Scale Structure and Similarity of Melodies
Author(s): James C. Bartlett and W. Jay Dowling
Source: *Music Perception: An Interdisciplinary Journal*, Vol. 5, No. 3, Cognitive and Perceptual Function (Spring, 1988), pp. 285-314
Published by: University of California Press
Stable URL: http://www.jstor.org/stable/40285401
Accessed: 04-12-2015 23:43 UTC

Music Perception
Spring 1988, Vol. 5, No. 3, 285–314

Scale Structure and Similarity of Melodies

JAMES C. BARTLETT & W. JAY DOWLING
University of Texas at Dallas

Four experiments explored an asymmetry in the perceived similarity of melodies: If a first-presented melody is "scalar" (conforms to a diatonic major scale), and is followed by a second melody slightly altered to be "nonscalar" (deviating from a diatonic major scale), subjects judge similarity to be lower than if the nonscalar melody comes first. Experiment 1 produced evidence that asymmetric similarity is *not* due simply to more strongly scalar melodies having greater memorability. Experiment 2 ruled out the hypothesis that asymmetric similarity depends on a task-specific strategy reflecting demand characteristics. Experiments 3 and 4 replicated asymmetric similarity while controlling the number of one-semitone intervals in scalar versus nonscalar melodies. The data are consistent with Garner's principles that stimuli are perceived with reference to sets of alternatives and that good stimuli have few alternatives. Specifically, when melodies are presented in scalar–nonscalar order, but not when presented in nonscalar–scalar order, the first melody evokes a small set of alternatives which the second melody often violates.

Introduction

A basic fact about melodies is that their tonality, and in particular their use of a restricted set of notes from a diatonic scale, is a highly salient attribute. Even the most nonmusical listener will notice "wrong notes" in atonal melodies, and will judge such melodies as less "pleasant" than strongly tonal melodies. Some aspects of tonality are enormously subtle, and probably perceivable only by experts. However, the conformity of a melody to a diatonic scale—what can be termed its "scalar structure" (Cross, Howell, & West, 1985)—appears to be detectable by virtually all listeners brought up in Western culture, even at an early age.

The detection of scalar structure of melodies has been supported by several investigators (see Dowling & Harwood, 1986), including Cross, Howell, & West (1983). Cross et al. generated artificial melodies that varied systematically in degree of scalar structure, and they found that musician as well as nonmusician subjects gave generally higher "liking" ratings to the

Requests for reprints may be sent to Dr. James C. Bartlett, Program in Human Development & Communication Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688.

more scalar melodies. Indeed, the ratings of musicians and those of nonmusicians were not detectably different.[1]

A related study by Krumhansl and Keil (1982) had adults and young children aged 6 to 12 years rate melodies for "goodness." The findings suggested that children's sense of tonality, and in particular their sense of which scalar notes were particularly suitable for ending a melody, grew more refined with age. However, even the youngest (6-year-old) subjects gave higher goodness ratings to scalar melodies than to nonscalar melodies. Indeed, Dowling (1988) found that children as young as 3 years old can be more likely to judge that a melody sounds "funny" if it is nonscalar rather than scalar.

Although the foregoing evidence supports a sense of scalar structure emerging early in development, the origins of this sense are not yet understood. One approach to understanding scalar structure is based on the argument (e.g., Dowling, 1978) that children's exposure to diatonic melodies allows them to abstract some sort of scale schema, which thereafter is employed in the encoding of new melodies. Presumably, a melody's fittingness with an internal scale schema—the ease with which it is assimilated to this schema—affects how "good" (or non-funny) it sounds. A second approach to understanding scalar structure assumes that there are purely physical aspects of scales that render scalar melodies "better" or less "funny." Specifically, diatonic scales include a reasonably large number of pitch classes (seven) while minimizing the occurrence of semitone intervals—the major diatonic scale (do–re–mi–fa–so–la–ti–do) follows an interval structure (2–2–1–2–2–2–1) containing only two one-semitone intervals. No other scale using seven pitch classes out of 12 equal-tempered semitones has fewer one-semitone intervals, and melodies that violate diatonic scales will generally have more one-semitone intervals than those that match such scales. Since tones that are less than two semitones apart appear to produce interference in memory (Deutsch, 1982), scalar melodies might sound "better" *not* because they are readily assimilated to schemata, but rather because they contain small numbers of one-semitone intervals.

A third approach to understanding scalar structure is to assume that such structure contributes to the "goodness" of melodies as gestalts. Due to their organization into phrases and larger structural units, as well as their transposability in pitch, melodies have long been viewed as prototypical gestalts (e.g., Boring, 1942; Dowling & Harwood, 1986). Moreover, research on

1. Cross et al. (1983) initially found that their least scalar melodies evoked high liking ratings. However, this effect was due to the perceived rhythm of the melodies, which tended to place stress on a set of scale-compatible notes in the lowest-scalar-structure condition. When Cross et al. removed this confound by altering the rhythm of their stimuli, the least scalar melodies received low liking ratings, as would be expected.

the factors affecting gestalt "goodness" has clear implications for the perception of melodies. Following Garner's (1970, 1974) conception that "good patterns have few alternatives," we recently suggested that a melody, like other simple visual and auditory patterns, is perceived as belonging to a set of alternative stimuli (Dowling & Bartlett, 1981; Bartlett, 1984). The smaller the set of alternative stimuli, the better (more pleasant or stable) the melody. The "wrong notes" heard in a nonscalar melody reflect a subject's perception that, because of these notes, the set of perceived alternatives is large.

How might the scalar structure of a melody affect its perceived alternatives? Our notion is twofold: First, the perception of a melody evokes a set of alternatives that share various of its features (e.g., contour, general pitch range, loudness, timbre). This is consistent with the "norm theory" of Kahneman and Miller (1986), who claim that: (a) "the experienced facts of reality [i.e., stimuli] evoke counterfactual alternatives and are compared to these alternatives," and (b) some features of a stimulus are treated as "immutable" in that they "guide and constrain the spontaneous recruitment of alternatives to it" (p. 142).

Second, if a melody is perceived as scalar, the set of alternatives evoked by this melody is constrained to fit its musical key. In terms of norm theory (Kahneman & Miller, 1986), the scale from which a melody's notes are chosen functions as an immutable feature of that melody, constraining the set of alternative melodies to which it is (often implicitly) compared. Note that in the system of European tonal music, the most typical scale pattern is the diatonic scale of seven pitch classes (i.e., seven pitches and their functional equivalents at octave intervals), selected from the 12 pitch classes of the chromatic set (i.e., the European "tonal material," Dowling, 1978). Thus, we are suggesting that if a melody is perceived as being scalar, its perceived alternatives are restricted to melodies constructable from the seven pitch classes of a diatonic scale. In contrast, if a melody is perceived as nonscalar, its perceived alternatives are only weakly or not at all constrained by a diatonic scale—they might include melodies based on all 12 pitch-classes of the chromatic set. Since the 12 pitch classes of the chromatic set include the 7 pitch classes of any major diatonic scale, a nonscalar melody would evoke a larger set of alternative melodies than a scalar melody that is otherwise similar (same contour, length, etc.). Indeed, the set of alternatives of a scalar melody may be "nested" within that of a nonscalar melody that is otherwise similar to it (Garner, 1974).

The idea that scales constrain alternative sets for melodies can be visualized using spatial models of pitch, particularly the "melodic map" of Shepard (1982). One axis of Shepard's spatial model is the cycle of fifths of musical keys, each key generating a diatonic scale. Based on the musical interval of the fifth, the cycle of fifths moves through all possible pitch classes of the chromatic set, making no repeats (if one starts with F, the cycle continues with C, G, D, A, E, B, F♯, C♯, G♯, D♯, A♯ [or B♭], and then returns to F). More important, any set of seven notes from contiguous locations along the cycle constitute the notes of a diatonic major scale. Thus, if we begin with F and follow the cycle to B (i.e., F, C, G, D, A, E, B), we have the seven notes of the C major scale. When the cycle-of-fifths is "cut open" at some (arbitrary) point to form a one-dimensional axis, and combined with a second axis representing pitch height (frequency), we have a two-dimensional "map" for representing melodies, as shown in Figure 1 (note that A♯ is actually proximal to F around the cycle of fifths).

