Why Do Skips Precede Reversals? The Effect of Tessitura on Melodic Structure

PAUL VON HIPPEL
Ohio State University and Stanford University

DAVID HURON
Ohio State University

In melodies from a wide variety of cultures, a large pitch interval tends to be followed by a change of direction. Although this tendency is often attributed to listeners' expectations, it might arise more simply from constraints on melodic range or tessitura. Skips tend toward the extremes of a melody's tessitura, and from those extremes a melody has little choice but to retreat by changing direction.

Statistical analyses of vocal melodies from four different continents are consistent with this simple explanation. The results suggest that, in the sampled repertoires, patterns such as "gap-fill," "registral direction," and "registral return" (L. Meyer, 1956, 1973; E. Narmour, 1990) are mere side effects of constraints on melodic tessitura.

Received December 1, 1998; accepted for publication, February 17, 2000.

For at least 400 years, students of Western music have been taught that a large melodic interval tends to be followed by a change of direction (e.g., Nanino & Nanino, ca. 1600; Fux, 1725/1943; Prout, 1890; Kostka & Payne, 1995). This tendency—which we call post-skip reversal—is illustrated by the German folk song in Figure 1. This melody contains 12 skips, and 10 of them are followed immediately by a reversal.

1. To prevent confusion, we should say that we are not using the term reversal in the technical sense proposed by Narmour (1990). When we say reversal, all we mean is that the melodic contour changes direction—from upward to downward, or from downward to upward.

Address correspondence to Paul von Hippel, School of Music, Ohio State University, 1866 College Ave., Columbus, OH 43210. (e-mail: von-hippel.1@osu.edu)
ISSN: 0730-7829. Send requests for permission to reprint to Rights and Permissions, University of California Press, 2090 Center St., Ste. 303, Berkeley, CA 94704-1223.
Post-skip reversal is sometimes viewed as a convention for late-Renaissance polyphony (e.g., Jepesen, 1931/1939), but, as Figure 1 suggests, the tendency has been observed in a wide variety of musical styles. For example, the first empirical study of post-skip reversal—performed in 1924 by Henry Watt—compared Schubert’s Lieder to folk songs by the Ojibway Indians of Wisconsin and Minnesota. Despite cultural differences, in each repertoire Watt found that, as the size of an interval increases, the chance of a subsequent reversal increases as well. Watt’s results for Schubert’s Lieder are summarized in Figure 2.

More recent measurements have found post-skip reversal in masses by Palestrina, chorale harmonizations by J. S. Bach, Lieder by Schumann and Schubert, and folk songs from Bohemia (Miller, 1992), Thompson & Stallman, 1995/1996, 1998; Yi, 1990). Moreover, similar tendencies have been observed informally in the music of northern India and the Arab

![Graph of proportion of intervals followed by a change of direction](image)

**Fig. 2.** In a study of Schubert’s Lieder, Watt (1924) found that larger intervals were more likely to precede a change of direction. He obtained similar results for the folk songs of the Ojibway Indians. (This plot is based on Watt’s Tables I and III. There is no data point corresponding to 11 semitones because there were no 11-semitone intervals in Watt’s sample.)

world (Fox Strangways, 1914; Lachmann & Fox Strangways, 1936). In short, it seems that post-skip reversal may be a universal property of melodic structure.

In this article, we review three theoretical accounts of post-skip reversal. Two of these accounts attribute the pattern to hypothetical laws governing listeners’ expectations (Meyer, 1956; Narmour, 1990). By contrast, a third account attributes the pattern to simple constraints on melodic range or tessitura (Watt, 1924). Using a variety of statistical methods, we compare and evaluate these accounts by analyzing vocal melodies from four different continents. To foreshadow the conclusions, we will find that accounts based on expectation do not seem necessary to explain post-skip reversal. Instead, the results fit best with the view that post-skip reversal arises from constraints on melodic tessitura.

**ACCOUNTS BASED ON EXPECTATION**

The best-known accounts of post-skip reversal refer to the expectations of listeners. After a skip, listeners from a variety of musical and cultural backgrounds tend to expect that a melody will change direction (Cuddy & Lunney, 1995; Krumhansl, 1995a, 1995b, Schellenberg, 1996, 1997; Schmuckler, 1989; Thompson, Cuddy, & Plaus, 1997). Thus post-skip reversals have been viewed as evidence that melodicists try first to arouse and then to gratify expectations—possibly as a way of engaging listeners’ thoughts and emotions (Meyer, 1956).

