-4: 65–90. https://doi.org/10.2307/3686642

C2P126 Schoenberg, Arnold. 1994. *Coherence, Counterpoint, Instrumentation, Instruction in Form = Zusammenhang, Kontrapunkt, Instrumentation, Formenlehre*. Edited and with an introduction by Severine Neff. Translated by Charlotte M. Cross and Severine Neff. Lincoln: University of Nebraska Press.

C2P127 Schoenberg, Arnold. 1995. *The Musical Idea and the Logic, Technique, and the Art of Its Presentation*. Edited, translated, and with a commentary by Patricia Carpenter and Severine Neff. New York: Columbia University Press.

C2P128 Spitzer, Michael. 2004. *Metaphor and Musical Thought*. Chicago and London: University of Chicago Press.

C2P129 Straus, Joseph N. 2016. *Introduction to Post-Tonal Theory*. 4th ed. New York: W. W. Norton.

C2P130 Swett, Nicky. 2022. "Functional Analogies: Learning by Comparison in Wordless Functional Analysis." Paper presented at the joint meeting of the American Musicological Society, Society for Ethnomusicology, and Society for Music Theory. November 11.

C2P131 Teitelbaum, Richard. 1965. "Intervallic Relations in Atonal Music." *Journal of Music Theory* 9/1: 72–127. https://doi.org/10.2307/843150

C2P132 Tversky, Amos. 1977. "Features of Similarity." *Psychological Review* 84/4: 327–352. https://doi.org/10.1037/0033-295X.84.4.327

C2P133 Velardo, Valerio, Mauro Vallati, and Steven Jan. 2016. "Symbolic Melodic Similarity: State of the Art and Future Challenges." *Computer Music Journal* 40/2: 70–83.

C2P134 Yust, Jason. 2020. *Organized Time: Rhythm, Tonality, and Form*. New York: Oxford University Press.

C2P135 Zbikowski, Lawrence M. 1999. "Musical Coherence, Motive, and Categorization." *Music Perception: An Interdisciplinary Journal* 17/1: 5–42.

C2P136 Zbikowski, Lawrence M. 2002. *Conceptualizing Music: Cognitive Structure, Theory, and Analysis*. New York: Oxford University Press.

C2P137 Zbikowski, Lawrence M. 2017. *Foundations of Musical Grammar*. New York: Oxford University Press.

C2P138 Ziv, Naomi, and Zohar Eitan. 2007. "Themes as Prototypes: Similarity Judgments and Categorization Tasks in Musical Contexts." *Musicae Scientiae* 11/1_suppl: 99–133. https://doi.org/10.1177/1029864907011001051