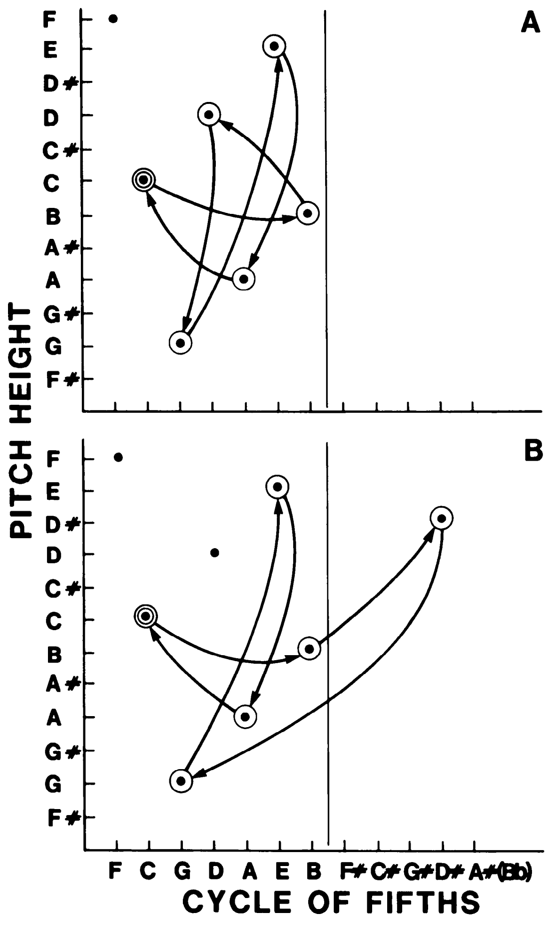

Fig. 1. A portion of Shepard's (1982) melodic map, unrolled onto a two-dimensional surface with the dimensions of pitch height (ordinate) and key distance around the cycle of fifths (abscissa). Filled circles represent the seven pitches of the C-major scale, and the vertical line separates scalar from nonscalar pitches. Open circles and arrows indicate (a) the scalar melody C–B–D–G–E–A–C, and (b) the nonscalar melody C–B–D♯–G–E–A–C.

The melody shown in Figure 1A begins and ends on C, and contains neither flats nor sharps, and so it clearly matches the key of C. This fact is highlighted on the melodic map by means of a vertical line delimiting the notes of the C-major scale. All notes of the melody fall to the left of the line, and this captures the notion that scalar structure limits the set of alternatives perceived for a melody. If the melody were transposed into a different key (e.g., G), its map-position would be changed, but its "shape" would be the same and again it would fall into a compact map region. Thus, strongly scalar melodies in any diatonic key occupy restricted regions of the melodic map. The hypothesis we are forwarding here is that the perceived alternatives for strongly scalar melodies occupy the same restricted map regions as the melodies themselves. In contrast, as shown in Figure 1b, weakly scalar and nonscalar melodies will tend to cover the full lateral extent of the map, and we suggest that their perceived alternative melodies will also cover the entire map range. Thus, we have a way of visualizing the notion that a set of alternatives for a scalar melody is smaller than that of a nonscalar melody and is "nested" within that of a nonscalar melody.

The idea that scalar melodies evoke small sets of perceived alternatives, and that the alternative sets evoked by scalar melodies can be nested within those evoked by nonscalar melodies, has been supported by a finding of asymmetric similarity (Tversky, 1977). Dowling and Bartlett (1981) constructed pairs of five-note melodies, each pair consisting of a scalar melody accompanied by a relatively nonscalar mate. The scalar melody was a random permutation of the first five notes of the C-major scale, always beginning on middle C (262 Hz). Its nonscalar mate was physically identical, except that one note was changed by one semitone to fall outside of the C-major scale (e.g., a D might be replaced by D♯). Subjects' task was either to rate each pair for similarity, or to rate their confidence that each pair was "same" versus "different" (although in fact all pairs were different). Both similarity and same–different ratings were made using a six-point scale, and average ratings were higher (i.e., judged similarity/sameness was greater) when the nonscalar item came first (nonscalar–scalar pairs) than when the scalar item came first (scalar–nonscalar pairs). Asymmetric similarity has been found outside the music domain (Handel & Garner, 1966; Tversky, 1977), and even within the music domain the effect is not restricted to melodies—asymmetric similarity has been found with single pitches presented in a tonal context (Krumhansl, 1979), as well as with chord sequences (Bharucha & Krumhansl, 1983). However, the effect obtained with melodies is most directly relevant here.

In interpreting asymmetrical similarity of melodies, we follow Garner (1974; Handel & Garner, 1966) in assuming that asymmetries reflect a "nesting" relation in the sets of alternatives evoked by two stimuli. That is, asymmetric relatedness between two stimuli occurs when the set of alternatives evoked by one stimulus includes the second stimulus, but not vice versa. Turning to the paradigm of Dowling and Bartlett (1981), the idea is that if a first melody evokes a set of alternatives, and if a subsequent melody falls outside this set of alternatives, an impression of dissimilarity results. This state of affairs should be much more probable with scalar–nonscalar pairs—the initial melodies of such pairs have sets of alternatives constrained by their key, and the second melodies in such pairs should frequently fall outside of these sets. In contrast, violations of alternative sets should be rare in the case of nonscalar–scalar pairs. The initial melodies of such pairs are likely to have large sets of alternatives, including not only other nonscalar melodies, but also scalar melodies. And since the second, scalar, melodies share many features of their first-presented mates, they are likely to fall within the sets of alternatives evoked by their mates.

Toward the general goal of evaluating a perceived-alternatives hypothesis for effects of scalar structure of melodies, the present research was a further exploration of asymmetric similarity of scalar and nonscalar melodies. Our aim was to determine whether asymmetric similarity was better explained by a perceived-alternatives hypothesis than by three competing views. First, we addressed the possibility that asymmetric similarity results from scalar melodies being generally more memorable than nonscalar melodies (Experiments 1 and 2). Second, we examined the notion that asymmetric similarity is the product of a task-specific strategy as opposed to the intrinsically perceptual process of inferring alternative sets for melodies (Experiments 2 and 4). Third, we tested the hypothesis that the greater number of one-semitone intervals in nonscalar melodies than in scalar melodies is sufficient to explain asymmetric similarity (Experiments 3 and 4). Since the four experiments that tested these views were fairly similar in design and procedure, it is useful to present a General Method section prior to details of the individual experiments.

General Method

SUBJECTS

Undergraduates at the University of Texas at Dallas served in these experiments for partial course credit. This population has a mean age of 29.4 years, and consists of 64% females. Based on prior findings that moderate amounts of prior musical experience affect certain aspects of melody processing (Dowling & Harwood, 1986), subjects in each experiment were separated into two groups, "moderately experienced" and "inexperienced." Moderately experienced subjects had 2 years or more of musical training, including lessons on an instrument or voice, and playing in instrumental ensembles (but not choral singing). Those subjects had an average of about 5 years training approximately 10 to 15 years prior to the experiment. Inexperienced subjects had less than 2 years musical training.

STIMULI

The stimuli consisted of pairs of seven-note melodies. On each trial, subjects would hear a pair of melodies in which one of the notes was altered in pitch by one semitone. Melodies were either scalar or nonscalar, and were combined in four types of pairs as shown in Table 1. In Experiments 1 and 2, all four possible types of pairs were used: pairs in which both melodies were scalar (SS), pairs in which both melodies were weakly scalar or nonscalar (NN), pairs in which the first melody was scalar and the second was not (SN), and pairs in which the second melody was scalar and the first was not (NS). In Experiments 3 and 4 only SN and NS pairs were employed. Note that for both SN and NS pairs, the one-semitone change was such as to alter the scale structure of the melody (A♯ might be substituted for A). For SS and NN pairs however, the one-semitone change did not affect scalar structure (E might be substituted for F).

The stimuli were based on 24 or 48 seed melodies, all beginning and ending with middle C (fundamental frequency, 262 Hz). The intervening five pitches were all different and consisted of five pitches of the C-major diatonic scale other than C. Either E or F was included in this set of pitches, but not both. Since there are only seven pitches in the scale, this meant that the five pitches in the middle of the melody were drawn from the set A–B–D–G plus either E or F. These five possible intervening pitches gave $5! = 120$ possible melodies. For each experiment we randomly selected the seed melodies from the set of 120 possible melodies. A constraint on selection was that the melodic contour (pattern of ups and downs in pitch) was different for each seed melody.

An SN pair was constructed from a seed melody by recording the seed melody twice, first with no change (scalar) and then with a nondiatonic note substituted for a diatonic note (nonscalar). The diatonic and nondiatonic notes always differed by one semitone. An NS pair was constructed in the same way as an SN pair, except it was the second melody that was recorded unchanged and the first that was recorded with the nondiatonic note. An SS pair was constructed by recording the seed melody twice, once with no change and once with an E substituting for the F or vice versa (depending on whether the seed melody contained an E or F). An NN pair was constructed in precisely this same fashion, except that a diatonic pitch (besides the E or F) was replaced by a nondiatonic pitch in both members of the pair (for example, A♯ for A). Note that the two melodies comprising each pair always differed by only one semitone in a single serial position. This one-semitone change did not cause a change in the melodic contour of the melodies.

Each experiment consisted of from two to five blocks of 24 melody pairs. Each of the pairs within a block were constructed from a unique seed melody. There was random assign-

TABLE 1

Examples of Melody Pairs Used in Experiment 1

Item Type	First Melody	Second Melody
SS	C B D G E A C	C B D G F A C
NN	C B D♯ G E A C	C B D♯ G F A C
SN	C B D G E A C	C B D♯ G E A C
NS	C B D♯ G E A C	C B D G E A C

NOTE: S = scalar, N = nonscalar.

ment of seed melodies to condition (for example, SS, NN, SN, NS), with the constraint that each condition occurred equally often in each block. The melody pairs were presented over loudspeakers using high quality stereophonic equipment.