This account raises the question of why a skip would cause listeners to expect reversal. At least two cognitive theories have been advanced to explain this effect. The first explanation is Leonard Meyer’s (1956, 1973) theory of “gap-fill.” In brief, Meyer proposed that a skip creates a “structural gap” that listeners expect to hear “filled.” This proposal was anticipated by Fox Strangways (1914, p. 329), who wrote that “to pass over a note immediately creates a desire for it.” Meyer (1956, pp. 86–87, 128–135), however, developed the idea into an argument that relies on four cognitive hypotheses:

1. Listeners—at least “practiced or cultivated” listeners—infer the scale from which a melody’s pitches are drawn.
2. Listeners find a melody most “satisfactory” when it uses every degree of its scale. This claim is consistent with a preference for completeness—which Meyer attributes to the Gestalt law of Prägnanz, or good figure.
3. When a melody’s use of its scale is incomplete—that is, when one or more scale degrees are omitted—listeners’ dissatisfaction persists for some time after the omission.
4. This dissatisfaction is relieved when the missing scale degrees finally occur.
There is now some experimental evidence that trained listeners can infer the scale of a brief musical fragment (Cohen, 1991). The other links in Meyer’s argument, however, remain largely untested.

An alternative cognitive theory was proposed by Meyer’s onetime student, Eugene Narmour (1990). In Narmour’s theory of “implication and realization,” post-skip reversals are attributed to two distinct principles: registral direction and registral return. According to the principle of registral return, listeners expect an interval to be opposite in direction, but similar in size, to the interval that preceded it. This pattern has been attributed to the Gestalt rule of similarity (Narmour, 1990). Another way to describe registral return is that listeners expect an interval to land close to where the previous interval began. Seen in this way, registral return has been ascribed to the Gestalt rule of proximity (Krumhansl, 1993a).

Narmour’s second principle, registral direction, relies on a distinction between large and small intervals. After a small interval, the rule predicts that listeners expect further motion in the same direction. By contrast, after a large interval, the rule predicts that listeners expect a change of direction. In Gestalt terms, the first prediction fits the law of good continuation, whereas the second prediction fits a law of “symmetry or near-symmetry” (Krumhansl, 1993a, p. 221).

From a conventional Gestalt perspective, there is no reason why a different law should apply after large intervals than after small ones. To explain this discontinuity, Narmour has proposed two novel hypotheses. According to the first hypothesis, when successive events are similar (as when the pitches forming an interval are close together), listeners expect further similarity (e.g., further intervals in the same direction). According to the second hypothesis, when successive events are different (as when the pitches forming an interval are far apart), listeners expect further difference (e.g., a change of direction).

With respect to post-skip reversal, the first of these hypotheses may be unnecessary. Cognitive experiments have not generally indicated that small intervals cause listeners to expect continued motion in the same direction. On the contrary, some results suggest that listeners expect reversal after large and small intervals alike (Schmuckler, 1989). In a simplification of Narmour’s model, Schellenberg (1997) has derived a principle that treats interval size more continuously. According to Schellenberg’s rule of “pitch reversal,” listeners expect reversal after intervals of any size—although the expectation is stronger after a large interval than after a small one.

To summarize, Narmour and Meyer have explained post-skip reversals as an attempt to satisfy listeners’ expectations. These expectations, in turn, have been attributed to Gestalt and Gestalt-like rules of pattern formation.

Although there has been some disagreement over the precise form of these rules, all of this disagreement has taken place within a common Gestalt-influenced framework.

Like any venerable tradition, Gestalt psychology is subject to a number of venerable criticisms. The most relevant criticism here is that Gestalt rules can be overly flexible in their application (Goldstein, 1996, pp. 190–193; Pomerantz, 1986, pp. 7–8). There are a large number of Gestalt rules—particularly under the heading of Pragmatics—and if concepts such as similarity, symmetry, and completeness are defined broadly enough, a rule can be found to explain nearly any pattern of results. In the case of melodic structure, for example, Gestalt principles fit post-skip reversal no better than they would fit the opposite tendency. That is, if melodies tended not to reverse after a skip, their structure would fit the Gestalt rule of good continuation. In short, the case for post-skip reversal rests not so much on theoretical grounds as on the pattern’s remarkable consistency in the structure of melodies themselves.

**AN ACCOUNT BASED ON TESSITURA**

As we remarked earlier, the first empirical demonstration of post-skip reversal was carried out by Scottish psychologist Henry J. Watt (1924). Watt also proposed a simple explanation for the pattern, and his explanation did not refer at all to listeners’ expectations. Instead, Watt proposed that post-skip reversals arise from constraints on melodic range.²

Watt’s argument can be illustrated by a simplified melody that is confined to a range of three adjacent pitches—for example, A, B, and C. In such a melody, the only skip available is between A and C. Upward, this skip must land at the top of the range; downward, it must land at the bottom. After a skip, therefore, there is no way for a melody to continue moving in the same direction. On the contrary, two of the three available pitches can be reached only by reversal.

