Music Perception
Summer 2004, Vol. 21, No. 4, 457–498

Musical Forces and Melodic Expectations: Comparing Computer Models and Experimental Results

STEVE LARSON
University of Oregon School of Music

Recent work on "musical forces" asserts that experienced listeners of tonal music not only talk about music in terms used to describe physical motion, but actually experience musical motion as if it were shaped by quantifiable analogues of physical gravity, magnetism, and inertia. This article presents a theory of melodic expectation based on that assertion, describes two computer models of aspects of that theory, and finds strong support for that theory in comparisons of the behavior of those models with the behavior of participants in several experiments.

The following summary statement of the theory is explained and illustrated in the article:

> Experienced listeners of tonal music expect completions in which the musical forces of gravity, magnetism, and inertia control operations on alphabets in hierarchies of embellishment whose stepwise displacements of auralized traces create simple closed shapes.

A "single-level" computer program models the operation of these musical forces on a single level of musical structure. Given a melodic beginning in a certain key, the model not only produces almost the same responses as experimental participants, but it also rates them in a similar way; the computer model gives higher ratings to responses that participants sing more often. In fact, the completions generated by this model match note-for-note the entire completions sung by participants in several psychological studies as often as the completions of any one of those participants matches those of the other participants.

A "multilevel" computer program models the operation of these musical forces on multiple hierarchical levels. When the multilevel model is given a melodic beginning and a hierarchical description of its embellishment structure (i.e., a Schenkerian analysis of it), the model produces responses that reflect the operation of musical forces on all the levels of that hierarchical structure.

Statistical analyses of the results of a number of experiments test hypotheses arising from the computer models' algorithm (S. Larson, 1993a) for the interaction of musical forces as well as from F. Lerdahl's similar (1996) algorithm. Further statistical analysis contrasts the explanatory power of the theory of musical forces with that of E. Narmour's (1990, 1992) implication-realization model.

Address correspondence to Steve Larson, University of Oregon, School of Music, 1225 University of Oregon, Eugene, OR 97403-1225. (e-mail: steve@uoregon.edu)
ISSN: 0730-7829. Send requests for permission to reprint to Rights and Permissions, University of California Press, 2000 Center St., Ste. 303, Berkeley, CA 94704-1223.

The striking agreement between computer-generated responses and experimental results suggests that the theory captures some important aspects of melodic expectation. Furthermore, the fact that these data can be modeled well by the interaction of constantly acting but contextually determined musical forces gives support to the idea that we experience musical motions metaphorically in terms of our experience of physical motions.

Received February 9, 1996, accepted December 13, 2003

Listening to music is a creative process in which we shape the sounds we hear into meanings tempered by our nature and experience. When we are engaged in that process, experienced listeners of tonal music make predictions about what will happen next—about "where the music is going." In this sense, the recent growth of interest in melodic expectation, both in theoretical and experimental research, responds to central questions about musical experience—questions about meaning and motion.

This article draws on approaches from a variety of fields to make a contribution to our understanding of melodic expectation and its role in musical meaning and musical motion: at the intersection of current music theory and philosophy of meaning, it begins with a theory of metaphorical "musical forces"; from cognitive science, especially from artificial intelligence, it adapts the approach of building computer models of cognitive processes; and from music psychology and statistical analysis, it borrows techniques for analyzing the results of psychological experiments. This article thus synthesizes all these fields; the theory of melodic expectation presented here is implemented in computer models whose behavior is compared with that of participants in music-cognition experiments.

A Theory of Melodic Expectation

The theory of melodic expectation presented here claims that experienced listeners of tonal music have expectations about how melodic beginnings will be completed, and it claims that important aspects of those expectations are captured in the following summary statement:

> Experienced listeners of tonal music expect melodic completions in which the musical forces of gravity, magnetism, and inertia control operations on alphabets in hierarchies of embellishment whose stepwise displacements of auralized traces create simple closed shapes.

The following paragraphs explain what is meant by this summary statement, restate portions of it as specific rules, and then describe computer implementations of certain aspects of the theory. A complete specification of the theory and its complete implementation in a single computer program are long-range goals. However, for now, the success of two partial implementations is enough to suggest its power to account for the results of several different psychological experiments.

ENTIRE COMPLETIONS VERSUS MERE CONTINUATIONS

The first claim in this summary statement is that listeners expect *completions*. This emphasis on completions, as opposed to mere continuations, separates this claim from many current experimental and theoretical studies of music perception.

Some experimental work in music perception restricts attention to the first new note that listeners expect in a melodic continuation. Carlsen and his collaborators asked participants to sing continuations of two-note beginnings (Carlsen, 1981; Carlsen, Divenyi, & Taylor, 1970; Unyk & Carlsen, 1987). Cuddy and her collaborators have tested listeners' judgments (Cuddy & Lunney, 1995) and listeners' produced continuations (Thompson, Cuddy, & Plaus, 1997) of two-note beginnings. Lake (1987) and Povel (1996) also asked participants to produce continuations of (respectively) two- or one-note beginnings, but for each beginning, they first established a major-key context. In Povel's experiment, participants were allowed to add only one note. In the rest of these experiments in which participants produced continuations, the participants often added more than one new note, but the experimenters analyzed only the first added note.

Experiments by Krumhansl and others that ask listeners to judge how well a single probe tone or chord fits an established context may also be regarded as asking listeners to rate the degree to which that tone or chord is expected in that context (Krumhansl, 1990). Some of these studies (Cuddy & Lunney, 1995; Krumhansl, 1995; Schellenberg, 1996, 1997) used the probe-tone technique to test predictions of the bottom-up component of Narmour's (1990, 1992) "implication-realization model" of melodic expectancy. Such probe-tone experiments also limit analysis to only the first new element expected by listeners.

However, other experiments (Larson, 1997a) suggest that we should regard the melodic expectations of participants as expectations not so much for *continuations*, but as expectations for entire *completions*. (By "entire completions," I mean *all* the notes sung by an experimental participant.) In these experiments, most participants' responses agreed note-for-note with another participant's *entire* response—and the total number of different entire responses was small. (Some responses failed to agree note-for-note with another response, but since our interest lies in understanding shared musical intuitions, the elimination of such outliers may be regarded as a positive feature of considering entire completions.) Furthermore, these other experiments demonstrate that sorting responses by just the first added notes can confuse the data—by making clearly similar responses look different and by making clearly different responses look similar. Larson (1997a) shows, for example, that responses that end in the same key and have the same essential structure can have different first added notes, and that responses in different keys and ending on different notes can have the same first added notes.

Until now, theories of melodic expectation have not offered a testable explanation of how listeners generate entire completions. For example, Narmour's implication-realization model suggests that melodic continuations result from the interaction of "bottom-up" and "top-down" components. However, tests of the bottom-up component of the implication-realization model (cited above) suggest that it is little more than a description of what is statistically true of first added notes in general—like claiming that "added notes are usually close in pitch to one of the two preceding notes" and "large leaps are usually followed by a change in direction." In fact, Schellenberg (1996, 1997) finds that the bottom-up component of Narmour's model can be simplified to something like these two statements without any loss of accuracy in describing experimental results.

It might appear that including the top-down component of Narmour's model would allow it to generate entire completions. In fact, Krumhansl (1995) found that, in order to account well for the data in one of her tests of Narmour's model, she had to add a factor she called "tonality" (which she modeled with her major-key profile)—and that factor turned out to be one of the most important in explaining the data. (In her experiment with tonal melodies, this was true even though the experiment considered only diatonic continuations. Had a more complete test been done, also considering nondiatonic tones, this factor may well have played an even greater role.) However, a close reading of Narmour's books suggests that the top-down component—including factors like "the influence of intra-opus style" and "the influence of extra-opus style"—includes more than Krumhansl's "tonality." Despite the elegance of Krumhansl's experiments, the top-down component of Narmour's model is still not codified as a set of rules capable of generating entire completions.

In fact, by itself, the bottom-up component of Narmour's model leads to two seemingly contradictory conclusions: the first is that the larger an implicative interval, the *stronger* its implications; the second is that the larger an interval, the greater the number of implied continuations (that is, possible "reversals"), thus the *less specific* our expectations. Since, in experience, stronger implications are usually *more* specific, this also suggests that, if Narmour's model is to capture important aspects of melodic expectation, then the top-down component may need to be included.

In order to generate entire completions, a theory of melodic expectation would have to be articulated as a specific set of rules—which could be implemented as a computer program—capable of generating a list of entire completions (and, ideally, assigning them ratings that suggest their likelihood). The closer that program comes to matching note-for-note all and only the continuations produced or approved by experimental participants, the more successful it will appear (and the closer the ratings attached to the computer-generated responses correlate with the frequency with which those responses are produced, or with the judgments made about them, the more successful it will appear.)

This article proposes just such a theory, describes two computer models that implement aspects of that theory, and shows that a comparison of the results of that model with the results of psychological experiments offers strong support for that theory.

MUSICAL FORCES

The summary statement given above claims that musical forces play an important role in generating melodic completions. A number of recent papers and presentations suggest that musical forces shape musical experience.

> Three of these forces I call "gravity" (the tendency of an unstable note to *descend*), "magnetism" (the tendency of an unstable note to move to the *nearest* stable pitch, a tendency that grows stronger the closer we get to a goal), and "inertia" (the tendency of a pattern of musical motion to continue in the *same* fashion, where what is meant by "same" depends upon what that musical pattern is "heard as"). (Larson 1997b, p. 102)

The idea of musical forces was inspired by Rudolf Arnheim's applications of gestalt psychology to the perception of visual art (Arnheim, 1966, 1974, 1986; Larson, 1993c); has illuminated aspects of Schenkerian theory (Larson, 1994a, 1997b); has improved the pedagogy of aural skills, counterpoint, and harmony (Hurwitz & Larson, 1994; Larson, 1993d, 1994a; Pelto, 1994); has been used in the analysis of pieces by Chopin, Brahms, and Varese (Brower, 1997–98, 2000); and has helped to explain the phenomenon of "swing" in jazz (Larson, 1999b). The idea of musical forces has also been used to generate a small, well-defined set of three-, five-, and seven-note patterns—and a comparison of that set with the set of patterns discussed in published accounts of hidden repetition in tonal music, with a set of patterns said to structure fugue expositions, and with a set of patterns described as first-level Schenkerian middlegrounds gives strong support to the claim that this set of patterns is privileged in tonal music (Larson, 1997–98). A recent article in this journal (Larson, 2002) relates aspects of the theory of musical forces to important work in jazz theory, analyzes recorded jazz compositions and improvisations in light of the theory, and finds that the distribution of melodic patterns in that music gives further empirical support to the theory.

The theory argues that we experience musical motions metaphorically in terms of our experience of physical motions. According to the theory, the metaphor of musical motion is neither optional nor eliminable. This metaphor and its entailments (which include the musical forces) shape musical experience in direct, profound, and consistent ways. This view of metaphor as constitutive of experience agrees with some current work in philosophy and cognitive linguistics (Johnson, 1987; Lakoff & Johnson, 1980, 1999). It is also supported by recent studies of metaphor and embodiment by music analysts (Aksnes, 1997, 2002; Coker, 1972; Cox, 1999; Guck, 1981, 1991; Kassler, 1991; and the work on musical forces cited above). However, this theory of musical experience should not be taken as making any assertions about whether these "forces" are learned, innate, or some combination thereof.

In some situations, the musical forces agree: In a context where we expect melodies to move within the major scale and where we experience the members of the tonic triad as stable pitches, musical gravity suggests that the melodic beginning $\hat{5}$-$\hat{4}$-? will continue by going *down*, $\hat{5}$-$\hat{4}$-$\hat{3}$; musical magnetism suggests that the same beginning will continue by going to the *nearest* stable pitch, $\hat{5}$-$\hat{4}$-$\hat{3}$; and musical inertia suggests that the same beginning will continue by going in the *same* direction, $\hat{5}$-$\hat{4}$-$\hat{3}$.

In other situations, the forces may disagree: In a context where we expect melodies to move within the major scale and where we experience the members of the tonic triad as stable pitches, musical gravity and musical magnetism suggest that the melodic beginning $\hat{5}$-$\hat{6}$-? will continue by going down and to the nearest stable pitch, $\hat{5}$-$\hat{6}$-$\hat{5}$; but musical inertia suggests that the same beginning will continue by going in the same direction, $\hat{5}$-$\hat{6}$-$\hat{7}$-$\hat{8}$.