PROCEDURE

Subjects served in group sessions and were introduced to the experiment by means of instructions that included four examples of the types of melody pairs they would hear. Subjects were asked to listen to each pair of melodies carefully and to rate (for example) the similarity of the melodies in the pair on a six-point scale going from "6" ("very similar") to "1" ("very different"). Subjects responded on an answer sheet printed with blanks numbered with the appropriate number of trials. Subjects also completed a brief questionnaire concerning age, sex, and musical experience.

Experiment 1

The aim pursued in Experiment 1 was to determine if the finding of asymmetric similarity could be explained by a simple *memorability* hypothesis. It frequently is claimed that more strongly tonal melodies are more easily remembered than less tonal or atonal melodies, although relevant data are scarce (but see, e.g., Francès, 1958/1988). Making the working assumption that this claim is true, and that it extends from tonal structure in general to *scalar* structure in particular, the asymmetric similarity effect can be easily explained. First, consider scalar–nonscalar pairs: Since the notes of the first (scalar) melody are very well remembered, the alteration in a note of the second (nonscalar) melody is easily detected, producing an impression of low similarity. Now, consider nonscalar–scalar pairs: Since the notes of the first (nonscalar) melody are remembered relatively poorly, the alteration in a note of the second (scalar) melody is *not* easily detected, producing an impression of relatively high similarity.

The memorability hypothesis is compatible with either a schema-assimilation view or the memory-interference view of effects of scalar structure on processing of melodies (see Introduction). That is, the low memorability of nonscalar melodies might be attributed either to their less than perfect match to existing scale schemata, or to the semitone intervals—and memory interference—such melodies produce. Hence, in testing the memorability hypothesis, we are providing information relevant to these broader hypotheses for scalar structure of melodies.

In order to test the memorability hypothesis, Experiment 1 included the SN and NS conditions used in our prior studies, along with two additional conditions; SS and NN (see Table 1). Although all melody pairs were changed by only one semitone in one serial position, the memorability hypotheses predicts that this semitone mismatch will be more easily detected when the first melody is scalar than when it is nonscalar. Hence, the rated similarity of SS pairs should be as low as that of SN pairs and lower than that of NS and NN pairs.

The perceived-alternatives hypothesis does not make the same predictions. Since the melodies of an SS pair are in the same key, and also are identical save for one note changed by a semitone, the second melody in such a pair need not mismatch the perceived alternative set of the first. If there is no such mismatch, similarity ratings made to SS pairs should be higher than those made to SN pairs and at least as high as those made to NS pairs, in direct contradiction to the memorability hypothesis. On the assumption that two nonscalar melodies do not produce alternative-set violations, the rated similarity of NN pairs also should be high.

As a test of generality of whatever findings emerged, we followed Dowling and Bartlett (1981) in asking half of the subjects to rate the similarity of each melody pair and the remainder to rate their confidence that the melodies of each a pair were exactly the same versus changed. Subjects were not informed that all melodies were changed in one note.

METHOD

Subjects

Twenty-eight subjects served in Experiment 1, with 13 in the inexperienced group and 15 in the moderately experienced group.

Stimuli

Stimuli were produced on a freshly tuned Steinway piano. Timing was controlled by the clicks of a metronome set at 72 beats/min, which were barely audible on the tape. Melodies proceeded at a rate of three notes per metronome beat, or 3.6 notes/sec. Melodies were played "legato," with little time between note offsets and onsets, and the seventh note of each melody was held for a full metronome beat. (That is, the last note was three times as long as each of the first six notes.) One metronome beat separated the standard and comparison melodies on a trial. Each trial began with a warning consisting of the spoken trial number two beats (1.67 sec) before the onset of the first melody. Subjects had eight beats (6.67 sec) to respond before the start of the next trial.

Melodies were constructed from the set of pitches between the F♯ below middle C and the F above. Standard and comparison melodies on a trial always differed in exactly one pitch, which was altered by one semitone. There were four types of trials: SS, NN, SN, and NS. The melodies on SS and NN trials differed by the interchange of the pitches E and F. The melodies on SN and NS trials differed by the substitution of nondiatonic pitches for diatonic ones: A♯ for A, C♯ for D, or F♯ for G (randomly determined and occurring equally often). These substitutions of nondiatonic pitches had the effect of making the N melodies either strictly nonscalar (in that the melody could not be interpreted in any diatonic scale), or relatively nonscalar (in that the melody could not be interpreted within the C major scale implied by all melodies of the experiment, all of which began and ended on C). The substitutions involving A♯ and C♯ produced strict nonscalarity through the creation of adjacent one-semitone intervals (A♯–B–C and B–C–C♯). The substitutions involving F♯ produced strict nonscalarity in those melodies containing F due to the creation of two tritone (six-semitone relationships within the set of pitches (i.e., B–F and C–F♯). The substitutions involving F♯ produced relative nonscalarity in those melodies containing E instead of F. This is because F♯ in conjunction with the pitches C, D, E, G, A, and B is compatible with the G-major diatonic scale, and hence not strictly nonscalar. However, F♯ in that context is perceptually jarring due to its interval of a tritone with the initial and final Cs of the melody, thereby producing relative nonscalarity. Thus, strict nonscalarity occurred in five-sixths of the N melodies, whereas relative nonscalarity occurred in one-sixth.

Procedure

There were 96 trials arranged in four blocks of 24 each. Each block consisted of a random arrangement of six occurrences each of the four trial types (SS, NN, SN, NS). Subjects served in group sessions, and different sessions performed the blocks in different order, counterbalanced so that approximately equal numbers of subjects performed each block in each ordinal position. The same seed melodies were repeated in each block, but a given seed melody was tested in each of the four different types of trial in different blocks. Seven of the subjects at each experience level rated the stimuli for similarity (on the six-point scale described above), and the remainder judged the degree to which the comparison stimulus on each trial appeared to be the same or different (with six response categories: very sure different, sure different, different, same, sure same, very sure same, going from "1" to "6").

RESULTS

Each subject's responses were used to compute the average similarity or same–different rating for SS, SN, NS, and NN items (S = scalar, N = nonscalar). These individual-subject data were subjected to an analysis of variance (ANOVA) which included the between-subjects variables of Rating type (similarity/same–different) and Musical experience (moderate/none), and the within-subjects variables of First melody type and Second melody type. The alpha level was .05.

The ANOVA showed main effects for both First melody and Second melody, $F(1,24) = 40.2$ and 39.1, $MSe = .069$ and $.215$, respectively. It also showed interactions between First melody and Musical experience, $F(1,24) = 4.96$, $MSe = .069$, Second melody and Musical experience, $F(1,24) = 5.98$, $MSe = .215$, and between First and Second melody, $F(1,24) = 14.6$, $MSe = .147$. The rating-type variable produced no main effect and participated in no reliable interactions. As shown in Table 2, in which the data have been collapsed over the rating-type variable, the SS and NS pairs produced the highest ratings, and the SN pairs produced the lowest ratings, with the NN pairs falling in between.

A set of planned F tests confirmed that there was no reliable difference between the ratings to SS and NS pairs ($F < 1$), whereas the ratings to both reliably exceeded the ratings to SN pairs, ($F(1, 24) = 36.6$ and 54.3 for the SS–SN and NS–SN comparisons, respectively, with $MSe = .260$ and $.192$). The ratings to the NN pairs differed reliably from ratings to each of the other three pairs ($F(1, 24) = 8.21$, 10.1, and 29.3, and $MSe = .092$, $.101$, and $.167$, for the comparisons with SS, NS, and SN pairs, respectively).

TABLE 2
Average Ratings of Melody Pairs by Subjects in Experiment 1

	Type of Melody Pair			
Musical Experience	Scalar–Scalar	Scalar–Nonscalar	Nonscalar–Scalar	Nonscalar–Nonscalar
Inexperienced	4.41	3.85	4.39	4.28
Moderately Experienced	4.65	3.56	4.76	4.32
Mean	4.53	3.71	4.58	4.30

NOTE: Data from the Similarity Rating Group and Same–Different Rating Group have been collapsed in this table (there being no significant effect of rating task in the ANOVA). Both similarity and same–different ratings were based on a 1-to-6 scale (6 = most similar or most confidence in identity of melodies).

Out of the six planned *F* tests, only one produced a reliable effect involving the musical experience variable. This was the *F* test comparing the NS and SN conditions, in which the Melody pair type × Musical experience interaction gave $F(1, 24) = 7.70$, $MSe = .192$, replicating the findings of Dowling and Bartlett (1981). As shown in Table 2, the size of the difference between the NS and SN conditions was greater among the more experienced subjects.

DISCUSSION

Contrary to the memorability hypothesis, the similarity/same–different ratings for SS pairs were significantly higher than those for SN pairs, and were *as* high as the ratings for NS pairs. This pattern fits well with the perceived-alternatives hypothesis: SS pairs, like NS pairs, *should not* produce mismatches with perceived alternative sets. Hence, both sorts of pair *should* be rated as more similar than SN pairs, which should produce mismatches with perceived alternative sets.