Although most melodies have a wider range, they are subject to the same basic argument. Every melody has a limited range of pitches, and many melodies seem to favor the center of that range (von Hippel, 2000). In describing a melody’s pitch distribution, therefore, it is prudent to deemphasize the likelihood of pitches that are very high or very low. In musical terms, this emphasis on the heart of the range is expressed by the word **tessitura**.

² The tendency toward post-skip reversals, Watt wrote, “doubtless finds its raison d’être in the limitations of range of the human voice or of any instrument. Having moved in one direction, the voice is more likely to move or to have moved just previously in the other direction. Its average position will presumably be the centre of its range. The greater the leap in one direction, the greater will be the probability of its moving in the opposite direction” (Watt, 1924, p. 386).
Melodic Sample

In the interest of replicating earlier results, we began by sampling the repertoires used in Watt's (1924) study of post-skip reversals. Specifically, we chose 42 melodies at random from a collection of Ojibway folk songs (Densmore, 1910, 1913; computer data files available in von Hippel, 1998) as well as the vocal lines from a collection of 35 Lieder by Schubert (Friedlaender, n.d.). To broaden the study's cultural scope, we also chose a popular collection of 24 folk songs from South Africa (Makeba, 1971), the first 30 folk songs in a collection from China (Chung-kuo yin yueh yen chiu so [Chinese Music Research Institute], 1959), and a random sample of 80 European folk songs from the Essen Folksong Collection (Schaffrath, 1995) and the Child anthology of British folk ballads (Bronson, 1959).

This European sample consisted of four songs chosen at random from each of 20 different regions—including Alsace, Austria, Britain, former Czechoslovakia, Denmark, France, Germany, Italy, Lorraine, Luxembourg, Holland, Hungary, Romania, Poland, Sweden, Switzerland, the Tyrol, former Yugoslavia, the Ukraine, and other parts of the former Soviet Union. Unfortunately, this European sample is not as diverse ethnically as it is geographically. Although the Essen Folksong Collection includes melodies from regions not under German or Austrian government, in many of those regions the collection is biased toward the songs of ethnic Germans.

In the Ojibway and South African folk songs, as in the Lieder by Schubert, the notation often indicates that a stretch of melody should be sung more than once. Such repeated sections present a dilemma for data collection. On the one hand, including a section more than once reduces the independence of the sampled data. On the other hand, including a repeated section only once can violate a melody's structure. In the Ojibway folk songs, for example, a crucial interval is often formed when the last note of one section is followed by the first note of an earlier section.

For the analyses in this article, we chose to preserve musical structure by sampling repeated sections as many times as was indicated in the score. To compensate for the resulting loss of independence, we designed one set of analyses (Study 2) so that long melodies—including melodies lengthened by repetition—would carry no more weight than short melodies.

Study 1: Classifying Skips

Watt's (1924) explanation for post-skip reversal depends in part on the observation that a melody tends to retreat from the extremes of its tessitura (see von Hippel, 2000, for a demonstration). Operationally, we might say that a melody tends toward the bulk of its pitch distribution—in other
words, that it tends toward its median pitch. With respect to the median pitch, skips may be classified into four types. These types are illustrated in Figure 3:

- **Median-departing skips** move away from the median without crossing it.
- **Median-crossing skips** cross from one side of the median to the other.
- **Median-landing skips** land precisely on the median.
- **Median-approaching skips** move toward the median without reaching it.

By classifying skips in this way, we can distinguish competing hypotheses about post-skipping reversal. In particular, hypotheses based on expectation would predict that any skip, wherever it lands in the tessitura, tends to be followed by a reversal. By contrast, a hypothesis based on tessitura would predict only reversals that take a melody toward its median pitch. That is, a tessitura hypothesis would predict reversals only after skips that depart or cross the median—not after skips that approach the median or land on it. More specifically, after a median-approaching skip, a tessitura hypothesis would predict a continuation in the same direction. And after a median-landing skip, a tessitura hypothesis would predict a balance of tendencies—that is, an equal likelihood for continuation or reversal.

![Diagram](https://example.com/diagram.png)

**Fig. 3.** With respect to the median, skips may be classified into the four types shown above. If post-skipping reversal arises from constraints on tessitura, only median-departing and median-crossing skips would tend to precede reversals. A median-approaching skip would tend to initiate continued motion in the same direction. And a median-landing skip would be balanced in its tendencies—equally likely to precede reversals or continuations. (Only upward skips are displayed here. The classification, however, applies to downward skips as well.)

---

5. The median pitch should not be confused with the median or third scale degree.

6. Among the sampled repertoires, only the European folk song contains explicitly notated phrase marks. The other repertoires contain other clues to phrasing—such as harmony, contour, regularity, note duration, and rests. Except in the case of rests, however, these other clues were not used in our definition of interval.