The interaction of the forces may be contextually determined (e.g., a specific context may heighten listeners' attention to the effects of gravity while another context may lessen the impact of the same force). However, for the purposes of computer modeling, the theory offers an algorithm for their interaction (Larson, 1993a). The general form of that algorithm is

$$F = w_G G + w_M M + w_I I$$

This equation shows how, for a particular note in a particular context, the computer can represent the forces felt to impel that note to another specified note. It combines the scores it gives to gravity (G is 1 for patterns that give in to gravity and 0 for motions that do not), magnetism (M is calculated according to a formula discussed in more detail later), and inertia (I is 1 for patterns that give in to inertia, –1 for patterns that go against inertia, and 0 if there are no inertial implications) in a proportion (represented by the weights w_G, w_M, and w_I, respectively) to represent their cumulative effect (F). Just as we experience physical motions as shaped by an interaction of constantly acting but contextually determined physical forces, so, the theory argues, we experience musical motions as shaped by an interaction of constantly acting but contextually determined musical forces.

Earlier computer models of musical forces (Larson, 1993a, 1994b, 1999a) represent the magnetic pull toward the closest stable pitch as the difference between the pull of that pitch (expressed as the inverse square of its distance in semitones) minus the pull of the closest pitch in the other direction (also expressed as the inverse square of its distance in semitones). In this way, the algorithm models the human tendency to be more strongly drawn to closer goals. Theoretically, one could calculate magnetic pulls from all stable pitches, but as a practical matter, only the closest notes in each direction are considered in this calculation. Thus, the computer models represent magnetism as

$$M = 1/d_{to}^2 - 1/d_{from}^2$$

where M indicates the magnetic pull on a given note in a given context in the direction of a specified goal; d_{to} is the distance in semitones to that goal; and d_{from} is the distance in semitones to the closest stable pitch (potential goal) in the other direction. In the context just described (where we expect melodies to move within the major scale and where we experience the members of the tonic triad as stable pitches), the distance (d) from $\hat{4}$ to $\hat{5}$ is 2 semitones, and the distance from $\hat{4}$ to $\hat{3}$ is 1 semitone. Thus, the magnetic pull exerted on $\hat{4}$ by $\hat{5}$ is represented as 0.25 ($1/d^2 = 1/2^2 = 1/4$), the magnetic pull exerted on $\hat{4}$ by $\hat{3}$ is represented as 1 ($1/d^2 = 1/1^2 = 1$), and their combined magnetic effect on $\hat{4}$ is 0.75 in the direction of $\hat{3}$ ($1/d_{to}^2 - 1/d_{from}^2 = 1/1^2 - 1/2^2 = 1 - 1/4 = 3/4$). In the same context, the magnetic pull exerted on $\hat{6}$ by $\hat{8}$ is represented as 0.11 ($1/d^2 = 1/3^2 = 1/9$), the magnetic pull exerted on $\hat{6}$ by $\hat{5}$ is represented as 0.25 ($1/d^2 = 1/2^2 = 1/4$), and their combined magnetic effect on $\hat{6}$ is 0.14 in the direction of $\hat{5}$ ($1/d_{to}^2 - 1/d_{from}^2 = 1/2^2 - 1/3^2 = 1/4 - 1/9 = 5/36$). Again, the computer model considers only the effects of the two closest stable pitches (above and below the unstable note).

If we take the general algorithm given above and substitute this more specific representation of magnetism, then we get the first of the three following algorithms.

$$F = w_G G + w_M(1/d_{to}^2 - 1/d_{from}^2) + w_I I \quad \text{(Larson, 1993a)}$$

$$F = w_M(a_{to}/d_{to} - a_{from}/d_{from}) + \ldots \quad \text{(Bharucha, 1996)}$$

$$F = w_M(s_{to}/d_{to}^2 - s_{from}/d_{from}^2) + w_I I \quad \text{(Lerdahl, 1996)}$$

The first of these algorithms is the one used by my computer models (Larson, 1993a, 1994b, 1999a).

The second is Bharucha's (1996) "yearning vector," with magnetism described by his "tonal force vector" (the ellipsis indicates that there may be other unspecified forces). Bharucha (1996) describes the sum of all forces acting on a given note in a given context as its "yearning vector." He describes this net force as an expectation, and explains it in terms of attention (modeled by neural-net activations) and the "anchoring" of dissonances. He does not specify what all the components of such a vector might be. If those forces were gravity, magnetism, and inertia, then his yearning vector would be equivalent to the algorithm used by my computer models to evaluate the interaction of musical forces. However, his article describes only one force—musical magnetism (which he calls the "tonal force vector"). As in my computer's algorithm, Bharucha quantifies the net magnetic pull on a note as the difference between the pulls of the closest note above and the closest note below that note. As in my computer's algorithm, he also asserts that the magnetic pull on a note from either attractor is inversely proportional to the distance between that note and its attractor. (However, while Bharucha offers a *linear* function, my computer algorithms model this pull as inversely proportional to the *square* of that distance.) He also suggests that the magnetic pull on a note from either attractor is directly proportional to the activation (represented here as a, which, in his article, is equated with the stability) of its attractor.

The third is Lerdahl's (1996) "tendency algorithm," which also quantifies the net magnetic pull on a note (which he calls "the resultant attraction") as the difference between the pulls of the closest note above and the closest note below that note. However, in a later book, Lerdahl (2001, p. 170) conjectures that "it may turn out to be more accurate just to take attractional values as correlating directly with degrees of expected continuation," which excludes the addition of opposing vectors. Nevertheless, he retains the addition of opposing vectors to quantify what he calls "the power of implicative denial."

Like my computer model's algorithm, Lerdahl quantifies this pull as inversely proportional to the square of that distance. Also, like Bharucha, Lerdahl suggests that the magnetic pull on an unstable note from either attractor is directly proportional to the stability of its attractor. However, while Bharucha considers only the stability of the attractor, Lerdahl also considers the stability of the unstable note (s_{to} is the ratio of levels of embedding in a tonal pitch space between the closest stable pitch and the unstable note; s_{from} is the ratio of levels of embedding between the closest pitch in the other direction and the unstable note). This requires calculating the stability values for three pitches: the unstable pitch and the attractors in both directions. However, instead of using the basic pitch space given in his original article (Lerdahl, 1988), Lerdahl's (1996) algorithm uses a different pitch space—one that does not distinguish between the stability of the mediant and that of the dominant. He omits gravity. Inertia (which he calls "directed motion") is modeled in similar ways in both our algorithms (however, my algorithm allows for three cases: patterns that give in to inertia score 1, patterns that go against inertia score –1, and patterns that do not have inertial implications score 0—Lerdahl's algorithm considers only the first two of these cases).

A comparison of these algorithms suggests six hypotheses concerning the ways in which the expectations of experienced listeners might reflect their intuitions about musical forces: (1) expectations are influenced by the stability of the unstable note (the attracted note); (2) expectations are influenced by the stability of the goal to which that unstable note is most strongly attracted (the attractor); (3) expectations are influenced by the stability of the closest stable pitch in the other direction (the opposing attractor); (4) such stabilities are better represented by Lerdahl's 1996 rather than by his 1988 values for pitch-space embedding; (5) magnetic pulls are better represented as inversely proportional to distance rather than inversely proportional to the square of distance; and (6) gravity does not play a necessary role in understanding melodic expectations.

Lerdahl (2001, p. 191) makes the last of these hypotheses explicit:

> However, gravity appears to be dispensable: in the major scale, except for the leading tone, the strongest virtual attractions of nonchordal diatonic pitches are by stepwise descent anyway. If there is any downward tendency beyond what is accounted for by attractions, it may reside in the fact that the most relaxed register for vocal production lies in a rather low range (though not at the bottom). That is, the cause may be more physical than cognitive. Besides, what is "down" to us may be "up" or "away" in another culture. The use of spatial metaphors is universal in talking about music, but spatial orientations are not. There is reason, then, to drop gravity as a musical force.

There is an interesting alternative statement of this last hypothesis. If, as Lerdahl suggests, the relative stabilities of each of the tones in the tonal system results in a tendency for less stable tones to descend, then we might ask whether the tonal system itself has evolved (here the term "teleologically" would be appropriate) so that its magnetic pulls imitate the effects of physical gravity. (One could, of course, answer this question in the affirmative and still find that gravity retains explanatory value as a separate musical force.) Note also that the use of Lerdahl's 1996 values for pitch space embedding instead of his original 1988 values increases the sense in which those stability values tend to imitate the effects of gravity (because of the resultant downward attractions).

The computer models and statistical analyses described in this article allow us to test all of these hypotheses.

The claim that musical forces shape melodic expectations is, of course, not new. Rothfarb (2001) describes the long history of the metaphors of motion and force in music discourse. Lerdahl (2001) also cites precedents to our work on magnetism. Inertia has been discussed as "good continuation" (Meyer, 1956, 1973), is related to "process" in the implication-realization model (Narmour, 1990, 1992), and appears central to the expectancy model of Jones (1981a, 1981b).[1]

OPERATIONS ON ALPHABETS

The summary statement above claims that musical forces "control operations on alphabets." Deutsch and Feroe (1981), drawing on the work of Simon and Sumner (1968), advance a model of music cognition that describes musical passages in terms of alphabets (e.g., the chromatic scale, the major scale, and specific chords) and operations (e.g., repetition, or motion to the next higher or lower member) that create motions through those alphabets.

Figure 1, taken from their article, shows a passage of music and a representation of how it might be encoded as the result of an operation

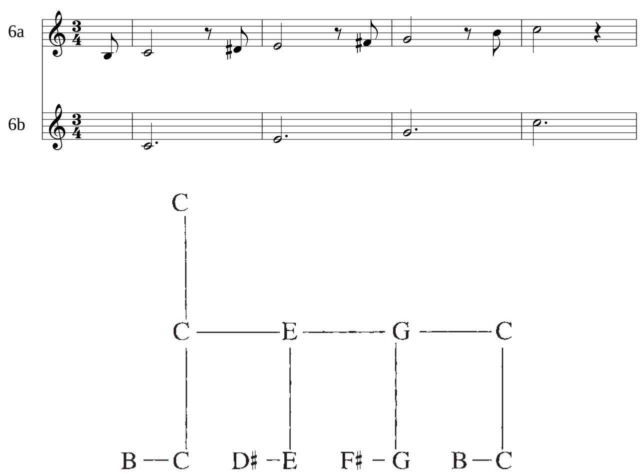

Fig. 1. Deutsch and Feroe (1981) describe passages in terms of alphabets and operations.

1. The claim that inertia controls melodic expectations may be regarded as a special case of

$$F = w_G G + w_M M + w_I I$$

in which weights for gravity (w_G) and magnetism (w_M) are equal to 0.

(involving successorship in the chromatic scale) applied to a major-triad arpeggio (itself the result of an operation on the alphabet of that chord).

Although a complete specification of the theory might describe how alphabets may be built up and internally represented, the present description simply assumes their internal representation and its computer implementation simply lists candidate alphabets and provides a simple set of rules for choosing appropriate alphabets in generating completions.

HIERARCHIES OF EMBELLISHMENT

The summary statement of the theory given above claims that these operations on alphabets create "hierarchies of embellishment." Deutsch and Feroe note that the operations on alphabets they describe create hierarchical structures like those described by Schenker (1935/1979)—what Bharucha (1984b) calls "event hierarchies." Although Schenker's claims about relationships between distant pitches have proven controversial, some more recent authors have recast his ideas as claims about the perceptions of skilled listeners (Deutsch & Feroe, 1981; Larson, 1997b; Lerdahl & Jackendoff, 1983; Westergaard, 1975). Others report experiments that support these claims about perception (Dibben, 1994; Marvin & Brinkman, 1999).

A Schenkerian analysis represents a hierarchy of embellishments (Larson, 1997b). Furthermore, the theory of musical forces claims that the way in which those embellishments are internally represented influences the way in which we expect melodies to be completed.