A possible difficulty for the perceived alternatives view is that NN pairs received lower ratings than SS and NS pairs. This effect was not predicted, though the SS-versus-NN difference fits a number of prior studies by Krumhansl and her colleagues (e.g., Krumhansl, 1979; Bharucha & Krumhansl, 1983). Using single notes as well as chords and chord sequences, these studies have shown greater similarity among stimuli central to a prevailing key than among those distant from a prevailing key.

Experiment 2

The aim pursued in Experiment 2 was to determine if the asymmetric similarity effect is due to *automatic processes* versus *task-specific strategies*. Faced with the data from Experiment 1, it is possible to argue that the low similarity and same–different ratings made to SN pairs reflects a specific strategy that subjects adopt in order to cope with the demand characteristics of certain tasks. For example, charged with rating pairs of melodies for similarity or sameness/difference and finding all pairs approximately equally similar or confusable (all have just a one-semitone change in one note), subjects might: (a) attempt to extract a sense of key from the first melody, (b) judge whether this key is violated by the notes of the second melody, and (c) take key violations as evidence for dissimilarity or difference, *all as a task-specific strategy*. Since key violations would be more frequently detected in SN pairs than in the other types of pairs, the low ratings of the former would be handled. To evaluate this strategy argument, we decided to exert more control over subjects' processing strategies through use of a well-defined task: Subjects were asked to judge pairs of melodies as same versus different in melodic contour (the pattern of ups and downs in pitch), under conditions in which accuracy could be meaningfully assessed.

Contour of melodies is a highly salient feature, as evidenced by the fact that even inexperienced subjects can make same–different judgments with respect to this feature with high levels of accuracy (Dowling & Fujitani, 1971; Dowling & Harwood, 1986). Furthermore, sameness or difference in the contour of two melodies is physically independent of their tonalities or keys. Thus, there is simply no reason why listening for changes in key should be used as a strategy for making contour judgments. However, if the alternatives evoked by a first-presented melody are necessarily constrained by key, and if a second-presented melody that violates these alternatives necessarily creates an impression of difference, relations of scale structure might affect contour judgments. Specifically, SN pairs, in which violations of the alternative set are most likely to occur, should be judged more often as different in contour than NS and SS pairs, in which set violations are unlikely to occur. This pattern should hold even in cases in which contour actually is unchanged.

METHOD

Subjects

Fifty-two subjects served in Experiment 2, with 35 in the inexperienced and 17 in the moderately experienced groups.

Stimuli

Stimuli were produced using the "clarinet" voice of a Yamaha Portasound synthesizer, which was connected directly to a tape recorder. Timing was controlled by the built-in metronome of the synthesizer set to blink at 60 beats/min, which produced no audible signal on the tape. Melodies proceeded at a rate of two notes per metronome beat, or 2.0 notes/sec. Melodies were played "legato," with little time between note offsets and onsets, and the seventh note of each melody was held for a full metronome beat. (That is, the last note was twice as long as each of the first six notes.) Two metronome eats (2 sec) separated the standard and comparison melodies on a trial. Subjects had 5.0 sec to respond before the start of the next trial. The seed melodies were constructed using the pitches ranging from the G below middle C to the G above. A given melody contained either the high G or the low G, not both.

As in Experiment 1, there were four tonality conditions: SS, NN, SN, and NS. The melodies of SS and NN pairs differed by the interchange of the pitches E and F. The melodies of SN and NS pairs differed by the substitution of nondiatonic pitches for diatonic ones: A♯ for A, or D♯ for D (randomly determined and occurring equally often). While all melody pairs contained a one-semitone change of one note, half of these pairs also contained a contour change. Changes in contour were accomplished by substituting the G above middle C for the G below, or vice versa (randomly determined and occurring equally often). Same contour trials used either the high or low G, randomly determined.

Procedure

Subjects received 48 melody pairs arranged in two blocks of 24. Each block consisted of a random arrangement of three occurrences each of the eight pair types (SS, NN, SN, and NS, all with same or different contour). Items were also counterbalanced across two lists with respect to scalar structure of comparison melody, so that a seed tested as a SS trial in one list was tested as a SN trial in the other list; and similarly for NN and NS trials. Subjects participated in group sessions, and their task was to rate each melody pair for similarity, and also for sameness versus difference in contour, using a six-point scale in both cases (6 = most similar and sure-same-contour, 1 = least similar and sure-different-contour). The contour judgments are of most interest, but we wanted to test the replicability of the pattern of similarity ratings found in Experiment 1.

RESULTS

The analysis of similarity ratings was similar to that of Experiment 1. The major change was that same-versus-different Contour of pairs was an additional factor in the ANOVA. Further, due to the slight unevenness in cell sizes, the counterbalancing variable of List was also included as a factor.

The outcome of the ANOVA resembled that of Experiment 1 in that there were reliable main effects of First melody and Second melody, $F(1, 48) = 5.96$ and 37.6, $MSe = .271$ and $.302$, respectively, as well as a First × Second melody interaction, $F(1, 48) = 5.95$, $MSe = .284$. It can be seen in Table 3 that similarity ratings were once again highest in the SS and NS conditions, lowest in the SN condition, and intermediate in the NN condition. It also can be seen in the table that Contour had a large main effect, F

TABLE 3
Average Similarity Ratings of Melody Pairs by Subjects in Experiment 2

	Contour and Scalar Structure Condition							
	Same Contour				Different Contour			
Musical Experience	SS	SN	NS	NN	SS	SN	NS	NN
Inexperienced	4.54	4.21	4.52	4.35	3.19	2.78	3.26	3.01
Moderately Experienced	4.59	4.11	4.64	4.27	3.35	2.62	3.24	3.15
Mean	4.57	4.16	4.58	4.31	3.27	2.70	3.25	3.08

NOTE: SS stands for scalar–scalar pairs, SN for scalar–nonscalar pairs, NS for nonscalar–scalar pairs, and NN for nonscalar–nonscalar pairs.

(1, 48) = 197.0, *MS*e = .816: Same-contour items received substantially higher similarity ratings than different-contour items. Although the counterbalancing variable of List produced a main effect, F (1, 48) = 4.12, *MS*e = 1.93, it did not enter into reliable interactions.

A difference from Experiment 1 was that there were no statistically reliable effects involving the Musical experience variable. Table 3 shows only a very small tendency for the difference between NS and SN pairs to be greater in the moderately experienced group than in the inexperienced group.

A second ANOVA was performed on the average contour ratings, which also were made on a six-point scale (6 = sure same, 1 = sure different). The outcome was somewhat different than that performed on the similarity ratings: Whereas we once again obtained a main effect of Second melody, F (1, 48) = 50.9, *MS*e = .289, there was no main effect for First melody and no First × Second melody interaction. There was of course a Contour main effect, F (1, 48) = 173.2, *MS*e = .176, which in this ANOVA was qualified by a Contour × Second melody interaction, F (1, 48) = 5.90, *MS*e = .380. The counterbalancing variable of List produced no main effect, and although it interacted with First melody, F (1, 48) = 4.52, *MS*e = .270, it did not qualify any of the aforementioned statistically reliable effects. Again, Musical experience entered into no reliable effects.

Although average contour ratings have the advantage of being comparable to average similarity ratings, it is useful also to examine proportions of "same-contour" judgments (ratings of 4, 5, or 6) to same-contour items (hits) and different-contour items (false alarms). An ANOVA was performed on the proportions of same-contour decisions, and the results were highly similar to that performed on contour ratings. The main effect of Second melody was reliable, as also was that of Contour, $F(1, 48) = 38.3$ and 174.0, $MSe = .023$ and .080, respectively). A minor difference from the ANOVA of contour ratings was that the Second melody × Contour interaction was only marginally reliable, $F(1, 48) = 3.37$, $MSe = .031$. There were no effects involving either List or Musical experience.

Despite some differences in the outcomes of ANOVAs, the actual pattern of means of the similarity ratings, contour ratings, and proportions of same-contour judgments were actually quite similar. As shown in Table 4, contour ratings and same-contour judgments resembled similarity ratings in showing highest scores in the SS and NS conditions, and lowest scores in the SN condition. The principal difference was that with the contour ratings and same-contour decisions, the four pair-conditions were closer together with different-contour items than with same-contour items (this was reflected by the Second melody × Contour interactions in the contour-judgments ANOVAs, but not in the similarity-judgment ANOVA). We note that an ANOVA of the contour-ratings *to different-contour items only* showed a main effect of Second melody tonality, $F(1, 48) = 11.3$, $MSe = .241$. A similar ANOVA of same-contour-judgment proportions also showed this effect.