7. Our threshold for statistical significance is $p < .05$ (two-tailed). However, values near this threshold should be taken with a grain of salt. This article contains 122 statistical tests, many of which relate to quite similar hypotheses. In the context of such multiple tests, we should be skeptical of a value that would occur by chance 1 time in 20. It seems likely that this article contains a number of type 1 errors; that is, some results may indicate a significant excess of post-skipping reversals where in fact there is none.

Under ordinary circumstances, we would reduce these type 1 errors by a correction for multiple tests (see, e.g., Darlington, 1990). However, such a correction would increase the chance of a type 2 error; that is, we would be more likely to overlook a significant excess of post-skipping reversals when it is actually present. Because we interpret change levels of post-skipping reversal as evidence for the tessitura hypothesis, we wish to avoid such type 2 errors as well. We have therefore chosen not to correct for multiple tests. Had we used a correction, our results would appear even more consistent with the tessitura hypothesis.
should be kept in mind that 18 of the 20 results in this table fit a consistent interpretation. On balance, that is, the results fit best with the claim that post-skip reversals arise from constraints on tessitura.

These results depend on the assumption that a skip is any interval larger than 2 semitones. This threshold, however, may be too small for some of the sampled repertoires. Chinese and Ojibway folk songs, for example, often use a pentatonic scale in which certain scale degrees are 3 semitones apart. In such a scale, a 3-semitone interval does not skip over any pitches; a 3-semitone interval cannot be a gap, in Meyer's (1956) sense, because there would be no way to fill it. Even in Western melodies, the proper skip threshold may be quite large. According to Narmour's (1990) principle of registral direction, listeners expect a reversal only when an interval exceeds 6 semitones (although this threshold may vary according to the musical context).

It would be easy enough to repeat the analyses using a threshold of 6 semitones. With such a high threshold, however, there are very few skips that approach the median or land on it. In fact, excluding skips that are followed by a repeated note, there are just 100 such skips in all the samples put together. These skips, moreover, are not distributed evenly across the sampled melodies. For example, of the 30 median-approaching skips, 11 occur in the sample of Schubert's Lieder, and 6 occur in a single Lieb ("Dithyrambe"). On closer inspection, these 6 skips boil down to just 2 skips in a strophe that is repeated three times. With so little data exhibiting so little independence, it is doubtful that we could make valid inferences about the sampled repertoires.

### Study 2: Multiple Regression

The difficulties encountered in our first study highlight the problems that result from using skip thresholds, as well as problems that result from treating the individual skip as the unit of analysis. To avoid these problems, in our second study we will treat the individual melody as the unit of analysis, and we will avoid using skip thresholds entirely.

The avoidance of skip thresholds may not hurt our description of melodic structure. Inspection of Figure 1 does not suggest a threshold where intervals become dramatically more likely to precede reversals. Instead, the likelihood of reversal seems to increase continuously with interval size. The effect of tessitura is continuous as well: as a pitch grows more extreme, the following interval grows more likely to retreat toward the center of the tessitura (von Hippel, 2000).

---

8. Despite his use of a skip threshold, Narmour occasionally suggests that the relationship between interval size and reversal is continuous. (See, for example, Narmour, 1992, p. 18, Figure 1.2).
To model these relationships formally, we represented each melodic interval with a signed number; the sign of this number represents the interval's direction (upward or downward), and the magnitude represents the interval's size in semitones. We also used signed numbers to represent melodic pitches; here the size of the number represents distance from a melody's median pitch, and the sign of the number indicates whether a pitch is above the median or below it. We call this number our measure of pitch extremity.

In terms of these measurements, we can define post-skip reversal as a negative linear correlation between successive intervals. After an upward interval, especially a large one (e.g., +9 semitones), this definition would predict a smaller downward interval (e.g., −2 semitones). Likewise, after a downward interval, especially a large one, the definition would predict a smaller upward interval. The effect of tessitura constraints can be defined in similar terms, as a negative linear correlation between the size of an interval and the extremity of its starting pitch. After a high pitch (an extremely high pitch might be +9 semitones above the median) this definition would predict a downward interval (e.g., −2 semitones), while after a low pitch, this definition would predict an upward interval.

Figure 5 illustrates these correlations for the pitches and intervals of the German folk song that appeared in Figure 1. In Figure 5, the left panel plots each interval against the interval that precedes it. (As in our first study, we have restricted the sample to interval successions that do not cross phrase marks or rests.) Although the correlation between successive intervals is not significant, its negative sign (i.e., its downward slope) is consistent with a tendency toward post-skip reversals, \( r(30) = -0.25, p = 0.17 \). Similarly, the right side of Figure 5 plots each interval against the extremity of its starting pitch. Here the significantly negative correlation provides evidence for the effect of tessitura constraints, \( r(30) = -0.41, p = 0.02 \). All 211 of the sampled melodies exhibit both of these negative correlations.