A complete specification of the theory might describe how hierarchical representations of embellishment structure may be built up and internally represented. However, the first computer model described here deals effectively only with short sequences of tones that may be thought of as representing a single level of such a hierarchy. Although the second computer model described here uses hierarchical descriptions to generate continuations, it does not create such hierarchical descriptions.

THE STEPWISE DISPLACEMENT OF AURALIZED TRACES

The general statement above claims that in these hierarchies of embellishment, "stepwise displacements of auralized traces create simple closed shapes." To auralize means to hear sounds internally that are not physically present. The term trace means the internal representation of a note that is still melodically active. In a melodic step (meaning a half step or a whole step), the second note tends to displace the trace of the first, leaving one trace in musical memory; in a melodic leap (meaning a minor third or larger), the second note tends to support the trace of the first, leaving two traces in musical memory.[2]

For the purposes of this article, I will not define "simple closed shapes." I believe that a complete understanding of melodic expectation requires that this term be defined and its ramifications explored. However, the points I will make in this article can rely on the shared intuitions of this journal's readers about the meaning of "simple closed shapes."

STEP COLLECTIONS

Because of the importance of the step-leap distinction, a special group of alphabets, called "step collections," seems to play a central role in music cognition (Larson, 1992; Hurwitz & Larson, 1994). A "step collection" is a group of notes that can be arranged in ascending pitch order to satisfy the following two conditions: (1) every adjacent pair of notes is a step (that is, a half step or a whole step) apart; and (2) no nonadjacent pair of notes is a step apart. The second condition can be modified slightly to produce a third condition, true of all "proper" step collections; (3) no two pitches—nor any of their octave equivalents—that are not adjacent in the list (except the first and last) are a step apart. The first condition ensures that the collection can be heard as a complete filling in of a musical space (this follows from recognizing that melodic leaps tend to leave the "trace" of a note "hanging" in our musical memories). The second condition ensures that no note will be heard as redundant in the filling of that space (this also reflects our desire to avoid confusion and the fact that either a whole step or half step can be heard as a step). The third condition grants a role to octave equivalence, ensuring that adding octave equivalents to a proper step collection can result in a proper step collection. All proper step collections satisfy the condition of maximal evenness and coherence as described in recent publications on scale theory (e.g., Balzano, 1980, 1982; Clough & Douthett, 1991; Clough & Myerson, 1985; Cohn, 1996; Gamer, 1967), the "non-adjacent half-step hypothesis" tested by Pressing (1978), and the "semitone constraint" described by Tymoczko (1997).

2. Bregman (1990) offers evidence concerning auditory streaming that supports this step/leap distinction. The "critical bandwidth" separates steps and leaps. Some theories of tonal music (see especially Bharucha, 1984a, on "melodic anchoring") grant important status to this distinction. For discussions of the step-leap distinction within the theory of musical forces, see Larson (1997b, 2002).

Gjerdingen (1994) has explored how neural-net models of aural perception may explain how we hear discrete pitches as forming a single melody or a compound melody. However, the questions "How do these notes break into different groups?" and "How do these notes displace the tensions represented by traces?," while related, are not identical questions.

Thus, step collections play a central role in the computer models described here. Most of the reference alphabets are step collections, and the programs encourage their selection as reference alphabets.

THE THEORY AS A SET OF INSTRUCTIONS

To give an idea of what a complete implementation of this theory might look like (and to suggest what a computer would have to do in order to model that theory), it may be restated as a set of instructions for producing a completion from a cue. Here is such a set of instructions:

1. Build up an internal representation of the cue (the "analysis") that includes the key, the mode, the meter, and a hierarchical representation of the embellishment functions and rhythmic attributes of each note or group of notes and the traces they leave. Evaluate the quality of that analysis in terms of its simplicity and order—a kind of confidence rating.
2. For each appropriate level of that hierarchy, determine the alphabet within which motion might continue (the "reference alphabet") and allow more basic levels of structure to determine the alphabet of pitches that will serve as potential goals of that motion (the "goal alphabet").
3. List inertia predictions by continuing successorship motion within the reference alphabet until a member of the goal alphabet is reached. Preserve musical patterns in inertia predictions by applying the same embellishment structures at analogous levels of structure. Consider alternative descriptions of structure wherever these may facilitate the creation of analogous structures.
4. For gravity predictions, allow pitches that are described as "above stable reference points" to descend within their reference alphabet until that reference point or a member of the goal alphabet is reached.
5. For magnetism predictions, move through the reference alphabet to the closest member of the goal alphabet.
6. Evaluate potential completions according to the degree that they give in to the musical forces (their "rating") and enter them into a lottery in which their chance of being chosen is a reflection of their rating, the confidence rating of the current analysis, and other factors affecting the urgency of choosing a completion.

Although I have numbered these instructions, we should consider them as taking place in parallel, and influencing one another, until a potential completion is chosen.

This set of instructions resembles computer models of analogous cognitive tasks (pattern finding, sequence extrapolation, and analogy making) created at the Center for Research on Concepts and Cognition (Hofstadter et al., 1995). The success and sophistication of those programs suggests that a complete implementation of this set of instructions, though a very complex task, is possible and will most likely be quite informative about music cognition.

However, for our present purposes, two much simpler computer models will suffice. They model limited but important aspects of this instruction set, and their success in producing entire completions provides strong support for the theory. This article will not explicate every detail of the complete instruction set just given, but it will explain all the assumptions that are implemented in the computer programs described next.

A Single-Level Computer Model

Some aspects of this theory have been implemented in what I will call a "single-level" computer model (elsewhere, I have called this model "Next Generation"). The model is a simple one—every aspect of the model is described in the following pages—no additional assumptions or mechanisms are "hidden" in the code. This model, when given a cue in a specified key, returns a rated list of completions. For example, if we ask it to assume the key of C and give it the beginning G–A, it predicts that roughly half of participants will respond with G–A–G (giving in to gravity and magnetism), that roughly half will respond with G–A–B–C (giving in to inertia), and that none will respond with anything else. To calculate the ratings that it gives to each completion it generates, it uses the algorithm given above (Larson, 1993a), with a separate factor added for the stability of the final note.

To make its predictions, the single-level model first chooses a pair (or pairs) of reference and goal alphabets from the list in Figure 2. To do so, it follows three simple rules: (1) diatonic cues may not use chromatic alphabets, (2) reference alphabets must include the last pitch of the cue, and (3) goal alphabets must not include the last pitch of the cue (an exception is noted below for unrated inertia predictions). Combinations other than those listed in Figure 2 are possible, and the responses of some experimental participants suggest expectations that can be described in terms of other combinations, but this simple set is well-defined and accounts well for the experimental results we will examine. (One simplification is that for every reference alphabet in Figure 2, there is only one goal alphabet, and that every goal alphabet in Figure 2 is the largest of the alphabets in Figure 2 that is a proper subset of its reference alphabet.)

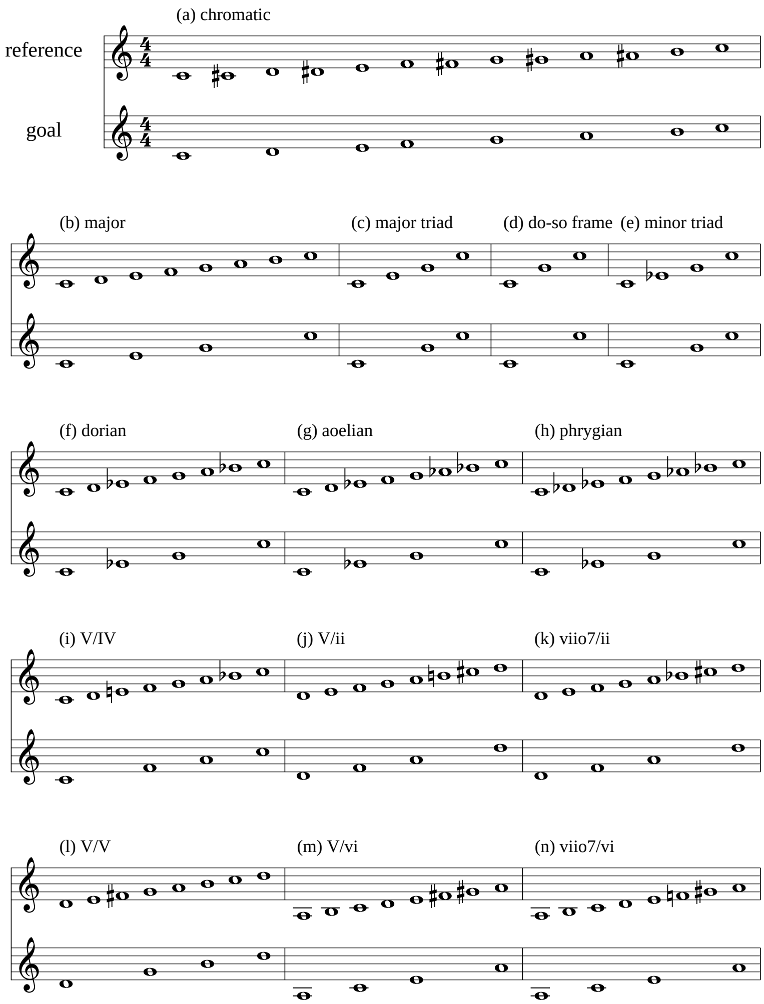

Fig. 2. Reference and goal combinations in the single-level model.

To illustrate the choice of reference and goal alphabets, consider the case of the melodic beginning G–A–?. According to the first rule, none of the chromatic alphabets (not choices a, nor e–n, in Figure 2) may be chosen as reference alphabets (because the cue contains no chromaticism); this leaves the combinations in Figure 2b–2d. According to the second rule, the triad and the frame (choices c and d) are eliminated as reference alphabets (they do not contain A, the last note of the cue); together with the first rule, this leaves only the combination shown in Figure 2b. According to the third rule, all of the combinations whose goal alphabets include A (choices a, i–k, and m–n in Figure 2) are eliminated (because this would suggest that the goal is already reached, hence no notes need be added); this still allows the combination shown in Figure 2b.

Once the combination of reference and goal alphabets is chosen, the single-level model makes its predictions by moving within the reference alphabet until it arrives at a member of the goal alphabet. The direction of motion is determined by the musical forces.

Figure 3 illustrates the resultant magnetism prediction for the beginning G–A–?. In Step 1, the reference and goal alphabets are chosen. In Step 2, the distances to the closest stable pitches, both up and down, are calculated (the "stable pitches" are simply the notes contained in the goal alphabet). For G–A–?, the closest stable pitches are G (2 semitones from A) and C (3 semitones from A). In Step 3, the computer chooses motion to G (because it is closer than C) and moves through the reference alphabet to that G. The result is that the computer has turned the cue G–A–? into the completion G–A–G.

Step 1: The goal and reference alphabets are chosen.

reference alphabet
(scale) ----->

C D E F G A B C

goal alphabet
(chord)

C E G C

Step 2: The distances to the closest stable pitches are calculated (in half steps).

reference alphabet
(scale) <--2- ---3--->

C D E F G A B C

goal alphabet
(chord)

C E G C

Step 3: The prediction is for motion (through the reference alphabet) to the closest stable pitch (G).

reference alphabet
(scale) <-----

C D E F G A B C

goal alphabet
(chord)

C E G C

Resultant prediction: G–A–G.

Fig. 3. A magnetism prediction for G–A–?.

Figure 4 illustrates the resultant inertia prediction for the same beginning. In Step 1, the reference and goal alphabets are chosen. Because the last two notes of the cue are adjacent in the reference alphabet, there is a pattern of motion (ascending through the reference alphabet) that can be continued. In Step 2, that pattern of motion is continued. G–A–?, has now become G–A–B–?. Because the B is unstable (i.e., it is not contained in the goal alphabet), we go on to Step 3, where the motion is continued further to C. Because the C is stable (i.e., it is contained in the goal alphabet), the computer stops. The result is that the computer has turned the cue G–A–? into the completion G–A–B–C.