DISCUSSION

A strong prediction of the perceived-alternatives hypothesis was confirmed in this experiment. We replicated the finding that SS and NS pairs have higher similarity than SN pairs, and found in addition, as predicted, that average same-versus-different contour ratings as well as proportions of same-contour judgments showed generally this same pattern. Given the

TABLE 4

Average Contour Ratings and Proportions of Same-Contour Decisions in Experiment 2

	Contour and Scalar Structure Condition							
	Same Contour				Different Contour			
Type of Rating	SS	SN	NS	NN	SS	SN	NS	NN
Contour Ratings	4.10	3.53	4.24	3.70	2.58	2.27	2.46	2.29
Same-Contour Decisions	.68	.53	.71	.59	.28	.20	.26	.21

NOTE: SS stands for scalar–scalar pairs, SN for scalar–nonscalar pairs, NS for nonscalar–scalar pairs, and NN for nonscalar–nonscalar pairs.

very high salience of the contour attribute in immediate memory for melodies, we find it impressive that contour judgments were sensitive to our manipulations of scalar structure. The fact that contour judgments were sensitive in this way suggests quite strongly that SN pairs, as opposed to SS and NS pairs, produce strong impressions of dissimilarity, apart from subjects' conscious strategies.

Just as in Experiment 1, NN pairs in this experiment apparently were perceived as reliably less similar than SS and NS pairs. Since it is not obvious that NN pairs would produce violations of perceived sets of alternatives, this outcome is somewhat puzzling—we take up this matter in the General Discussion.

Experiment 3

In Experiment 3 we question whether the physical property of semitone intervals explains the asymmetric similarity effect. We have mentioned that one-semitone intervals are more prevalent in nonscalar melodies than they are in scalar melodies. One-semitone intervals might have a variety of effects, including that of reducing memorability of melodies. Experiment 1 was a test of the hypothesis that low memorability of nonscalar melodies, possibly resulting from their one-semitone intervals, explains the asymmetric similarity phenomenon. Although we failed to find evidence in support of this hypothesis, other effects of one-semitone intervals might yet account for asymmetric similarity. For example, assume that (a) numerous one-semitone intervals make a melody sound abnormal, and (b) a relatively abnormal-sounding stimulus tends to be perceived as dissimilar not only to other stimuli in general, but to the stimulus that preceded it in particular. This would explain why nonscalar melodies, in which one-semitone intervals are relatively frequent, are perceived as dissimilar to scalar melodies when the scalar–nonscalar order is used.

There are other possible accounts of asymmetric similarity that point to the prevalence of one-semitone intervals in nonscalar melodies. In order to evaluate this entire class of accounts, we examined whether asymmetric similarity would generalize to conditions in which *neither* the scalar *nor* nonscalar melodies contained *any* one-semitone intervals. The perceived-alternatives hypothesis—which makes no specific reference to one-semitone intervals—predicts that asymmetric similarity should generalize to melodies lacking one-semitone intervals. In contrast, if one-semitone intervals are critical to the finding of asymmetric similarity, such generalization should not occur.

The ideal materials to use in this study would be melody pairs in each of which: (a) one melody was scalar whereas the other was strictly nonscalar, that is, inconsistent with *all* diatonic major scales (regardless of key), (b) neither melody contained any semitone intervals, and (c) the two melodies differed in just one note altered by one semitone. Unfortunately, we were unable to develop such pairs, and so we adopted the strategy of relaxing constraint "a," while maintaining constraints "b" and "c."[2] Specifically, the "nonscalar" melodies used in this experiment all were merely *relatively* nonscalar as opposed to *strictly* nonscalar (see Method section of Experiment 1). That is, all "nonscalar" melodies were compatible with a diatonic major scale (G major), but they were incompatible with the particular scale (C major) suggested by the starting and ending notes (C), as well as all other melody pairs in the study. Although this solution is less than ideal, note that to the extent that "nonscalar" melodies might be perceivable as scalar, the study was biased *against* the finding of asymmetric similarity. Hence, if Experiment 3 supports asymmetric similarity, the perceived alternatives hypothesis will be strongly favored.

METHOD

Subjects and Procedure

Fifty-six subjects served in Experiment 3, with 26 in the inexperienced group and 30 in the moderately experienced group. They served in group sessions, and performed the task of listening to each of 120 melody pairs, and rating the similarity of each pair on a six-point scale.

Stimuli

Stimuli were produced by a Commodore 64 computer via its 6581 Sound Interface Device and recorded directly on tape. Timing was controlled by the computer with an accuracy of .017 sec. Melodies proceeded at a rate of 2.6 notes/sec. Each note was 0.32 sec in duration, with an onset time of about 24 msec followed by a decay to a level 0.27 of peak amplitude (decay constant = 48 msec). Internote intervals were 0.06 sec in length. A 2-sec interval separated the standard and comparison melodies on a trial, and subjects had 8 sec to respond before the start of the next trial.

Each of the 24 seed melodies used the six pitches of the key of C ranging from the E below middle C to the E above, excluding B and F (i.e, E [low], G, A, C, D, E [high]). Note that this pitch set includes no semitone intervals. The seven-note seed melodies always began and ended on C, which never occurred within the five "internal" notes.

2. Producing strictly nonscalar melodies by shifting just one pitch (constraint "c") while satisfying the constraint of no semitone intervals ("b"), with melodies containing six different pitches, is virtually impossible. The two principal ways to make a melody strictly nonscalar are: (1) to include two adjacent one-semitone intervals (violating constraint "b"), and (2) including *two* pairs of tones in the tritone (6-semitone interval) relationship. Since we had eliminated the note B from all melodies (required by constraint "b"), which is the only note standing in a tritone relation (with F) within C major, our scalar melodies had no tritone intervals. Thus, altering just one pitch (constraint "c") could produce a pattern containing at most *one* tritone. These considerations demanded that we settle for *relative* rather than *strict* nonscalarity.

Unlike Experiments 1 and 2, here we used only two types of melody pairs, SN and NS. The two melodies comprising an SN or NS pair always differed in one note—the G in the scalar melody was replaced by an F♯ in its "nonscalar" mate. Since none of the melodies contained an F, the use of F♯ in the "nonscalar" melodies could *not* produce any semitone intervals. Thus, there were no semitone intervals in either scalar or "nonscalar" melodies.

Because we were using only two types of melody pair, we were able to employ more extensive counterbalancing than in Experiments 1 and 2. Each of the 24 seed melodies was used to construct five SN pairs and five NS pairs. Within each set of five pairs, the serial positions of notes 2 through 6 were shifted so that the altered note (G to F♯ or vice versa) occurred once at serial position 2, once at serial position 3, once at serial position 4, once at serial position 5, and once at serial position 6. Each of the five pairs from each seed melody was assigned to one of five blocks of the list. Hence, the list included 120 melody pairs, organized into five successive blocks, each block including one melody pair made from each of the 24 original seed melodies. Across the five blocks, each seed melody was represented by a pair with the changed note occupying each of the five internal serial positions. Further, a seed was tested either as an SN pair in three blocks and an NS pair in two, or as an SN pair in two blocks and an NS pair in three.

There were two different versions of the 120-pair list, each version used with half (28) of the subjects. The lists were identical save for the ordering of pairs: Each NS pair on version 1 of the list was played as an SN pair on version 2, and vice versa.

RESULTS

An initial look at the data showed that the average similarity rating for SN pairs was 4.09, while that for NS pairs was 4.34. The difference appeared small, and we wondered if it might be clarified by considering the position in the melodies of the altered notes (the changed-note position varied from 2 through 6—positions 1 and 7 always contained Cs). Hence, we performed an ANOVA including the within-subjects variable of Changed note position along with the within-subjects variable of Melody pair type (SN vs. NS), and the between-subjects variables of List version and Musical experience. Neither the List nor Musical experience variables produced reliable main effects or entered into reliable interactions. However, the ANOVA showed a main effect for Melody pair type, F (1, 52) = 15.2, MSe = .517, a main effect for Changed note position, F (4,208) = 3.22, MS = .172, and a Melody pair type × Serial position interaction, F (4,208) = 3.92, MSe = .143. As shown in Table 5, the asymmetrical similarity effect was statistically reliable for changed notes at serial positions 3 through 6, though not at serial position 2.

Since the asymmetric similarity effect was relatively small, we were concerned with its generalizability over different melody pairs. Consequently, we performed a second ANOVA on the effects of Changed note position and Melody pair type using items rather than subjects as the random variable. Melody pair type was a within-items variable, whereas Changed note position was a between-items variable; there were 24 melody pairs in each Changed note position condition. The main effect of Changed note position was no longer reliable, but the main effect of Melody pair type and the Mel-

TABLE 5
Average Similarity Ratings for Melody Pairs Changed in Serial Positions 1 through 6 in the Scalar–Nonscalar and Nonscalar–Scalar Conditions

	Scalar Structure Condition		
Serial Position	Scalar–Nonscalar	Nonscalar–Scalar	Diff.
2	4.25	4.27	−.02
3	4.02	4.23	−.21[a]
4	4.16	4.44	−.28[a]
5	4.02	4.33	−.31[a]
6	4.02	4.42	−.40[a]

[a]Difference gave $p < .05$ by t test.

ody pair type × Changed note position interaction both were once again significant, F (1,115) = 39.8 and F (4,115) = 2.65, respectively, MSe = .099 in both cases.