These reliable correlations indicate that interval size can be predicted by two different variables: (1) the extremity of the interval's starting pitch and (2) the size of the previous interval. These predictors, however, are far from independent. Because large intervals tend to land on extreme pitches, the predictive value associated with the previous interval may be only a side effect of the predictive value associated with pitch extremity.

To measure the unique contribution of each predictor, we performed a multiple regression analysis in which both predictors were used to estimate interval size in our illustrative German folk song. The analysis found that, when pitch extremity is controlled, there remains only a tiny semipartial correlation between successive intervals, \( sr(29) = -0.06, p = 0.73 \). By contrast, when the previous interval is controlled, the semipartial correlation between interval size and pitch extremity is over five times as large, \( sr(29) = -0.41, p = 0.02 \).
Finally, when both predictors are used together, their multiple correlation with interval size, \( R(29) = .42, p = .06 \), is only marginally stronger than the simple correlation between interval size and pitch extremity alone, \( r(30) = -.41, p = .02 \). In sum, the results suggest that the previous interval makes little or no unique contribution to the accuracy of the prediction.

Because this analysis predicted only 32 intervals from a single melody, its results merely border on statistical significance. Analyses of the other melodies, however, fit the same general pattern. First, the unique contribution of pitch extremity is highly reliable. For every one of the 211 sampled melodies, the extremity of an interval’s starting pitch has a negative semipartial correlation with the interval’s size. In contrast, the semipartial correlations between successive interval sizes are not reliably negative. No consistent sign is evident in the Chinese folk songs (19 positive, 11 negative, sign-test \( \chi^2 = 1.6, p = .20 \)), the Ojibway folk songs (18 positive, 24 negative, sign-test \( \chi^2 = .6, p = .44 \)), or the Lieder by Schubert (20 positive, 15 negative, sign-test \( \chi^2 = 0.5, p = .50 \)). In the South African folk songs, a significant majority of these semipartial correlations are negative (6 positive, 18 negative, sign-test \( \chi^2 = 9.0, p = .02 \)), but in the European folk songs, a significant majority are positive (26 positive, 24 negative, sign-test \( \chi^2 = 12.0, p = .0005 \)). On balance, the results suggest that, after controlling for the effect of tessitura, there is no reliable tendency for larger intervals to precede reversals.

These analyses presume that any correlation between successive intervals would be linear—that is, that the tendency toward reversal would increase in simple proportion to the size of the previous interval. This assumption may be simplistic. One possible complication is that the tendency toward reversal could depend on an interval’s direction. Analyses by Eitan (1997) suggest that listeners are more likely to expect reversal when the previous interval moves downward. If melodies were designed to satisfy such expectations, downward intervals would be more likely to precede reversals. In our regression analysis, this possibility can be crudely modeled by adding a predictor that represents the squared size of the previous interval. If the sign of this squared term is positive, while the sign for the unsquared previous interval remains negative, the resulting curve might be consistent with the asymmetry observed by Eitan: as illustrated in Figure 6, the slope describing reversal would be steeper following downward intervals. Alternatively, if the squared term were negative, the slope would be steeper following upward intervals.

To test this possibility, we performed a second set of analyses in which interval size was regressed against three predictors—one predictor representing the extremity of the interval’s starting pitch and two predictors (one squared, one linear) representing the size of the previous interval. If

**Fig. 6.** If reversals were more common after downward skips, the relationship between successive intervals could be modeled by a quadratic equation with a positive squared term and a negative linear term. When tessitura is controlled, however, there is no consistent evidence for such a relationship. (The precise equation plotted here, \( y = .007x^2-.25x \), is used for illustration only. A variety of quadratic equations would have a similar contour.)
term were negative, while the sign of the uncubed previous interval became positive, the resulting curve would have the general shape displayed in Figure 7. For large predictor intervals, the curve would slope downward—indicating a tendency toward reversals—but for small predictor intervals, the curve would slope upward—indicating a tendency to continue in the same direction.

To test this possibility, we performed a final set of analyses in which interval size was once more regressed against three predictors—one predictor representing the extremity of the interval’s starting pitch and two predictors (one cubed, one linear) representing the size of the previous interval. If the data fit the hypothesized curve, the cubed term would be consistently negative. Again, however, no such consistency was evident. Although the European folk song analyses returned a significant majority of negative cubed terms (22 positive, 58 negative, \( \chi^2 = 15.3, p = .0001 \)), no consistent sign was evident for the cubed terms in the Chinese folk songs (13 positive, 17 negative, \( \chi^2 = 0.3, p = .58 \)), the Ojibway folk songs (23 positive, 19 negative, \( \chi^2 = 0.2, p = .64 \)), the South African folk songs (15 positive, 9 negative, \( \chi^2 = 1.0, p = .31 \)), or the Lieder by Schubert (15 positive, 20 negative, \( \chi^2 = 0.5, p = .50 \)). In short, the analyses contradicted both halves of the principle of registral direction. When tessitura constraints were controlled, the results did not suggest that skips precede reversals, nor did they suggest that small intervals initiate further motion in the same direction.