Because the goal alphabet is always a subset of the reference alphabet, the computer always knows when to stop. Although Narmour's implica-

Step 1: The goal and reference alphabets are chosen.

```
reference alphabet
(scale)                 ----->
C     D    E   F     G      A      B  C

goal alphabet
(chord)
C          E         G                C
```

Step 2: Motion is continued in the same direction within the reference level, first to B (which is unstable).

```
reference alphabet
(scale)                        ----->
C     D    E   F     G      A      B  C

goal alphabet
(chord)
C          E         G                C
```

Step 3: Since B is also unstable, motion is again continued in the same direction within the reference alphabet, now to C (which is stable).

```
reference alphabet
(scale)                               -->
C     D    E   F     G      A      B  C

goal alphabet
(chord)
C          E         G                C
```

Resultant prediction: G–A–B–C.

Fig. 4. An inertia prediction for G–A–?.

tion-realization model does not clearly specify how long a continuation might go on, the computer model always does.

In the single-level model, inertia is represented only when the last two notes of the cue are adjacent in the reference alphabet (so that the pattern is "motion by adjacency within the reference alphabet" and inertia can thus continue "in the same fashion").

The idea of traces is implemented in one feature of the program. In situations where the unstable second-to-last note of a cue leaps to a more stable pitch, the computer assumes that the leap has left the unstable note "hanging." It thus generates a continuation from that note rather than from the second note.

The contextual quality of musical gravity is implemented in another feature of the program. If a melody descends only a half step below a stable base (imitating the physical motion that we make when we crouch)—so that we experience no natural lower position for gravity to take us to—the computer will not attempt to give in to gravity.

As reported elsewhere (Larson, 1994b), disabling these features weakens the performance of the computer model; enabling them makes the program agree better with participants' responses.

Some inertia predictions are not assigned ratings by the algorithm. For the beginning F–E–?, the single-level model uses the tonic triad as reference alphabet and the tonic-dominant frame as goal alphabet (as noted earlier, goal alphabets must not include the last pitch of the cue, and the goal alphabet must be the largest one contained in the reference alphabet). The result is the continuations F–E–G and F–E–C, which give in to magnetism and gravity, respectively. It thus rates these two continuations accordingly (computing the differences in magnetic pulls from E to G above and E to C below). However—like physical inertia—musical inertia does not depend on stability. Thus, for the beginning F–E–?, if it is heard as a stepwise descent through the reference alphabet of the major scale, inertia predicts the continuation F–E–D–C. The single-level model will produce such inertial completions, but it is not clear what distances should be used in the algorithm for computing their magnetic pulls; thus the comparison of ratings for these continuations could seem arbitrary. In what follows, such inertial completions are suppressed whenever ratings are compared.

The single-level model can also make "recursive" predictions. For the melodic beginning, E–F–?, it predicts E–F–G and E–F–E. It can then take one of these predictions, say E–F–E, and then reinterpret it as the new beginning E–F–E–?. It thus adds the predictions E–F–E–C, E–F–E–G, and E–F–E–D–C. Again, because more than one rating could be assigned, comparing ratings seems arbitrary. Such recursive predictions are also suppressed when ratings are compared.

Two additional examples, shown in Table 1, illustrate the operation of the single-level model.

Consider the cue C–C♯–?. Because the cue ends with a C♯, the reference alphabet must include that note. (Of course, the same note may also be spelled D♭—the computer, like the human listener, may hear that note either way. Here, I will spell such notes according to music-theoretical conventions, that is, contextually.) This allows the reference and goal combinations given in Figures 2a, h, j, and k. For the combination in Figure 2a, the gravity prediction is C–D♭–C (going down in the reference alphabet until we hit a member of the goal alphabet). For the combination in Figure 2a, there are either two magnetism predictions or no magnetism predictions (the closest stable pitches, C and D, are equidistant from C♯). For the combinations in Figures 2a and h, the last two notes of the cue are adjacent in their alphabets, so they lead to inertial predictions (of C–C♯–D and C–D♭–E♭, respectively). The reference and goal combinations in Figures 2j and k do not allow an inertial prediction. For the combination in Figure 2h, the gravity and the magnetism predictions are the same: C–D♭–C (because one hits C by going down with gravity or by going to the closest stable pitch). For the combinations in Figures 2j and k, the magnetism predictions are C–C♯–D. For the reasons mentioned earlier, the combinations in Figures 2j and k do not produce gravity predictions (because these would take us below the "ground" of C).

Consider the cue D–G–? As noted earlier, when the second-to-last note of the cue leaps to a more-stable pitch, the single-level model computes the continuation from that note. The only available combination in Figure 2 that includes D in its reference alphabet but not in its goal alphabet is Figure 2b. Because the cue is a single note, there is no inertia prediction. Because the closest stable pitches, E and C, are equidistant from D, there

Table 1

Two Additional Examples Illustrate the Single-Level Model

Cue	**Reference and Goal Combination (see Figure 2)**	**Gravity**	**Magnetism**	**Inertia**
C–C♯–?	(a) Chromatic	C–D♭–C	Both C–D♭–C and C–C♯–D, or neither	C–C♯–D
	(h) Phrygian	C–D♭–C	C–D♭–C	C–D♭–E♭
	(j) V/ii	NA	C–C♯–D	NA
	(k) viio7/ii	NA	C–C♯–D	NA
D–G–?	(b) major	D–G–C	Both D–G–C and D–G–E, or neither	NA

are either two magnetism predictions (D–G–C and D–G–E) or no magnetism prediction. The gravity prediction is D–G–C.

A Multilevel Computer Model

Further aspects of this theory have been implemented in what I will call a “multilevel” computer model (elsewhere, I have called this model “Voyager”). When given a melodic beginning and a Schenkerian analysis of that beginning, this second model returns a list of possible completions. Although the single-level model deals better with short beginnings than with longer ones, the multilevel model appears to respond effectively to melodic beginnings of any length.

To make its predictions, the multilevel model takes the top (most basic) level of the analysis it has been given and calls on the single-level model to produce a completion at that level. It then fills in that “skeleton” by choosing notes (at levels closer to the “surface” of the melodic beginning) that give in to inertia.

Figures 5 and 6 illustrate the operation of the multilevel model. Given the melodic beginning E–D–C–D in the key of C (Figure 5a), the single-level model can give in to gravity by moving down to C (Figure 5b) or it can give in to inertia by continuing to move stepwise up the C major scale to E (Figure 5c). However, if one hears this melodic beginning in triple meter (as suggested by the notation of Figure 5d), then the continuation in Figure 5e seems natural.

This is the continuation that the multilevel model produces when given the analysis shown in Figures 6a–c. Each level of such an analysis shows how its structural notes (stemmed) are embellished with affixes

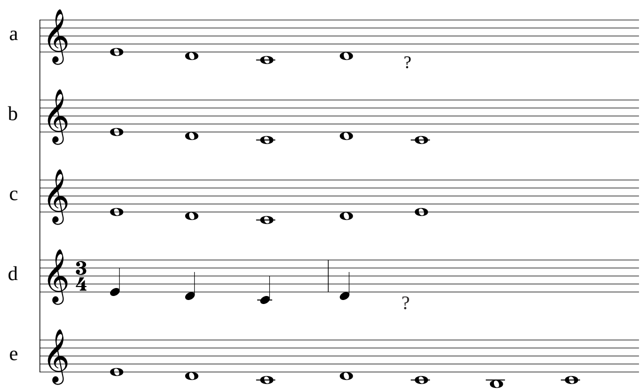

Fig. 5. Completions for E–D–C–D–?.

(unstemmed notes connected by slur to the note they embellish) or with connectives (unstemmed notes contained within slurs between the notes they connect). Notes that are unstemmed on one level do not appear on the next, more-abstract level. Figure 6c gives the melodic beginning already given in Figure 5a, but it also identifies the D as a connective (a passing tone). Figure 6b indicates that the C embellishes the E (as a suffix lower third). Figures 6d–f show how the model turns this analysis into a prediction. First, it adds to Figure 6a a note that continues the pacing of that level—and (by calling on the single-level model) it chooses C (gravity and inertia both point to C). This C is shown at the end of Figure 6d, and its presence there requires its appearance in the same location in Figures 6e and f. Second, it tries to give in to inertia by finding a suffix lower third for the D at the end of Figure 6b to answer the C before it—and the only candidate is the B shown in Figure 6e. Third, it tries to give in to inertia at the lower level too by finding a passing tone to answer the first D in Figure 6c—and the only candidate is the second C shown in Figure 6f. The result, Figure 6g, is the one we already listed in Figure 5e.

Further examples show that different analyses of the same beginning may lead to different completions. This reflects our intuition (and the results of some experiments, e.g., Kidd, 1984; Larson, 1997a) that hearing a passage in a different key or different meter will lead us to expect a different completion.

Figure 7a gives another melodic beginning. If given this beginning in the key of C, the single-level model produces the inertial and recursive

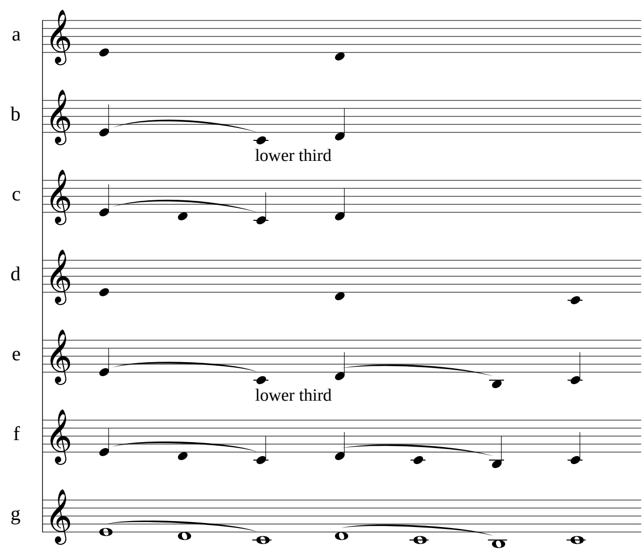

Fig. 6. The multilevel model's prediction for E–D–C–D–? depends on analysis.

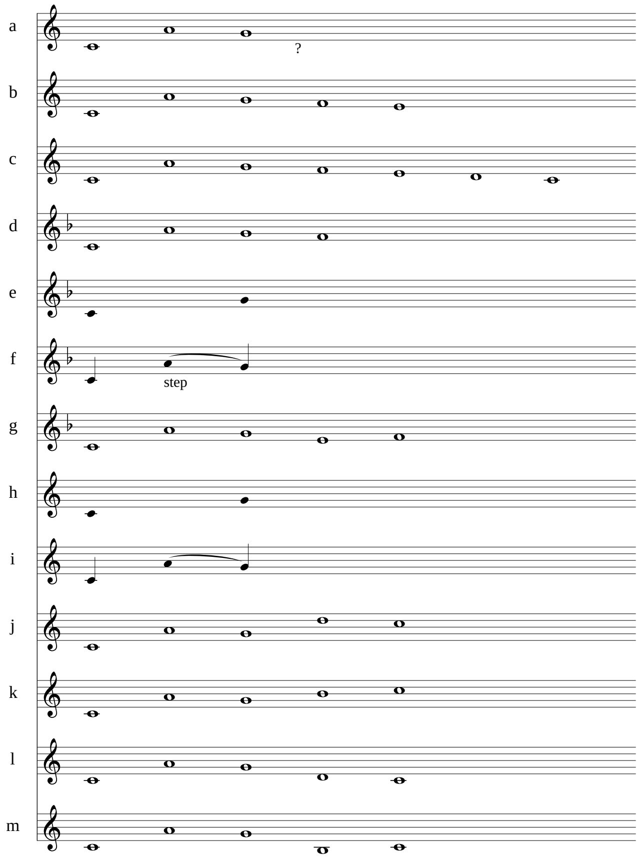

Fig. 7. The multilevel model's predictions for C–A–G–?.

predictions shown in Figures 7b and c. If represented in the key of F, it produces the prediction shown in Figure 7d. However, if the multilevel model is given the analysis in Figures 7e and f (i.e., if it is told that this beginning is in duple meter without a pickup note), then it produces the result given in Figure 7g. This is because, at the top level, gravity (neither magnetism nor inertia apply here) leads the G to F, and because, at the next level, inertia answers the prefix step to G with a prefix step to F (of course, that prefix step could be G or E, but the model also avoids adding repeated notes to melodic beginnings that do not have repeated notes in them, so it chooses the E). If the multilevel model is given the analysis shown in Figures 7h and i (same meter now, but a different key) it will produce the results given in Figures 7j–l. This is because the last note can be an upper C, giving in to inertia and magnetism, or a lower C, giving in to gravity and because, when asked to find a step, it will seek a lower or an upper step (if the analysis specifies that the A is an upper step, then the multilevel model will produce only the completions shown in Figures 7j and l).