DISCUSSION

The purpose of Experiment 3 was to examine the existence of asymmetric similarity between scalar and nonscalar melodies with the factor of semitone intervals removed. The results of the experiment confirmed asymmetric similarity, showing that the presence of semitone intervals in melodies, and the greater number of such intervals in nonscalar than scalar melodies, is not a necessary prerequisite for this effect to occur. This is consistent with the perceived-alternatives hypothesis, which makes no explicit reference to semitone intervals as a factor in asymmetric similarity.

Notwithstanding that semitone intervals in nonscalar melodies are not necessary to produce asymmetric similarity, it is noteworthy that the magnitude of asymmetric similarity was rather small in this experiment—averaging over serial position, its magnitude was only .25 on the 6-point scale. This could mean either of two different things. First, the prevalence of semitone intervals in nonscalar melodies might have contributed to asymmetric similarity in Experiments 1 and 2. If so, the absence of such intervals in the present experiment would have reduced the effect in magnitude. Second, as remarked previously, the "nonscalar" melodies used in this experiment were merely relatively nonscalar as opposed to strictly nonscalar. Hence, they occasionally might have been *heard* as scalar. This would have weakened asymmetric similarity.

Experiment 4

The general aim of this final experiment was to establish more firmly the critical finding of Experiment 2. That critical finding was that nonscalar–scalar pairs were not only judged more similar than scalar–nonscalar pairs, but also were more often judged as having the same contour. This effect observed with contour judgments was important in ruling out the idea that task-specific strategies produce asymmetric similarity. However, if the effect proved limited in its generalizability, a task-specific strategy hypothesis might be resurrected.

We examined two specific issues. The first was whether contour judgments would show the asymmetric similarity effect with the factor of semitone intervals controlled. Experiment 3 controlled semitone intervals, and obtained only a small asymmetry in similarity judgments. We wondered if controlling for semitone intervals would *remove* the asymmetry in *contour* judgments.

The second issue was whether contour judgments would show asymmetric similarity when subjects do *not* also make similarity judgments. Experiment 2 used a dual-task procedure whereby subjects responded to each melody pair with both a similarity rating and a same–different contour judgment. This dual-task procedure might have produced task demands such that rating pairs for similarity influenced contour judgments. If so, the asymmetry effect shown with the contour judgments might have been an artifact and not truly contrary to a task-specific strategy hypothesis.

METHOD

Subjects and Procedure

Forty subjects served in group sessions. There were 21 subjects in the inexperienced group and 19 in the moderately experienced group. They received a list of 120 melody pairs, and judged whether the melodies of each pair had the same or different contour, using a six-point scale as in Experiment 2 (6 = sure same contour, 1 = sure different contour).

Stimuli

Stimuli were produced as in Experiment 3, except for the following changes: First, all seed melodies included the pitch B as well as E, G, A, D, and C (serial positions 1 and 7). This meant that all melodies—both scalar and nonscalar—contained exactly one one-semitone interval; that between B and C. The note B was included because: (a) we wanted to maintain the use of six different pitches per melody as in Experiments 1 through 3, and (b) the contour changes (see below) prevented the use of both the upper and lower E within the same melody (the tactic of Experiment 3).

Second, half the SN and NS melody pairs contained a change in contour. Contour changes were accomplished by substituting the E above middle C in one melody for the E below middle C in the other, or vice versa (randomly determined and occurring equally often). Same contour trials used either the high or low E, randomly determined.

Third, the melodies were presented at the (slightly reduced) rate of 2.0 notes/sec. Each note was 0.44 sec in duration, with internote intervals of 0.06 sec. As before, a 2-sec interval separated the standard and comparison melodies on a trial, and subjects had 8 sec to respond before the start of the next trial.

The 120-trial list comprised five 24-trial blocks in which a melody pair derived from each of the 24 seed melodies appeared a single time. Across the five blocks, each seed melody was: (a) represented by a pair with the changed note occupying each of the five internal serial positions (2 through 6), and (b) presented in the SN order in three blocks and in the NS order in two, or as in the SN order in two blocks and the NS order in three (as in Experiment 3). In addition, a seed was tested three times as a same-contour item and twice as a different-contour item, or else vice versa. Finally, there were four different versions of the 120-trial list, and across these four versions, items were counterbalanced with respect to both melody-pair type (SN vs. NS) and contour condition (same vs. different). Each of the four list versions was used with a subgroup of 8 to 12 subjects.

After partitioning based on musical experience, there were between three and seven subjects in each of the eight cells (2 Experience levels × 4 List versions). Because of the varying cell sizes we report an ANOVA of the results using List version as a variable.

RESULTS

As in Experiment 2, we performed separate ANOVAs of subjects' average contour ratings, and of the average proportions of same-contour decisions (i.e., average proportions of ratings of 4, 5, and 6 on the six-point scale). Both ANOVAs included the counterbalancing variable of List version and as well as Musical experience of subjects, with Melody pair type (SN vs. NS) and same-versus-different contour as within-subjects variables. Due to the inclusion of the contour variable, we collapsed over the variable of changed note position to avoid the impractically small number of six observations per cell.

The ANOVA of the contour ratings produced a robust main effect for Melody pair type, $F(1, 32) = 21.1$, $MSe = .125$, a more robust main effect for same-versus-different Contour, $F(1, 32) = 221.4$, $MSe = .781$, and a strong interaction between these two variables, $F(1, 32) = 22.6$, $MSe = .050$. As shown in Table 6, contour ratings were higher in the nonscalar–scalar condition than in the scalar–nonscalar condition, especially for same-contour items. This is identical to the pattern that was shown in Experiment 2 (Table 4).

The ANOVA of contour ratings also showed a same-versus-different Contour × Musical experience interaction, $F(1, 32) = 4.44$, $MSe = .781$, reflecting the fact that the moderately experienced subjects exceeded the inexperienced subjects in average contour ratings to same-contour items (4.52 vs. 4.17), but fell below inexperienced subjects in average contour ratings to different-contour items (2.08 vs. 2.33). This pattern reflects simply slightly greater accuracy in making contour ratings on the part of the more experienced subjects.

TABLE 6

Average Contour Ratings and Average Proportion of Same-Contour Decisions in Experiment 4

Type of Rating	Same Contour: Scalar–Nonscalar	Same Contour: Nonscalar–Scalar	Different Contour: Scalar–Nonscalar	Different Contour: Nonscalar–Scalar
Contour Ratings	4.19	4.60	2.18	2.27
Same-Contour Decisions	.69	.77	.16	.18

A final outcome of the ANOVA of contour ratings was a three-way interaction of Melody pair type × same-versus-different Contour × List version, $F(3, 32) = 6.54$, qualified by the four-way interaction involving Musical experience, $F(3, 32) = 8.75$, $MSe = .050$ in both cases. Perusal of the data suggested that, for reasons we cannot explain, the moderately experienced subjects receiving one of the four lists did not show the tonality condition effect, whereas the inexperienced subjects receiving this same list showed a particularly large tonality condition effect. We view this pattern as a fluke and will not discuss it further.

The ANOVA of the proportions of same-contour judgments (Table 6, Row 2) produced the same pattern of effects as that of contour ratings. The only exception was that the Musical experience × Contour interaction did not attain significance ($p > .10$). Thus, the interaction obtained with the average contour ratings should be interpreted cautiously.

To confirm the reliability of our major findings over items, we performed an additional ANOVA on the contour ratings, using items as the random factor. The two factors were Melody pair type and Contour, both of which produced reliable main effects, $F(1, 119) = 48.4$ and 1405.3, and $MSe = .199$ and .364, respectively. The interaction was also reliable, $F(1, 119) = 13.5$, $MSe = .222$, confirming the pattern of the subjects-as-random ANOVAs.

General Discussion

These four experiments have extended prior evidence for asymmetric similarity of scalar and nonscalar melodies—pairs of more and less scalar melodies are perceived as more similar when the less scalar melody comes first. At the same time, the present studies have strengthened a perceived-alternatives hypothesis for the processes underlying a person's sense of scale.

According to the perceived alternatives hypothesis, the ordinary process of perceiving a melody involves its implicit comparison to a set of alternative melodies (cf. Garner, 1974). This set of perceived alternative melodies preserves some of the original melody's attributes (i.e., those of its attributes that are perceived as "immutable," Kahneman & Miller, 1986). To perceive a melody as scalar or "in key" is to perceive its key, or the set of pitch-classes that constitute this key, as one of its immutable attributes. Hence, to perceive a melody as scalar or in key is to compare it to a set of alternative melodies based on the pitch classes of its key. To perceive a melody as nonscalar, or as containing "wrong notes," is to compare it to a larger set of alternatives based on a larger set of pitch classes, perhaps all 12 pitch classes of the chromatic set.