More generally, when controls were placed on tessitura, our multiple regression analyses failed to verify any consistent function relating successive intervals—whether the function considered was linear, quadratic, or cubic. This failure corroborates our earlier results suggesting that post-skip reversals are a mere side effect of constraints on tessitura.

On a more positive note, the results suggest that tessitura constraints are a useful predictor of melodic intervals. As an index of this predictor's reliability, we might remind ourselves that every one of the 211 folk songs exhibits a negative correlation between the size of an interval and the extremity of its starting pitch (\( p < 10^{-6} \)). As an index of predictive strength, we can measure the size of this correlation for each sampled folk song. For the Ojibway folk songs, the median value of this correlation is \(-.30\); for the South African folk songs, it is \(-.41\); for the European folk songs, as for Schubert's Lieder, it is \(-.45\); and for the Chinese folk songs, the median correlation is \(-.53\).

**Study 3: Markov Twins**

The analyses so far have focused on the claim that skips precede reversals. Perhaps this claim is too broad. To find distinctive aspects of melodic structure, we may have to test claims of a more specific kind. For example, we could test the textbook rule that a skip tends to be followed by a step in the opposite direction (e.g., Nanino & Nanino, ca. 1600; Roberts & Fischer, 1967); this might be called the rule of contrary step. Alternatively, we could test whether melodies fill gaps, in Meyer's (1956, 1973) sense, that a skip tends to be followed by one of the notes that was skipped over. Finally, we could test Narmour's (1990) rule of "registral return," in which a skip (or smaller interval) tends to be followed by a return close to the skip's first note.

In order to test these specific claims, we need to know how melodies would look if the claims were false. Given constraints on tessitura and interval size, that is, melodies will simply by chance display a certain level of gap-fill, contrary step, and registral return. Only if these patterns exceed chance levels should we accept the related claims as principles of melodic structure. It is desirable, therefore, to contrast actual melodic practice with a control sample of "melodies" that ignore the hypothesized rules. The control melodies should resemble the originals in their distribution of pitches and intervals, but should be random in their interval successions. If the original and control melodies display similar levels of gap-fill, contrary step, and registral return, the patterns cannot be attributed to a rule for interval succession. Instead, they can only be a side effect of constraints on interval size and direction, notably the constraint on tessitura.
Suitable control melodies can be generated by a Markov model. The model begins by stripping a melody down to a series of tokens representing pitches, rests, and notated phrase marks (II), along with a token indicating that the melody has come to an end. In this format, the German folk song from Figure 1 would appear as follows:

\[
\]

Taking this representation as input, the model calculates the transition probabilities between successive tokens. Given the melody above, for example, the model finds that, following the pitch A, 13 different tokens might occur. Of these, the most likely are G, and A, which each occur 3 times, and so are assigned a probability of 23% each. Other possibilities include F, A, G, with a probability of 15%, and B, D, F, rest, or \{\}, with probabilities of 8% each. Because no other tokens follow any of the A's in this melody, all remaining tokens are assigned a probability of 0%.

Using these transition probabilities, the model generates a new melody, which we call a Markov twin. The Markov twin begins with the original melody's first pitch (along with any preceding phrase mark), then randomly chooses the next token according to the model's transition probabilities. In the case of the German folk song above, the Markov twin's first pitch would be A, and its next token would most likely be G, or A, though the other tokens listed above could also appear. Assuming that the next token were G, the token after that would be chosen at random from the tokens that followed G in the original melody, with greater probabilities for tokens that followed G more often. The model continues in this manner until it generates the "end" token. A Markov twin for our German folk song is printed below and notated in Figure 8:

\[
\{ A, A, F, E, D, C, B, G, A, A, G, A, \} \quad \{ E, F, D, A, C, F, A, D, \}
\]

This is just one of the Markov twins that can be derived from the given folk song. Unless the source melody is exceptionally predictable or short, an unlimited number of Markov twins are possible—although some are more likely than others. Ideally, we would like to contrast each original melody with the infinite set of Markov twins that can be derived from it. As a practical matter, however, we chose to generate 10 twins for each of the 211 sampled melodies.

The Markov model we have described is based on "first-order" transitions; that is, the choice of each token is constrained only by the token before it. Using such a model, it is impossible to describe a rule for post-

skip reversal, in which the choice of a pitch would be conditioned on the previous two pitches. Nevertheless, a Markov twin is likely to exhibit many post-skip reversals. A Markov twin inherits the input melody's pitch distribution—including its tessitura—and also preserves the input melody's first-order tendency to retreat from its tessitura's extremes. The patterns in a Markov twin, therefore, indicate how common post-skip reversals would be if tessitura were constrained but interval succession were not.