The same melodic beginning was used in a recent experiment (Larson, 1997a). Participants were given the melodic beginning C–A–G–?, but no clues about meter or key. The responses they sang most often are given in Figure 8. Notice that this list is almost identical to the list of computer-generated responses in Figure 7.

Experimental Evidence

The output of these computer programs also agrees with the results of a number of other experiments in ways that suggest that these programs may capture some important aspects of the expectations of experienced listeners of tonal music.

In what follows, I compare the behavior of these computer models with the behavior of participants in experiments that involve responses to one-note cues (Povel, 1996) and responses to two- and three-note cues (Lake, 1987; Larson, 1996, 1997a).

Although the multilevel model is a more complete implementation of the theory, it introduces complexities that prohibit the kind of straightforward quantification that allows the single-level model to give ratings to its completions. The following comparisons respect these differences between the models.

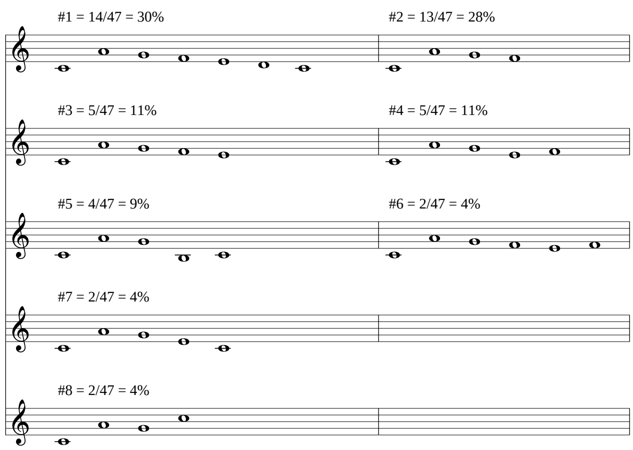

Fig. 8. Subject responses to C–A–G–? (Larson, 1997a).

SINGLE-NOTE CUES WITH SINGLE-NOTE RESPONSES (POVEL, 1996)

Povel (1996) established a major-key context, gave a single-note cue, and then asked participants to produce one added note by finding the expected note on the synthesizer. Povel separated his participants into three categories: the "tonal group" (whose responses to nontonic cues are shown in Table 2); the "fourth group" (whose responses tended to lie a fourth or fifth from the cue, regardless of its tonal context); and the "random group" (whose responses did not show a stable strategy).

In Table 2, the column on the left gives the cue, and the row at the top lists the response given by Povel's participants. Each cell gives the proportion of participants in Povel's "tonal group" that responded with the continuation listed. Povel omits values below 5% (thus, the sum of each row is 100%, minus the percentage of responses to any cues that were below 5%). This table also omits the tonic as cue (the computer responds to a single tonic with its equivalent of "Sounds fine by itself, why should I add anything?"). The table does not distinguish between ascending and descending melodies.

Single-tone cues tend to elicit single-note responses from the single-level model, too. Table 3 shows all 20 of the computer model's single-note responses to the 11 nontonic single-note cues. The computer model generates only three longer responses to these one-note cues (for one of the cues, the computer model generates one completion of more than one note; and for one other cue, it generates two completions of more than one note).

The striking agreement between the single-level model's predictions and Povel's participants' responses is summarized in Table 4. Out of 121 possible single-note responses, the single-level model produces only 20. All

TABLE 2

The "Tonal Group" in Povel's (1996) Experiment, Nontonic Cues, With C As Tonic

	To											
From	**C**	**C♯/D♭**	**D**	**D♯/E♭**	**E**	**F**	**F♯/G♭**	**G**	**G♯/A♭**	**A**	**A♯/B♭**	**B**
C♯/D♭	.69	.06	.07	0	0	0	0	0	0	.05	0	0
D	.48	0	0	0	.40	0	0	0	0	0	0	0
D♯/E♭	.51	0	0	0	.32	0	0	0	0	0	0	0
E	.81	0	0	0	0	0	0	.11	0	0	0	0
F	.15	0	0	0	.65	.09	0	0	0	0	0	0
F♯/G♭	.11	0	0	0	.07	0	0	.81	0	0	0	0
G	.76	0	0	0	.13	.05	0	0	0	0	0	0
G♯/A♭	.09	0	0	0	.07	0	0	.61	.05	.08	0	0
A	.28	0	.07	0	0	.07	0	.52	0	0	0	0
A♯/B♭	.24	0	0	0	0	0	0	.13	0	.39	0	.07
B	.95	0	0	0	0	0	0	0	0	0	0	0

TABLE 3
The Single-Level Model's Single-Note Responses to Single-Note Cues

	To											
From	**C**	**C♯/D♭**	**D**	**D♯/E♭**	**E**	**F**	**F♯/G♭**	**G**	**G♯/A♭**	**A**	**A♯/B♭**	**B**
C♯/D♭	1	0	1	0	0	0	0	0	0	0	0	0
D	1	0	0	0	1	0	0	0	0	0	0	0
D♯/E♭	1	0	1	0	1	0	0	0	0	0	0	0
E	1	0	0	0	0	0	0	1	0	0	0	0
F	0	0	0	0	1	0	0	0	0	0	0	0
F♯/G♭	0	0	0	0	0	1	0	1	0	0	0	0
G	1	0	0	0	0	0	0	0	0	0	0	0
G♯/A♭	0	0	0	0	0	0	0	1	0	1	0	0
A	0	0	0	0	0	0	0	1	0	0	0	0
A♯/B♭	1	0	0	0	0	0	0	0	0	1	0	1
B	1	0	0	0	0	0	0	0	0	0	0	0

TABLE 4
Agreements Between Povel's Subjects' One-Note Responses to All One-Note Cues and the Single-Level Model's One-Note Predictions for the Same Cues, With C As Tonic

	Number of Those Responses Produced by the Single-Level Model	**Number of Those Responses Not Produced by the Single-Level Model**	**Total**
Responses produced by ≥5% of Povel's participants	**18**	15	33
Responses produced by <5% of Povel's participants	2	**86**	88
Total	20	101	**121**

NOTE—$df = 1$, $\chi^2 = 47.53$, $p < .0001$.

but 2 of these 20 single-note responses are included in Povel's "tonal group." Povel's "tonal group" includes 15 responses not produced by the computer model, but many of those responses ended by simply adding a tonic (C) or by repeating the cue—and the remainder of those not produced by the computer model were among those sung by the smallest number of participants.

Lerdahl uses his algorithm to describe the attraction of any pitch to any other pitch (regardless of whether or not the pitches are adjacent in a relevant alphabet, and regardless of which pitch is more stable). Table 5 gives the values generated by Lerdahl's formula for "melodic attraction."

The values in Table 5 rely on Lerdahl's conjecture that "It may turn out to be more accurate just to take attractional values as correlating directly with degrees of expected continuation" (Lerdahl, 2001, p. 170). In other words, the values in Table 5 are calculated by multiplying the ratio of the stabilities of both notes by the inverse square of the semitone distance between them; opposing attractors are ignored. (Ignoring opposing attractors has an additional practical advantage here. It is not always clear for

TABLE 5

Lerdahl (2001) Uses His Melodic Attraction Algorithm to Describe the Magnetic Pull of Any Pitch to Any Other Pitch

	To											
From	**C**	**C♯/D♭**	**D**	**D♯/E♭**	**E**	**F**	**F♯/G♭**	**G**	**G♯/A♭**	**A**	**A♯/B♭**	**B**
C♯/D♭	4.00		2.00	0.25	0.33	0.13	0.04	0.08	0.04	0.13	0.11	0.50
D	0.50	0.50		0.50	0.38	0.11	0.03	0.06	0.01	0.04	0.03	0.11
D♯/E♭	0.44	0.25	2.00		3.00	0.50	0.11	0.19	0.04	0.06	0.04	0.13
E	0.08	0.04	0.17	0.33		0.67	0.08	0.11	0.02	0.03	0.01	0.03
F	0.08	0.03	0.11	0.13	1.50		0.50	0.38	0.06	0.06	0.02	0.03
F♯/G♭	0.11	0.04	0.13	0.11	0.75	2.00		3.00	0.25	0.22	0.06	0.08
G	0.05	0.01	0.03	0.02	0.11	0.17	0.33		0.33	0.17	0.04	0.04
G♯/A♭	0.25	0.04	0.06	0.04	0.19	0.22	0.25	3.00		2.00	0.25	0.22
A	0.22	0.03	0.04	0.01	0.06	0.06	0.06	0.38	0.50		0.50	0.25
A♯/B♭	1.00	0.11	0.13	0.04	0.08	0.08	0.06	0.33	0.25	2.00		2.00
B	2.00	0.13	0.11	0.03	0.06	0.03	0.02	0.09	0.06	0.25	0.50	

NOTE—Regardless of whether or not the pitches are adjacent in a relevant alphabet, and regardless of which pitch is more stable.

every cell in Table 5 which note should be considered to be the opposing attractor.) The diagonal is left blank because the algorithm does not produce values for the attraction of a note to itself. The remaining values correlate positively with those given in Table 2 ($r = .569$, $N = 121$, $p < .001$).

The multiple-regression analysis reported in Table 6 assesses the contributions of each of the factors in Lerdahl's 1996 algorithm. The three factors that go into Lerdahl's algorithm are the inverse stability of the cue, the stability of the response, and the distance between the cue and the response. Table 6 lists these factors in three different models. The first and third models use Lerdahl's 1988 values for stability. The second

TABLE 6

Summary of Multiple Regression Analysis for Factors In Lerdahl's (2001) Melodic Attraction Algorithm Predicting Distribution of Responses in Povel's "Tonal Group"

	Using 1988 Values for Stability and Inverse Square for Distance	**Using 1996 Values for Stability and Inverse Square for Distance**	**Using 1988 Values for Stability and Simple Inverse for Distance**
Summary of multiple regression	$R = .732$ $p < .001$ $N = 121$	$R = .706$ $p < .001$ $N = 121$	$R = .730$ $p < .001$ $N = 121$
Stability of response	$w = .659$ $p < .001$	$w = .631$ $p < .001$	$w = .658$ $p < .001$
Inverse stability of cue	$w = -.052$ $p = .415$	$w = -.054$ $p = .416$	$w = -.052$ $p = .416$
Inverse square of distance between cue and response	$w = .311$ $p < .001$	$w = .312$ $p < .001$	$w = .308$ $p < .001$

NOTE—Weights given as w are standardized coefficients.

model uses his 1996 values. The first and second models use the inverse square of the semitones between the cue and the response for that distance (as in Larson, 1993a and Lerdahl, 1996). The third model substitutes the inverse (instead of the inverse square) of the distance (as in Bharucha, 1996).

All three models show that two factors contribute significantly to account for these data: the stability of the response and the distance between cue and response. However, at least for Povel's data, the inverse stability of the cue does not help explain the result. Nor does the use of Lerdahl's 1996 values for stability improve the ability of the model to account for the data. Furthermore, substituting the simple inverse of the distance instead of the inverse square does not give a better result.

However, the theory proposed here goes beyond these algorithms for the interaction of musical forces. It also claims that melodic completions are created by operations on alphabets. Table 3, which listed the single-level model's force-driven operations on alphabets, makes a large number of "correct rejections" (the 86 cue-and-response combinations for which Tables 2 and 3 both give the value 0). This helps to explain the strong correlation between the values given in Tables 2 and 3 ($r = .773$, $N = 121$, $p < .001$).

The multiple-regression analysis reported in Table 7 further assesses the contributions of individual factors. The factors listed include musical gravity (scored 1 if the response is a gravity prediction, 0 if it is not), musical magnetism (scored 1 if the response is a magnetism prediction, 0 if it is not), and the stability of the response (using Lerdahl's 1988, rather than his 1996, values for depth of pitch-space embedding).