The preceding analysis implies that if two melodies are highly similar, but differ in their scalar structure, their sets of alternatives are likely to be "nested" (Handel & Garner, 1966). That is, the alternatives of the less scalar items are likely to include the scalar item, but not vice versa. Since nesting implies asymmetric relatedness, the phenomenon of asymmetric similarity is accommodated nicely. Specifically, we have argued that when a scalar and nonscalar melody are presented with the nonscalar melody first, the second melody is perceived as falling outside the alternative set of the first. The consequence is a global impression of difference which affects the perceived similarity of the melodies, as well as other judgments such as same-versus-changed contour.

The perceived-alternatives hypothesis made three clear predictions that distinguished it from other views. A first prediction concerned melody pairs which differed in the same way as scalar–nonscalar and nonscalar–scalar pairs, but which consisted of two scalar items: Such scalar–scalar pairs should produce *no* perception of second melodies violating perceived alternatives of first melodies. Hence, no global impression of difference should occur. In line with this reasoning, Experiment 1 showed that scalar–scalar pairs were rated as being equally as similar as nonscalar–scalar pairs, and as more similar than scalar–nonscalar pairs.

Confirming this prediction was important not only for supporting the perceived alternatives hypothesis, but for ruling out the competing idea that asymmetric similarity reflects the poor memorability of nonscalar melodies. Such poor memorability might be a result of two factors: (a) the assimilation of nonscalar melodies to internal scale schemata (Dowling, 1978), or (b) memory interference involving semitone intervals (Deutsch, 1982), which are more prevalent in nonscalar than scalar melodies. In either case, melody pairs in which both items are scalar should receive low similarity ratings. Since in fact such pairs received high similarity ratings, both the schema-assimilation and memory-interference notions were compromised by the study's findings. Although it is plausible that nonscalar melodies are more difficult to remember than scalar melodies, and that scale schemata as well as memory interference might contribute to this memorability difference, the available data rule against the idea that differential memorability provides the explanation for the asymmetric similarity effect. Thus, in rejecting the memorability hypothesis, we are not asserting that schemata and/or one-semitone intervals play no role in melody processing. We are simply suggesting that our data require the concept of perceived alternatives in addition to these other ideas.

A second prediction of the perceived alternatives hypothesis was that the asymmetric similarity effect, as well as the high similarity of scalar–scalar pairs, should be reflected not only in similarity judgments—which might be subject to task-specific strategies—but in judgments regarding an attribute of melodies that is logically independent of scalar structure. One such attribute is melodic contour, and Experiments 2 and 4 confirmed that nonscalar–scalar pairs were more often judged as sharing contour than scalar–nonscalar pairs, especially when contour was actually identical. Experiment 2 showed in addition that scalar–scalar pairs were judged to be matching in contour as often as nonscalar–scalar pairs and more often than scalar–nonscalar pairs. These findings were critical for establishing the validity of the similarity ratings used in Experiments 1 and 3. Without the converging findings of the contour judgments, it would have been arguable that a task-specific strategy for judging similarity was responsible for the asymmetric similarity effect. Armed with these findings, we can conclude that the processes underlying asymmetric similarity are automatic if not mandatory—they appear to affect perceived similarity regardless of the listener's task or intentions.

A third prediction of the perceived alternatives hypothesis concerns the troublesome factor of one-semitone intervals, which are generally more prevalent in nonscalar melodies than in scalar melodies. There probably exist a variety of reasons why an imbalance in one-semitone intervals might contribute to asymmetric similarity. However, the perceived-alternatives view implies asymmetric similarity when this imbalance is corrected. Indeed, Experiments 3 and 4 employed scalar and nonscalar melodies that were equated for one-semitone intervals, and showed once again that nonscalar–scalar melody pairs are rated as more similar, and are more often judged as having the same contour, as compared to scalar–nonscalar melody pairs. Although we were unable to use strictly nonscalar melodies in these studies, being forced to use melodies that were only relatively nonscalar (i.e., incompatible with the prevailing key of C), this (necessary) limitation of our experimental design could only weaken the finding of asymmetric similarity. In fact, we observed that the effect was rather weak, but that it was nonetheless reliable (Experiment 3), and even generalized to contour judgments (Experiment 4). It therefore appears that there is asymmetric similarity when semitone intervals are controlled. There clearly is more to asymmetric similarity than a confounding of scalar structure with semitone intervals.

Although the data in hand are largely consistent with a perceived alternatives notion, we see two potential problems for the survival of this notion. The first problem is empirical; it concerns the relatively low similarity of melody pairs in which both members are nonscalar. Such nonscalar–nonscalar melody pairs were examined in Experiment 1, in which their rated similarity was approximately intermediate between that of scalar–nonscalar pairs (which were the least similar) and that of scalar–scalar and nonscalar–scalar pairs (which were the most similar). Such pairs were also examined in Experiment 2, in which their rated similarity was relatively lower, that is, closer to scalar–nonscalar pairs than to scalar–scalar and nonscalar–scalar pairs. Indeed, looking at the contour ratings and proportions of same-contour judgments of Experiment 2 (Table 4), the nonscalar–nonscalar pairs appeared quite similar to the (least similar) scalar–nonscalar pairs. The ANOVAs of average contour ratings and proportions of same-contour decisions supported this observation—both ANOVAs supported a main effect of Second melody type, with no main effect of First melody type and no interaction (the ANOVA of the similarity ratings—in Experiment 2 as well as Experiment 1—supported all three effects). Thus, the contour ratings and same-contour judgments suggested less perceived similarity when the second melody was nonscalar, regardless of whether the first melody was scalar or nonscalar. This is not what a perceived alternatives hypothesis, as formulated heretofore, would lead one to predict—the simplest perceived-alternatives hypothesis implies no violations of perceived alternatives with nonscalar–nonscalar pairs. Hence, the perceived similarity of such pairs should be high, perhaps as high as nonscalar–scalar and scalar–scalar pairs.

In order to handle the relatively low similarity of nonscalar–nonscalar pairs, the perceived-alternatives hypothesis requires elaboration. Perhaps the simplest elaboration is to assume that nonscalar notes in otherwise scalar melodies are essentially ambiguous—they can be interpreted either as truly out-of-key notes, or as pitch-altered versions of in-key notes. Figure 2 uses Shepard's (1982) melodic map to illustrate this point. A nonscalar note *might* be interpreted as falling outside the scale-defined map region containing the other notes (Figure 2a), or it might be interpreted as a variant of a scalar note with its pitch altered by one semitone (Figure 2b). It is intuitively compelling that when a nonscalar note in an otherwise scalar melody is initially encountered, it has a jarring quality which indicates its interpretation as an out-of-key note (Figure 2a). However, after subsequent key-

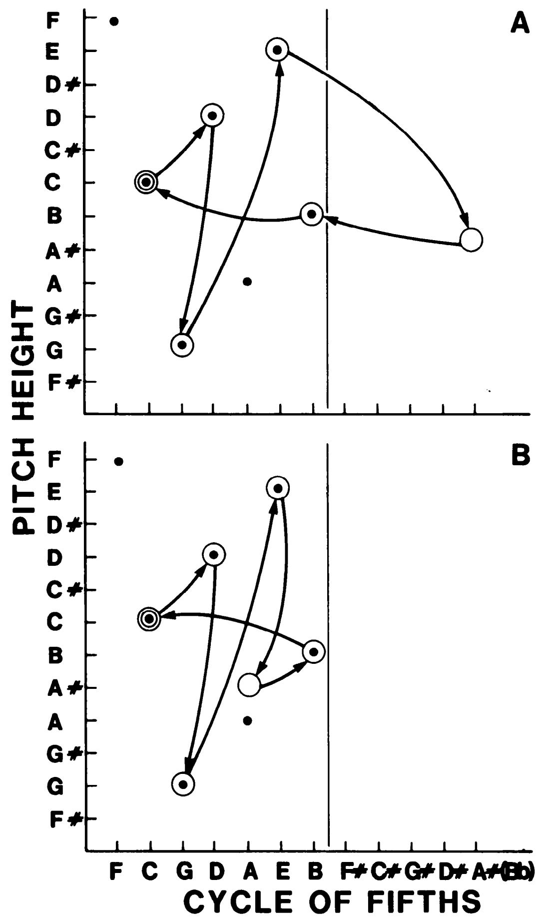

Fig. 2. Melodies represented as in Figure 1: (a) the melody C–D–G–E–A♯–B–C in which the A♯ is interpreted as an out-of-key pitch; (b) the same melody in which A♯ is interpreted as a within-key variant of A.

consistent notes, the nonscalar note tends to be remembered as a variant of an in-key note matching the prevailing key context. Bharucha's (1984) work on anchoring can be seen as illustrating a special case of this principle—the case where a nonscalar note is immediately or almost immediately followed by a scalar note only one-semitone removed in pitch.

Returning to ratings of nonscalar–nonscalar pairs, the preceding arguments have an interesting implication: At the time that the subject is comparing the two melodies (i.e., during or after presentation of the second melody), the nonscalar note of the first presented melody may have been interpreted as scalar, whereas the nonscalar note of the second-presented melody is likely to be interpreted as nonscalar, with the consequence that perceived similarity is low. Although this reasoning is post hoc, future research might fruitfully address the cognitive processes that produce and resolve the ambiguous "meanings" of nonscalar notes.