In Figure 9, we have plotted interval size against the probability of a subsequent reversal—much as Henry Watt did in 1924. Unlike Watt, however, we have plotted the relationship not only for Lieder and folk songs (represented by thick lines joining filled circles), but also for comparable sets of Markov twins (represented by thin lines joining open circles). If the original melodies were more likely to change direction after large intervals, we would expect all of the thick lines to rise above the thin lines as the graph proceeds from left to right. Instead, however, the thin lines seem just as likely to rise above the thick. In general, the lines follow one another closely; what divergence there is could be random variability due to the selectivity of certain intervals. In short, the figure does not suggest reliable differences between the original melodies and their Markov twins.

As a formal test, we can once again define a skip as an interval larger than 2 semitones, and ask in which melodies—the originals or their Markov twins—a skip is more likely to be followed by an immediate reversal. The results, displayed in Table 1, do not suggest that post-skip reversals are more common in the original melodies. To the contrary, three of the five comparisons do not reveal any significant differences at all. Moreover, of the two significant differences, only one suggests that reversals are more common in the original melodies; the other indicates that they are more common in the Markov twins. Because no rule was needed to generate post-skip reversals in the Markov twins, it would seem that no rule is needed to explain them in folk songs and Lieder.
As we remarked earlier, it may be that a skip threshold of 2 semitones is unduly small. In order to accommodate this possibility, we can repeat the analysis with a skip threshold of 6 semitones. The results, displayed in Table 2, are similar to those in Table 1—in fact, Table 2 provides even less evidence that the original repertoires are shaped by a rule for post-skip reversal. None of the five original repertoires has a significantly stronger tendency toward post-skip reversals than do the corresponding Markov twins.

In light of earlier results, these comparisons are not particularly surprising. We have already seen a number of results suggesting that tessitura constraints are the cause of post-skip reversals. The purpose of generating Markov twins was to test more specific hypotheses relating to gap-fill, contrary step, and registral return. To this end, we can define gap-fill as a skip followed by one of the pitches that has been skipped over. We can define contrary step as a skip followed by a 1- or 2-semitone interval in the opposite direction. And we can define registral return as a skip followed by a note within 2 semitones of the interval's starting pitch. (This is essentially Narmour's definition of registral return, except that Narmour applies it to small intervals as well as large ones.) Tables 3A and 3B measure the levels of gap-fill, contrary step, and registral return in the original melodies and their Markov twins. The upper table (3A) uses a skip threshold of 2 semitones, while the lower table (3B) uses a threshold of 6 semitones. Whichever threshold is used, and whichever type of reversal is considered, the results display the same general pattern: most
of the differences are insignificant. Furthermore, of the five significant differences in the tables, only one suggests that post-skip reversals are more common in the original melodies; the remaining four suggest that post-skip reversals are more common in the Markov twins. In short, the tables suggest that post-skip reversals in the original melodies are about as common as would be expected by chance.

To this point, all of our tests have focused on reversals that follow skips immediately. Perhaps this focus is too confining. Although many theorists restrict their attention to immediate reversals, Meyer and Narmour allow for delayed reversals as well. In fact, the suspense attending delay plays an important role in Meyer's (1956) theory of musical aesthetics.

In addition, both Meyer and Narmour suggest that post-skip reversals are not confined to a melody's low-level progress from note to note. At times, that is, post-skip reversal may occur at a higher level of a melody's hierarchical structure. A test for delayed reversal may be considered a crude test for certain types of high-level reversal as well.

Testing for delayed reversal can be problematic, as Lake (1987) has remarked, almost any melody will change direction if the listener waits long enough. It is therefore necessary to define limits on delay. Although any limits we set will be arbitrary, there is some merit in defining plausible limits a priori. To guide our definition, we can begin with Narmour's (1992, p. 269) claim that an interval can shape expectations for two or three times the length of the interval's second note. Unfortunately, because the Markov twins lack information on note length, we cannot use Narmour's definition in its original form. We can, however, slightly adapt it. Under our modified definition, a reversal follows a skip if it occurs within the next three notes. In addition, for reasons discussed earlier, the reversal must occur before the next phrase mark or rest.

Using this definition, we repeated our analyses, comparing the original melodies and their Markov twins with respect to simple reversal, gap-fill, contrary step, and registral return. We made these comparisons twice—once using a skip threshold of 2 semitones, once using a threshold of 6 semitones. All of the analyses included delayed reversals as well as immediate reversals. The results are summarized in Tables 4A and 4B.

The pattern in these tables should be familiar by now. Most of the differences are insignificant, and of the 9 significant differences, 4 suggest that post-skip reversals are more common in the Markov twins. In short, it seems that, given constraints on tessitura and interval size, even delayed post-skip reversals are no more common than would be expected by chance.