All the factors in Table 7—gravity, magnetism, and response-tone stability—contribute significantly to account for the data.

It would be interesting to test Narmour's implication-realization model on these data, too, but Narmour's model requires an implicative interval (two notes) and Povel's cues are only one note. Thus Narmour's model cannot be tested on Povel's data.

TABLE 7

Summary of Multiple Regression Analysis for Gravity and Magnetism in Predicting Distribution of Responses in Povel's "Tonal Group"

Summary of multiple regression	$R = .859$ $p < .001$ $N = 121$
Stability of response	$w = .327$ $p < .001$
Gravity	$w = .322$ $p < .001$
Magnetism	$w = .425$ $p < .001$

NOTE—Weights given as *w* are standardized coefficients.

To summarize, the single-level computer model comes close to predicting all and only the responses of Povel's participants. Furthermore, an algorithm that combines response-tone stability, gravity, and magnetism (inertia does not arise in these single-note cues) gives higher ratings to patterns that participants sing more often. The fact that these data can be modeled well by an interaction of constantly acting but contextually determined musical forces gives support to the idea that we experience musical motions metaphorically in terms of our experience of physical motions.

TWO-NOTE CUES WITH RESPONSES OF VARIED LENGTH (LAKE, 1987)

In his dissertation study, Lake (1987) asked music students at the University of Michigan to sing melodic continuations of two-note cues. First he established a major-key context by playing a tonic major chord and scale for them. He then played a two-note cue. Finally, he asked the students to sing that two-note beginning, "adding another tone or tones of your own choosing" (Lake, 1987, p. 31).

Lake's cues include all two-note combinations in which the first tone is diatonic to the key. Excluding cues that end on the tonic would give us 70 cues (7 diatonic first notes times the 10 chromatic pitches that exclude the tonic and the first note equals 70 two-note cues). However, perhaps because of the tone used for the cues (piano doubled in four octaves), for five of the cues, some the participants interpreted the cues as going up while the others interpreted them as going down. The result is a list of 75 different cues. Stating response frequency as a percentage of total responses to a given cue allows appropriate comparisons. Because each participant responded twice to each cue, 19 participants produced 2660 responses.[3]

To explore the shared intuitions of these participants, we consider only those responses that match note-for-note more than one other response to the same cue. The result is 1084 responses from 19 participants (41% of all the responses). These 1084 responses include 171 different responses, so each was sung an average of 6.34 times. This level of note-for-note agreement between participants suggests that these data are a reliable indication of shared musical intuitions about melodic expectations.

Table 8 summarizes the striking agreement between the single-level computer model's predictions and these data. The computer correctly predicts 129 (75%) of these 171 different responses, accounting for 934 (86%) of 1084 total responses. Furthermore, the responses predicted by

3. Lake began with 22 participants. Two were unable to sing back the cues accurately in their responses, so were excluded. Lake (private communication) urges the elimination of a third participant whose responses seem not to reflect the task. The method used here would eliminate most of this third participant's responses anyway, because they rarely match more than one other response note-for-note. This leaves 19 participants.

TABLE 8
The Single-Level Model Predictions Compared to Lake's Participants' Responses

	Number of Responses	Number of Different Responses	Average Number of Times Sung
All			
Participants' responses that exactly match more than one other response	1084	171	6.34
Hits			
Responses produced by the single-level model (with unrated inertia predictions and recursive predictions suppressed)	789	98	8.05
Additional responses produced as unrated inertia predictions or as recursive predictions	145	31	4.68
Total hits of these types	934	129	7.24
Misses			
Responses not produced by the single-level model	150	42	3.57
Responses produced by the single-level model but not produced by Lake's participants	45	45	0
Total misses	195	87	

NOTE—Including only those responses that exactly match more than one other response.

the computer are popular ones (they were sung an average of 7.24 times by these participants). The responses not predicted by the computer are less popular (they were sung an average of 3.57 times; and since we consider only those responses that match more than one other response, this is barely above the theoretical minimum of 3 times).

To further appreciate the level of agreement involved, it helps to describe "correct rejections." There are responses that are produced neither (a) by experimental participants nor (b) by the single-level model. We may regard these as responses that the computer model correctly predicts no one will sing. They are, however, responses to which both Lerdahl's and Narmour's theories give positive ratings.

Lake's participants added one note (roughly three fifths of the time), two notes (roughly one fifth of the time), or three notes (roughly one fifth of the time). Their responses usually (but not always) stay within a fifth of the last note of their beginning. For these 75 cues, there are 205,800 unique responses that meet these criteria. Even if we limit ourselves to diatonic responses within a fifth above and below the last note of a beginning (limits that Lake's participants did not adhere to), we still have 38,400 unique responses. These results are summarized in Table 9.

Furthermore, a closer look at some of the "misses" shown in Table 8 suggests that the theory proposed here also accounts for several of those responses, too.[4] For example, some of the responses may be regarded as

4. Readers who wish to examine the data themselves in more detail may get those data by e-mailing me at steve@uoregon.edu.

TABLE 9

Exact Matches Between Lake's Participants' Responses and Those of the Single-Level Model for All Possible Five-Note Responses

	Produced by the Single-Level Model	Adding Selected Recursive and Unrated Inertia Predictions	Not Produced by the Single-Level Model	Total
Sung by Lake's subjects	**98**	**129**	42	171
Not sung by Lake's subjects	45	45	**205,584**	205,629
Total	143	174	205,626	**205,800**

NOTE—Within the fifth above or octave below the final note of each beginning.

inertia predictions within alphabets not included as reference alphabets in the single-level model: for the descending cue C–A–?, the response C–A–F would be the inertial prediction if it included the subdominant chord as a reference alphabet; and for the ascending cue D–F♯–?, the response D–F♯–A would be the inertial prediction if it included V/V as a reference alphabet.

Other responses show the need for hierarchical descriptions of embellishment structure—and the multilevel model predicts their completion well. Consider the cues E–C♯–? and C–A♭—? and the respective "misses" E–C♯–D–B–C and C–A♭–G–B–C. If the cue E–C♯–? is given to the single-level model, one result is E–C♯–D. If that C♯ is then represented as a suffix lower third, then the multilevel model will produce the continuation E–C♯–D–B–C. If the cue C–A♭—? is given to the single-level model, the result is C–A♭–G. If that A♭ is then represented as a prefix half-step, then the multilevel model will produce the continuation C–A♭–G–B–C.

AN ARTIFICIAL-INTELLIGENCE APPRAISAL

Within the field of artificial intelligence, it is often suggested that the best test of a computer model is to enter it in a kind of contest where it competes against human participants. The "Turing Test" (which is supposed to test machine "intelligence") and computer chess tournaments (which are supposed to assess machine skill at chess) may be the best-known examples of such contests. In the Turing Test, a human judge at one terminal sends questions to both a human at another terminal and to a computer. Responses from both appear on the judge's screen. The object is for the human judge to determine which responses come from the computer program and which come from another human. If the human judge cannot make that determination, then the computer is deemed intelligent. No computer has passed the Turing Test yet. However, the computer chess program "Deep Blue" recently beat top-rated Garry Kasparov in a chess tournament.

By analogy, we ask what would happen if the single-level computer model were one of the participants in Lake's experiment—and we regard that experiment as a contest. As with the Turing Test, we regard the computer as successful the more it agrees with the human participants. Thus each participant (the computer program included) gets a score based on how often its responses agree—note for note—with those of other participants. The winner is the participant that gets the highest score.

To do this, we allow the rating that the single-level computer model assigns to each prediction to determine the frequency with which it produces that response. We then let it produce two responses to each cue (as did Lake's participants). Finally, we can ask how often it agrees note-for-note with Lake's other participants. The answer is that the single-level model can produce a set of responses that creates as many as 710 note-for-note matches with Lake's participants. When it does, the average number of such note-for-note matches for each of Lake's participants is 294 and the highest is 619. In other words, the computer model can match the experimental participants note-for-note as often as any one of the participants does.[5] The results are shown in Figure 9.

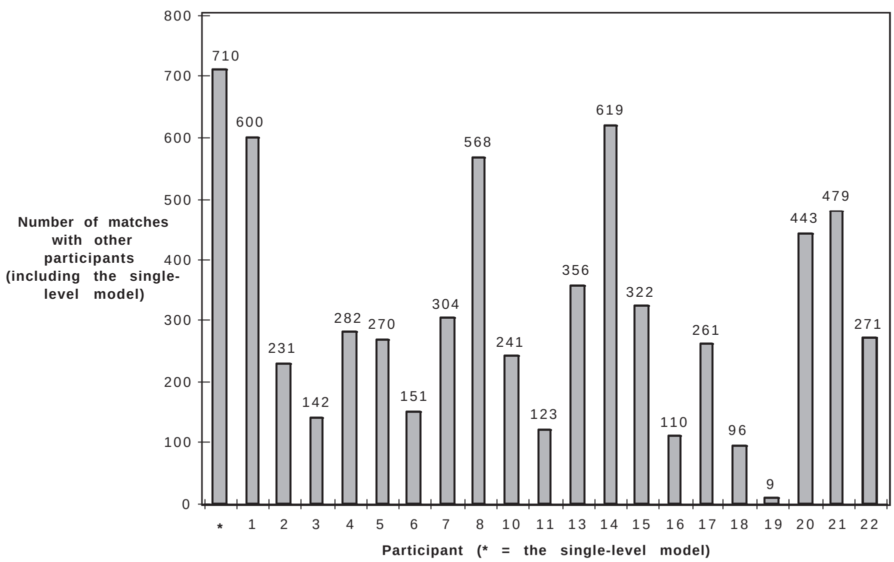

Fig. 9. The single-level model can produce a set of responses that exactly matches the entire responses of Lake's (1987) participants as often as do the responses of any one of those participants.

5. In order to choose among responses with probabilities that depend on ratings, the single-level model must call on a random number generator. The result is that each time the program is run, different choices may be made. In this sense, its behavior resembles that of the experimental participants, who sometimes give different responses to the same cues. I have run this "contest" many times, and, while the computer scored differently on different runs, it always came in first or second place.

To summarize, the single-level computer model predicts most of the responses of Lake's participants and predicts few responses not produced by those participants. Furthermore, an algorithm that combines gravity, magnetism, inertia, and stability gives higher ratings to patterns that participants sing more often. The fact that this model can exactly match the responses of participants in Lake's study as often as do those of any of the participants in that study gives strong support to the idea that we experience musical motions metaphorically in terms of our experience of physical motions.

USING LAKE'S (1987) DATA TO COMPARE THE IMPLICATION-REALIZATION MODEL WITH THE THEORY OF MUSICAL FORCES

In a recent article, Krumhansl (1995) compared the predictions of the bottom-up component of Narmour's (1990) implication-realization model with the judgments given to selected melodic continuations by her experimental participants. Her approach allows us to compare Narmour's model, as well as the theory proposed here, to the results of Lake's experiment.

To test Narmour's model, Krumhansl identified and quantified the five principles listed in Table 10. According to the implication-realization

TABLE 10

Five Principles of Melodic Expectancy Underlying the Bottom-Up Component of the Implication-Realization Model

Registral Direction
Small implicative intervals imply realized intervals in the same direction.
Large implicative intervals imply realized intervals in a different direction.

Intervallic Difference
Small implicative intervals imply realized intervals that are similar in size.[a]
Large implicative intervals imply realized intervals that are smaller.[b]

Registral Return
The interval formed by the first tone of the implicative interval and the second tone of the realized interval is no greater than a major second.

Proximity
Independent of the size and direction of the implicative interval, implied realized intervals are no larger than a perfect fourth.

Closure
Closure is strongest when (1) the implicative interval is large and the realized interval is smallerb and (2) registral directions of implicative and realized intervals are different.

NOTE—From Krumhansl (1995).

[a]Within a minor third, if registral direction of implicative and realized intervals is the same; within a major second, if registral direction is different.