A second threat to the perceived alternatives notion is more theoretical than empirical. A different and apparently contradictory perspective on asymmetric similarity melodies derives from Tversky's (1977) set-theoretic account. According to Tversky, asymmetrical similarity between two stimuli can occur if two conditions are met: First, the two stimuli must be unequal in "salience" such that the features of one must be more noticeable or prominent, or greater in number, than those of the other. Second, there must be "asymmetry of focus" (selective attention) on one of the two stimuli, so that its distinctive features are more heavily weighted in the judgment of similarity. If a pair of stimuli differ in salience, similarity should be lower if there is selective focus on the more salient of the two. Tversky goes further to argue that "good form" contributes to the salience of a stimulus (i.e., stimuli that are good gestalts have highly salient features), and also marshals evidence (Rothkopf, 1957; Wish, 1967) that when subjects make same–different judgments to pairs of auditory sequences, they selectively focus on the first sequence of a pair at the expense of the second sequence of a pair. If we assume that scalar melodies are possessed of good form, it follows they should also have a high degree of salience, which would produce the finding of lower similarity for scalar–nonscalar pairs—in which the more salient stimulus receives selective focus—than for nonscalar–scalar pairs—in which the less salient stimulus receives selective focus.

A problem with the set-theoretic argument, as formulated above, concerns the claim that a nonscalar melody is necessarily less salient than a scalar melody with which it is paired. Intuitively, the nonscalar melody of a scalar–nonscalar pair sounds strikingly wrong and ugly, which is not the same thing as having low salience. Rather than arguing that scalar structure affects salience, we propose that the set of alternatives evoked by a melody—the region of the melodic map that encapsulates this melody—can act as a "feature" affecting similarity judgments. A corollary claim is that while the small set of alternatives evoked by a scalar melody is violated by its nonscalar mate, the large set of alternatives evoked by a nonscalar melody is *not* violated by its scalar mate. Thus, the alternative-set of a scalar melody functions as a "distinctive feature," whereas the alternative-set of a nonscalar melody either functions as a "common feature" or does not function as a feature at all. If we follow Tversky (1977) in proposing that features of the first melody are weighted more heavily than features of the second, we have an account of asymmetric similarity of melodies: A more scalar melody and a less scalar melody are perceived as less similar if the more scalar melody, the one whose set of perceived alternatives functions as a distinctive feature, is presented first and therefore receives selective focus.

This account is compatible with Tversky's (1977) set-theoretic framework, the main change being that it does not appeal to "salience." It also is compatible with spatial models of musical pitch, especially that of Shepard (1982, Figure 1), suggesting that set-theoretic and spatial models need not be viewed as conflicting.

A final matter to discuss concerns the contour ratings and same-contour judgments in Experiments 2 and 4. Both studies showed that judged sameness of contour was higher for nonscalar–scalar pairs than for scalar–nonscalar pairs. However, these effects were apparently stronger for same-contour items than different–contour items (this was evidenced by a Second melody type × Contour interaction in Experiment 2, and by a Melody pair type × Contour interaction in Experiment 4). A possible explanation is that our scalar-structure manipulations affect perceived similarity of melodies, which in turn affects contour judgments in two ways. First, high perceived similarity causes a bias to judge two melodies "same." Second, perceived similarity actually improves the ability of subjects to compare the two melodies—melodies perceived as essentially similar might more accurately judged for same-versus-different contour than melodies perceived as different. These two effects would reinforce each other with same-contour items, but would work against each other with different-contour items (for which "same-contour" judgments are errors).

An implication of the preceding analysis is that the accuracy of discriminating same- from different-contour pairs should be higher in the nonscalar–scalar condition than in the scalar–nonscalar condition. To test this possibility, we used subjects' contour ratings to derive area-under-the-MOC scores (Swets, 1973) for the nonscalar–scalar and scalar–nonscalar conditions in Experiments 2 and 4 (Area under the MOC scores were computed from the rating data supplied by each subject within each condition). The resulting means are shown in Table 7, where there is a trend for better discrimination in the nonscalar–scalar condition. This trend is reliable. ANOVAs comparing the scalar–nonscalar and nonscalar–scalar conditions showed a main effect of Melody pair type in Experiment 2, F (1, 48) = 7.41, MSe = .015, as well as in Experiment 4, F (1, 45) = 14.0, MSe = .003 (the latter ANOVA showed interactions between Melody pair type and List, and among Melody pair type, List, and Musical Experience—these interactions reflect the same anomalous pattern described previously in the Results section of Experiment 4). Thus, the evidence is reasonably clear that the ordering of scalar and nonscalar melodies affects not only the overall tendency to make same-contour judgments, but the accuracy of those judgments as well. This is suggestive that strong perceived similarity between two melodies has two different effects; that of producing a bias toward "same" judgments, and that of improving comparisons of the melodies. Future research must further examine the ways in which tonality in-

TABLE 7

Mean Area under MOC Scores for Same- versus Different-Contour Discrimination in Experiments 2 and 4

	Scalar-Structure Condition	
Experiment	Scalar–Nonscalar	Nonscalar–Scalar
Experiment 2	.73	.78
Experiment 4	.81	.85

NOTE: Area under MOC scores generally vary from .50 (chance discrimination) to 1.0 (perfect discrimination).

formation enters in to contour judgments. It seems fair to say that no existing theories, aside from the perceived-alternatives hypothesis, would have predicted that scalar structure might be involved in such judgments.

Future research on scalar structure of melodies should include converging operations on the perceived alternatives for scalar and nonscalar melodies. Asymmetric similarity is but one source of evidence—albeit a strikingly counterintuitive source—that scalar melodies have few perceived alternatives.[3]

References

Bartlett, J. C. Cognition of complex events: Visual scenes and music. In W. R. Crozier and A. J. Chapman (Eds), *Cognitive processes in the perception of art.* Amsterdam: North Holland, 1984, pp. 225–251.

Bharucha, J. J. Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology,* 1984, *16,* 485–518.

Bharucha, J. J. & Krumhansl, C. L. The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition,* 1983, *13,* 63–102.

Boring, E. G. *Sensation and perception in the history of experimental psychology.* New York: Appleton-Century-Crofts, 1942.

Cross, I., Howell, P., & West, R. Preferences for scale structure in melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance,* 1983, *9,* 444–460.

Cross, I., Howell, P., & West, R. Structural relationships in the perception of musical pitch. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition.* New York: Academic Press, 1985, pp. 121–142.

Deutsch, D. The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music.* New York: Academic Press, 1982, pp. 271–316.

Dowling, W. J. Scale and contour: Two components of a theory of memory for melodies. *Psychological Review,* 1978, *85,* 341–354.

Dowling, W. J. Tonal structure and children's early learning of music. In J. Sloboda (Ed.), *Generative processes in music.* London: Oxford, 1988, pp. 113–128.

3. We thank Reid Hastie for valuable comments on this research, and Deborah Dayboch for initially calling our attention to the phenomenon.

Dowling, W. J. & Bartlett, J. C. The importance of interval information in long-term memory for melodies. *Psychomusicology,* 1981, ***1***, 30–49.
Dowling, W. J. & Fujitani, D. S. Contour, interval and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America,* 1971, ***49***, 524–531.
Dowling, W. J. & Harwood, D. L. *Music cognition*. New York: Academic Press, 1986.
Francès, R. *The perception of music* (translated by W. J. Dowling). Hillsdale, N.J.: Erlbaum, 1988. (Original French edition published 1958.)
Garner, W. R. Good patterns have few alternatives. *American Scientist,* 1970, ***58***, 34–42.
Garner, W. R. *The processing of information and structure*. Hillsdale, N.J.: Erlbaum, 1974.
Handel, S. & Garner, W. R. The structure of visual pattern associates and pattern goodness. *Perception and Psychophysics, 1966,* ***1***, 33–38.
Kahneman, D. & Miller, D. T. Norm theory: Comparing reality to its alternatives. *Psychological Review,* 1986, ***92***, 136–153.
Krumhansl, C. L. The psychological representation of musical pitch in a musical context. *Cognitive Psychology,* 1979, ***11***, 346–374.
Krumhansl, C. L., & Keil, F. C. Acquisition of the hierarchy of tonal functions in music. *Memory and Cognition,* 1982, ***10***, 243–251.
Rothkopf, E. Z. A measure of stimulus similarity and errors in some paired-associates learning tasks. *Journal of Experimental Psychology,* 1957, ***53***, 94–101.
Shepard, R. N. Geometrical approximations to the structure of musical pitch. *Psychological Review,* 1982, ***89***, 305–333.
Swets, J. A. The relative operating characteristic in psychology. *Science,* 1973, ***182***, 990–1000.
Tversky, A. Features of similarity. *Psychological Review,* 1977, ***84***, 327–352.
Wish, M. A model of the perception of Morse Code-like signals. *Human Factors,* 1967, ***9***, 529–540.