**Discussion**

It is sobering to find that post-skip reversals are just as common in random melodies as in folk songs and Lieder. This does not mean, of course, that folk songs and Lieder are random. Instead, it means that post-skip reversals will occur in any melody unless the composer has taken steps to avoid them. That is, it does not seem necessary to explain post-skip reversals in terms of listeners' expectations. An explanation that relies on tessitura constraints is largely sufficient.

This conclusion applies only to the sampled vocal repertoires. Other repertoires may well contain an excess of post-skip reversals. Given the centuries of Western composers who have been taught to follow a skip with a reversal, it seems inevitable that some of them have actually done so. If so, however, post-skip reversal would be specific to the style of specific composers; the pattern would not indicate a general principle of music
<table>
<thead>
<tr>
<th>Repertoire</th>
<th>No. of Intervals</th>
<th>Percentage of Intervals Followed by</th>
<th>Simple Reversal</th>
<th>Gap-Fill</th>
<th>Contrary Step</th>
<th>Regressal Return</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original Melodies</td>
<td>Markov Twins</td>
<td>Original Melodies</td>
<td>Markov Twins</td>
<td>Original Melodies</td>
<td>Markov Twins</td>
</tr>
<tr>
<td>Schubert's Lieder</td>
<td>1,581</td>
<td>16,747</td>
<td>88%</td>
<td>85%</td>
<td>56%</td>
<td>59%</td>
</tr>
<tr>
<td>European folk</td>
<td>806</td>
<td>6,905</td>
<td>85%</td>
<td>57%</td>
<td>53%</td>
<td>55%</td>
</tr>
<tr>
<td>songs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese folk</td>
<td>872</td>
<td>2,211</td>
<td>90%</td>
<td>32%</td>
<td>47%</td>
<td>44%</td>
</tr>
<tr>
<td>South African</td>
<td>1,526</td>
<td>12,246</td>
<td>83%</td>
<td>80%</td>
<td>42%</td>
<td>40%</td>
</tr>
<tr>
<td>folk songs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ojibway folk</td>
<td>643</td>
<td>1,130</td>
<td>68%</td>
<td>17%</td>
<td>26%</td>
<td>24%</td>
</tr>
<tr>
<td>songs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B. Following Intervals Larger Than 6 Semitones

<table>
<thead>
<tr>
<th>Repertoire</th>
<th>No. of Intervals</th>
<th>Percentage of Intervals Followed by</th>
<th>Simple Reversal</th>
<th>Gap-Fill</th>
<th>Contrary Step</th>
<th>Regressal Return</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original Melodies</td>
<td>Markov Twins</td>
<td>Original Melodies</td>
<td>Markov Twins</td>
<td>Original Melodies</td>
<td>Markov Twins</td>
</tr>
<tr>
<td>Schubert's Lieder</td>
<td>336</td>
<td>3,649</td>
<td>93%</td>
<td>93%</td>
<td>76%</td>
<td>63%</td>
</tr>
<tr>
<td>European folk</td>
<td>142</td>
<td>1,058</td>
<td>93%</td>
<td>87%</td>
<td>49%</td>
<td>53%</td>
</tr>
<tr>
<td>songs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese folk</td>
<td>151</td>
<td>1,445</td>
<td>96%</td>
<td>83%</td>
<td>43%</td>
<td>48%</td>
</tr>
<tr>
<td>South African</td>
<td>238</td>
<td>2,464</td>
<td>82%</td>
<td>71%</td>
<td>43%</td>
<td>48%</td>
</tr>
<tr>
<td>folk songs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ojibway folk</td>
<td>65</td>
<td>653</td>
<td>75%</td>
<td>52%</td>
<td>28%</td>
<td>33%</td>
</tr>
<tr>
<td>songs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05, **p < .01, ***p < .001, ****p < .0001 (χ²). Differences not marked with asterisks are not statistically significant.

**Condition**

Despite these reservations, our results fit remarkably simple interpretations. In the simple repetitions, it seems that post-anticipations arise from contour completion. After a skip that is a single pitch, there is no evidence that post-anticipations shape melodic structure. Instead, in this case, it seems more likely that melodic structure influences post-anticipations, rather than the other way around. This is consistent with studies of music cognition, which suggest that expectations are not static. Expectations are influenced by the context in which they occur. When a skip is approached, listeners may expect to hear a melody to continue, whether or not the pitch was approached by a skip. In other cases—whether or not the pitch was approached by a skip—listeners may expect to hear a melody to repeat, whether or not the pitch was approached by a skip.
post-skip reversal or to its underlying cause— it seems that some expectations can ultimately be traced to constraints on melodic tessitura.9

References


9. We thank the Center for Computer-Assisted Research in the Humanities for technical support and a congenial work environment. We also thank Jonathan Berger, Chris Chatfield editor Jay Dowling, and two anonymous reviewers for helpful comments on drafts of this article.