[b]Smaller by more than a minor third, if registral direction is the same; smaller by more than a major second, if registral direction is different.

model, the two-note beginnings (the notes x and y) used in Lake's experiment form the "implicative interval." When followed by a third note (z, the first response tone), the second and third notes form the "realized interval." According to the quantification that Krumhansl devised, every realized interval earns five scores, one for each of the principles listed in Table 10. (The scores for registral direction, intervallic difference, and registral return are "all or nothing" scores: if the realized interval follows the principle, it earns a score of 1; otherwise it earns a 0. The scores for proximity and closure earn "graded" scores: the score for proximity varies from 0 to 6; and the score for closure can be 0, 1, or 2.)

To Narmour's five principles, Krumhansl added two, which she called "unison" and "tonality." If the realized interval is a unison, it scores a 1 for that principle; otherwise it earns a zero (regardless of the implicative interval). For her experiment on tonal melodies (using excerpts from British folk songs), the tonality score is the position of the second tone of the realized interval (z, the first added note) in her tonal hierarchy (and thus varies from 1 to 7). Thus, Krumhansl's "tonality" is equivalent to what is called "stability" in this article.

Krumhansl then used the technique of multiple regression to assess how well the judgments made by her participants could be explained as a weighted sum of these seven scores. The results for her experiment on tonal melodies (using excerpts from British folk songs) are given in Table 11.

The coefficients given each of these factors tell us that proximity has the greatest predictive value for these data. Proximity—an element of magnetism—is also important in the theory of musical forces.

TABLE 11
Krumhansl's (1995) Test of the Implication-Realization Model

	British Folk Songs (Krumhansl, 1995)
Summary of multiple regression	$R = .846$ $p < .0001$ $N = 120$
Registral direction	$w_{RD} = .17$ $p < .005$
Intervallic difference	$w_{ID} = .17$ $p < .1$
Registral return	$w_{RR} = .31$ $p < .0001$
Proximity	$w_{PR} = .50$ $p < .0001$
Closure	$w_{CL} = .21$ $p < .005$
Unison	$w_{UN} = -.06$ *ns*
Tonality	$w_{TN} = .17$ $p < .005$

NOTE—Weights given as *w* are standardized coefficients.

In order to compare the implication-realization model to the theory of musical forces, we will apply Krumhansl's approach to Lake's data, comparing the contributions of factors derived from Narmour's theory and from the theory of musical forces. To be consistent with her approach, we consider all of Lake's data, but we limit our attention to the first response tone (the note z). For each cue (notes x and y, the "implicative interval"), we list all possible responses within an octave above and below the last note of the cue (y). There are 1800 such possible responses (75 cues times 24 responses—12 above and 12 below). For each resultant three-note pattern, we list the number of Lake's continuations that begin with those three notes. We calculate the same scores for Narmour's bottom-up factors that Krumhansl did (except that we again use levels of pitch-space embedding to represent stability). Then we calculate analogous scores for the musical forces. (For a given three-note pattern, if there is a gravity prediction for that cue that includes that first response tone, it scores a 1; otherwise it scores a 0. If there is a magnetism prediction for that cue that includes that first response tone, it scores a 1; otherwise it scores a 0. Here, the magnetism factor is "all or nothing"; proximity, included in the magnetism scores above, is listed here as the same separate factor coded by Krumhansl. If there is an inertia prediction for that cue that includes that first response tone, it scores a 2; if that cue does not generate an inertia prediction, it scores a 1; and if that cue does generate an inertia prediction, but not one that includes that first response tone, it scores a 0.) The results of this statistical analysis are given in Table 12.

For these data, the analysis fails to support four of Narmour's five factors (they are either statistically insignificant or negatively affect the performance of the model). Only proximity and the added factor of stability—both of which are also relevant in the theory of musical forces—contribute significantly to explaining Lake's data. (The cues in Lakes data feature mostly intervals smaller than a tritone. Had he included more larger cues, it is possible that more of Narmour's factors may have contributed positively to explaining the data.)

However, all factors in the theory of musical forces contribute significantly to explaining Lake's data. This gives additional support to the idea that we experience musical motions metaphorically in terms of our experience of physical motion.

Furthermore, although the stability of the goal pitch does turn out to be a significant factor, the analysis of these data also fails to support the other hypotheses suggested by a comparison of Larson's (1993a) and Lerdahl's (1996) algorithms: the stabilities of the other pitches are statistically insignificant, and gravity has significant predictive power (regardless of whether stability is included in the model).

The comparison also underscores important differences between my computer models of musical forces and the bottom-up component of Narmour's

TABLE 12

Multiple Regression Analysis Results for Lake (1987), Comparing the Implication-Realization Model to the Theory of Musical Forces

	Narmour's Five Bottom-Up Factors Plus Stability	Musical Forces Plus Stability, With Proximity Encoded Separately
Summary of multiple regression	R = .614 p < .0001 N = 1800	R = .773 p < .0001 N = 1800
Registral direction	w_{RD} = –.082 p = .0065	
Intervallic difference	w_{ID} =.049 p = .1151	
Registral return	w_{RR} = -.076 p = .0107	
Proximity	w_{PR} =.545 p < .0001	w_{PR} =.280 p < .0001
Closure	w_{CL} =.034 p = .1466	
Stability of first response tone (Lerdahl, 1988)	w_S =.314 p < .0001	w_S =.174 p < .0001
Gravity		w_G =.088 p < .0001
Magnetism		w_M =.432 p < .0001
Inertia		w_I =.187 p < .0001
Stability of last tone of cue		w = .009 p = .543

NOTE—Weights given as w are standardized coefficients.

implication-realization model. In order to compare these models, we must restrict our attention to just the first response tones. At least for Lake's data, the single-level model does a better job of predicting those first response tones than does Narmour's model. Far more important though, as noted earlier, my model can predict completions that match the experimental participants' *entire* completions *note-for-note* as often as any one of those participants does. The operational modeling of tonal relationships in terms of reference and goal alphabets and musical forces allows the computer models to predict listeners' expectations in a way that suggests that these models may capture important aspects of the underlying cognitive processes.

MORE STUDIES WITH RESPONSES OF VARIED LENGTH (LARSON, 1996, 1997A)

Two partial replications of Lake's study (Larson, 1996, 1997a) suggest that the responses of more experienced musicians may not only match those of other participants more often that do those of less experienced musicians, but may also match those of my computer models more often. In both experiments, musicians recruited from those attending national meetings of the Society for Music Theory were given melodic beginnings in notation, were asked to assume the context of C major, and were asked to write the responses that they thought would appear most often on other participants' pages.

Table 13 compares the responses of participants in the first of these experiments (Larson, 1996) with those of Lake's participants and with those of my computer models. Table 14 makes the same comparison for the second of these experiments (Larson, 1997a).

Four of these cues in Table 13 were the same as cues used in Lake's study (the fifth was an inversion of one of Lake's cues). Because Lake's 19 participants responded twice to each cue, and because Larson's 38 participants responded once to each cue, both groups supplied 38 responses to each cue, or 152 responses for these four cues. Out of their 152 responses, 92 (61%) of Larson's participants' appear in Table 13. Out of their 152 responses, 60 (39%) of Lake's participants' appear in Table 13. The number of cases (these 304 responses represent 14 unique responses to 4 different cues) is small enough to limit the conclusions that can be drawn, and the tasks differ (singing responses to heard sounds for the undergraduates versus writing responses to imagined sounds for the professional theorists). Nevertheless, the results are striking enough (61% versus 39%) to raise the question of whether the higher level of agreement might result

TABLE 13

Larson's (1996) Experimental Participants' Responses Compared With Lake's (1987), With the Single-Level Model and With the Multilevel Model

	Produced by Experimental Participants				Rating Assigned by the Single-Level Model		
	Larson (1996)		Lake (1987)				
Response: Letter Names	Number of Responses	Proportion of Responses to That Cue	Number of Responses	Proportion of Responses to That Cue	Numerical Ratings	Proportion of Ratings	Produced by the Multilevel Model
A E A	4	.24	0	.00			
A E F	5	.30	0	.00			√
A E F D C	5	.30	0	.00			√
A E G	3	.18	5	1.00	20.85	1.00	√
C A♭ G	14	.70	—	—	26.06	.58	√
C G♯ A	0	.00	—	—	18.59	.42	√
C A♭ G B C	3	.15	—	—			√
C A♭ G F E♭ D C	3	.15	—	—	i,r		√
D C♯ D	13	.68	9	.50	18.59	.43	√
D D♭ C	6	.32	9	.50	25.00	.57	√
E D C	27	.90	16	.72	23.13	1.00	√
E D C B C	3	.10	6	.27	r		√
G A C	0	.00	4	.20			
G A G E C	0	.00	3	.15			√
G A B	3	.12	0	.00			
G A B C	13	.50	9	.45	20.15	.49	√
G A G	7	.27	4	.20	20.85	.51	√
G A G F E D C	3	.12	0	.00	**r**		√

NOTE—Key: i = unrated inertia prediction, r = recursive prediction, — = not a cue used in Lake's study.

from greater musical training. Furthermore, Larson's participants come closer to the predictions of the single-level computer model. Once again, those predictions not listed for that model include many that would be produced as unrated inertia predictions, as recursive predictions, or by the multilevel model (if their hierarchical embellishment structures were represented).

Table 14 shows that the 61 participants in the second experiment (Larson, 1997a) produced 219 responses, in 26 unique responses, to these five cues, meaning that each unique response was sung an average of 8.42 times. Once again, these participants agreed more often with themselves and more often with the computer models than did Lake's participants.

The high levels of agreement between participants in these different studies: (1) argue that we are measuring robust musical behaviors, (2) raise the question of whether the differences between these participants may be due to differences in musical training, and (3) confirm the value of these data for testing the model presented here.

And once again, the computer programs come close to predicting all and only the responses of participants in various different experiments. Furthermore, where ratings may be assigned to predictions, the computer model gives higher ratings to patterns that participants sing more often.

Conclusions

This paper has argued that

> experienced listeners of tonal music expect completions in which the musical forces of gravity, magnetism, and inertia control operations on alphabets in hierarchies of embellishment whose stepwise displacements of auralized traces create simple closed shapes.

The instruction set given above suggests how this may happen. The two computer programs described here implement aspects of that instruction set.

The completions generated by the single-level model match note-for-note the entire completions produced by participants in several psychological studies as often as do the completions of any one of those participants. Furthermore, the ratings it gives to its completions correlate with the popularity of those responses; higher ratings are given to responses that participants sing more often. Those ratings are assigned by the computer's algorithm for the interaction of musical forces. While the best results are achieved by adding to that original algorithm a separate factor for the stability of the response tone, none of the studies reported here support any of the other hypotheses suggested by the differences between Lerdahl's (1996) and Larson's (1993a) algorithms.

TABLE 14

Larson's (1997a) Experimental Participants' Responses Compared With Lake's (1987), With the Single-Level Model and With the Multilevel Model

	Produced by Experimental Participants				Rating Assigned by the Single-Level Model		
	Larson (1997a)		Lake (1987)				Produced
Response: LetterNames	**Number of Responses**	**Proportion of Responses to That Cue**	**Number of Responses**	**Proportion of Responses to That Cue**	**Numerical Ratings**	**Proportion of Ratings**	**by the Multilevel Model**
Two-note cues							
D F E	11	.30	9	.24	21.88	1.00	√
D F E C	7	.19	5	.13	r		√
D F E D C	15	.41	5	.13	r		√
D F E G F D C	4	.11	0	.00			
F D B C	0	.00	4	.11			√
F D C	10	.23	5	.13	23.13	1.00	√
F D E	6	.14	4	.11			√
F D E C	12	.27	0	.00			√
F D E C D B C	5	.11	0	.00		√	
F D E D C	8	.18	0	.00			
F D E F E D C	3	.07	0	.00			
Three-note cues							
G E F D	3	.07	—	—			
G E F D C	19	.42	—	—			√
G E F D E D C	5	.11	—	—			
G E F E	4	.09	—	—	21.88	.63	√
G E F E D C	6	.13	—	—	r		√
G E F G	3	.07	—	—	13.13	.38	√
G E F G C	6	.13	—	—	r		√
G F E D C	44	.92	—	—	i, r		√
G F E F E D C	4	.08	—	—			√
C A G F E D C	14	.32	—	—	i		√
C A G F E	14	.32	—	—	i		√
C A G E F D C	5	.11	—	—			√
C A G E C	4	.09	—	—			√
C A G B C	4	.09	—	—			√
C A G C	3	.07	—	—	21.12	.48	√

√NOTE—Key: i = unrated inertia prediction, r = recursive prediction, — = not a cue used in Lake's study.

The multilevel model takes a melodic beginning, together with a description of its hierarchical embellishment structure (a Schenkerian analysis) and returns a completion that reflects all the levels of that structure. The success of the multilevel model suggests the importance of hierarchical structure in melodic expectation and illustrates how different interpretations of the same melodic beginning may lead to different expectations for its completion.

The striking agreement between computer-generated and participant-generated responses suggests that the theory captures some important aspects of melodic expectation. Furthermore, the fact that these data can be modeled well by the interaction of constantly acting but contextually determined musical forces gives support to the idea that we experience musical motions metaphorically in terms of our experience of physical motions.[6]

References

Aksnes, H. (1997, June). A cognitive approach to musical analysis: Metaphorical projection in music. In *Proceedings of the Third Triennial ESCOM Conference* (pp. 551–556). Uppsala, Sweden: Uppsala University.

Aksnes, H. (2002). Music and its resonating body. In *Dansk Årbog for Musikforskning 2001* [Danish Yearbook for Music Research 2001] (Vol. 29, pp. 81–101). Copenhagen: Engstrøm & Sødring Musikforlag.

Arnheim, R. (1966). The gestalt theory of expression. *Toward a psychology of art* (pp. 51–73). Berkeley: University of California Press.

Arnheim, R. (1974). *Art and visual perception: A psychology of the creative eye* (the new edition). Berkeley: University of California Press.

Arnheim, R. (1986). Perceptual dynamics in musical expression. In *New essays on the psychology of art* (pp. 214–227). Berkeley: University of California Press.

Balzano, G. J. (1980). The group-theoretic description of 12-fold and microtonal pitch systems. *Computer Music Journal, 4,* 66–84.

Balzano, G. J. (1982). The pitch set as a level of description for studying musical pitch perception. In M. Clynes (Ed.), *Music, mind, and brain* (pp. 321–351). New York: Plenum.

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

Bharucha, J. J. (1984b). Event hierarchies, tonal hierarchies, and assimilation. *Journal of Experimental Psychology, 113,* 421–425.

Bharucha, J. J. (1996). Melodic anchoring. *Music Perception, 13,* 383–400.

6. The Center for Research on Concepts and Cognition at Indiana University supported initial research on this project. My colleagues at CRCC, especially Doug Hofstadter, Gary McGraw, and James Marshall, offered thoughtful ideas on music and considerable help with computer programming.

A grant from the University of Oregon provided time in the summer of 1999 to explore the theory through additional experiments and computer models. Statistical consultants Robin High and Joseph St. Sauver of Academic Services at the University of Oregon helped with the data analyses.

Bill Lake supplied me with a copy of his dissertation and offered help and advice. Erick Carballo helped translate Lake's transcriptions of his participants' responses from musical notation into integers that the computer could read.

Special thanks to Fred Lerdahl, who heard my first public presentation of these ideas (Larson, 1993a) and has offered suggestions and encouragement on this work in various stages since then. At that same 1993 presentation, Carol Krumhansl and Jamshed Bharucha noted the similarity between my computer's algorithm for the interaction of musical forces and the general formula for multiple regression and encouraged me to learn about this method of statistical analysis. Thanks also to David Butler, Steven Demorest, Thomas Goolsby, Robert Hurwitz, Bunny Laden, John Rahn, and Sonja Rasmussen for reading drafts of this article.

This article has benefited immensely from the patient and helpful suggestions by Mari Riess Jones and two other anonymous readers, and by Lola Cuddy, editor of this journal.

My computer models were written in MacScheme. Scheme is a dialect of LISP invented by Guy L. Steele, Jr. and Gerald Jay Sussman. Scheme for the Macintosh was developed at MIT (© 1984). MacScheme is published by Lightship Software, Inc. (© 1992).

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: The MIT Press.

Brower, C. (1997–98). Pathway, blockage, and containment in “Density 21.5.” *Theory and Practice, 22-23,* 35–54.

Brower, C. (2000). A cognitive theory of musical meaning. *Journal of Music Theory, 44,* 323–379.

Carlsen, J. C. (1981). Some factors which influence melodic expectancy. *Psychomusicology, 1,* 12–29.

Carlsen, J. C., Divenyi, P. L., & Taylor, J. A. (1970). A preliminary study of perceptual expectancy in melodic configurations. *Council for Research in Music Education Bulletin, 22*, 4–12.

Clough, J., & Douthett, J. (1991). Maximally even sets. *Journal of Music Theory, 35,* 93–173.

Clough, J., & Myerson, G. (1985). Variety and multiplicity in diatonic systems. *Journal of Music Theory, 29,* 249–270.

Cohn, R. (1996). Maximally smooth cycles, hexatonic systems, and the analysis of late-romantic triadic progressions. *Music Analysis, 15,* 9–40.

Coker, W. (1972). *Music and meaning.* New York: The Free Press.

Cox, A. (1999). *The metaphoric logic of musical motion and space.* Unpublished doctoral dissertation, University of Oregon, Eugene.

Cuddy, L. L., & Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychophysics, 57,* 451–462.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 88,* 503–522.

Dibben, N. (1994). The cognitive reality of hierarchic structure in tonal and atonal music. *Music Perception, 12,* 1–25.

Gamer, C. (1967). Some combinatorial resources of equal-tempered systems. *Journal of Music Theory, 11,* 32–59.

Gjerdingen, R. O. (1994). Apparent motion in music? *Music Perception, 11,* 335–370.

Guck, M. (1981). *Metaphors in musical discourse: The contribution of imagery to analysis.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Guck, M. (1991). Two types of metaphoric transfer. In J. Robinson (Ed.), *Music and meaning* (pp. 201–212). Ithaca, NY: Cornell University Press.

Hofstadter, D., & the Fluid Analogies Research Group (1995). *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought.* New York: Basic Books.

Hurwitz, R. I., & Larson, S. (1994). Step collections in aural theory. In G. Karpinski (Chair), *Applying a theory of expressive meaning in the written- and aural-theory classrooms.* Panel session conducted at the national meeting of the College Music Society, Savannah, GA.

Johnson, M. (1987). *The body in the mind.* Chicago: University of Chicago Press.

Jones, M. R. (1981a). Music as a stimulus for psychological motion: Part I. Some determinants of expectancies. *Psychomusicology, 1,* 34–51.

Jones, M. R. (1981b). Music as a stimulus for psychological motion: Part II. An expectancy model. *Psychomusicology, 1,* 1–13.

Kassler, J. (1991). *Metaphor: A musical dimension.* Sydney: Currency Press.

Kidd, G. R. (1984). *The perception of tonal relationships in music.* Unpublished doctoral dissertation, Ohio State University, Columbus.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* New York: Oxford University Press.

Krumhansl, C. L. (1995). Music psychology and music theory: Problems and prospects. *Music Theory Spectrum, 17,* 53–80.

Lake, W. (1987). *Melodic perception and cognition: The influence of tonality.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought.* Chicago: University of Chicago Press.

Larson, S. (1992). *Scale-degree function: Cognition research and its application to the teaching of aural skills.* Paper presented at the meeting of the Society for Music Theory, Kansas City. (Available from Center for Research on Concepts and Cognition; 510 North Fess; Bloomington, IN 47408 as CRCC Technical Report #67.)

Larson, S. (1993a). *Computer models of melodic implication and key determination in tonal music.* Paper presented at the Society for Music Perception and Cognition, Philadelphia. (Available from Center for Research on Concepts and Cognition; 510 North Fess; Bloomington, IN 47408 as CRCC Technical Report #77.)

Larson, S. (1993b). *Modeling melodic expectation: Using three "musical forces" to predict melodic continuations.* Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Lawrence Erlbaum Associates.

Larson, S. (1993c). On Rudolf Arnheim's contributions to music theory. *Journal of Aesthetic Education, 27,* 97–104.

Larson, S. (1993d). Scale-degree function: A theory of expressive meaning and its application to aural-skills pedagogy. *Journal of Music Theory Pedagogy, 7,* 69–84.

Larson, S. (1994a). Another look at Schenker's *Counterpoint. Indiana Theory Review, 15,* 35–52.

Larson, S. (1994b). Musical forces, step collections, tonal pitch space, and melodic expectation. In *Proceedings of the Third International Conference on Music Perception and Cognition, Liège, Belgium* (pp. 227–229). (Available from Center for Research on Concepts and Cognition; 510 North Fess; Bloomington, IN 47408 as CRCC Technical Report #111.)

Larson, S. (1996). Expert expectations: Professional theorists' continuations compared with a computer model of musical forces. In *Current research in music cognition and aural training.* Poster presented at national meeting of the Society for Music Theory, Baton Rouge, LA.

Larson, S. (1997a). Continuations as completions: Studying melodic expectation in the creative microdomain Seek Well. In M. Leman (Ed.), *Music, gestalt, and computing: Studies in cognitive and systematic musicology* (pp. 321–334). Berlin: Springer-Verlag,.

Larson, S. (1997b). The problem of prolongation in tonal music: Terminology, perception, and expressive meaning. *Journal of Music Theory, 41,* 101–136.

Larson, S. (1997–98). Musical forces and melodic patterns. *Theory and Practice, 22-23,* 55–71.

Larson, S. (1999a). *Musical imagery and melodic expectation.* Paper presented at the Conference on Musical Imagery, Oslo, Norway.

Larson, S. (1999b). Swing and motive in three performances by Oscar Peterson. [In "Analysis forum" on Cole Porter's "Night and Day."] *Journal of Music Theory, 43,* 283–313.

Larson, S. (2002). Musical forces, melodic expectation, and jazz melody. *Music Perception, 19,* 351–385.

Lerdahl, F. (1988). Tonal pitch space. *Music Perception, 5,* 315–349.

Lerdahl, F. (1996). Calculating tonal tension. *Music Perception, 13,* 319–363.

Lerdahl, F. (2001). *Tonal pitch space.* New York: Oxford University Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

Marvin, E., & Brinkman, A. (1999). The effect of modulation and formal manipulation on perception of tonic closure by expert listeners. *Music Perception, 16,* 389–408.

Meyer, L. B. (1956). *Emotion and meaning in music.* Chicago: University of Chicago Press.

Meyer, L. B. (1973). *Explaining music.* Berkeley: University of California Press.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model.* Chicago: The University of Chicago Press.

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The implication-realization model.* Chicago: The University of Chicago Press.

Pelto, W. (1994). An alternative to rule memorization for written theory. In G. Karpinski (Chair), *Applying a theory of expressive meaning in the written- and aural-theory classrooms.* Panel session at the meeting of the College Music Society, Savannah, GA.

Povel, D. (1996). Exploring the fundamental harmonic forces in the tonal system. *Psychological Research, 58,* 274–283.

Pressing, J. (1978). Towards an understanding of scales in jazz. *Jazzforschung/Jazz Research, 9,* 25–35.

Rothfarb, L. (2001). Energetics. In T. Christensen (Ed.), *The Cambridge history of western music theory* (pp. 927–955). Cambridge: Cambridge University Press.

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition, 58,* 75–125.

Schellenberg, E. G. (1997). Simplifying the implication-realization model of melodic expectancy. *Cognition, 58,* 75–125.

Schenker, H. (1979). *Free composition: Volume III of New musical theories and fantasies* (E. Oster, Ed. & Trans.). New York: Longman. (Original work published 1935)

Simon, H. A., & Sumner, R. K. (1968). Pattern in music. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 219–250). New York: Wiley.

Thompson, W. F., Cuddy, L. L., & Plaus, C. (1997). Expectancies generated by melodic intervals: Evaluation of principles of melodic implication in a melody-completion task. *Perception & Psychophysics, 59,* 1069–1076.

Tymoczko, D. (1997). The consecutive-semitone constraint on scalar structure: A link between impressionism and jazz. *Intégral, 11,* 135–179.

Unyk, A. M., & Carlsen, J. C. (1987). The influence of expectancy on melodic perception. *Psychomusicology, 7,* 3–23.

Westergaard, P. (1975). *An introduction to tonal theory.* New York: W. W. Norton.