FOUNDATIONS OF MUSICAL GRAMMAR
Foundations of Musical Grammar
OXFORD STUDIES IN MUSIC THEORY

Series Editor Steven Rings

Studies in Music with Text, David Lewin

Music as Discourse: Semiotic Adventures in Romantic Music, Kofi Agawu

Metric Manipulations in Haydn and Mozart: Chamber Music for Strings, 1787–1791, Danuta Mirka

Songs in Motion: Rhythm and Meter in the German Lied, Yonatan Malin

A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice, Dmitri Tymoczko

In the Process of Becoming: Analytic and Philosophical Perspectives on Form in Early Nineteenth-Century Music, Janet Schmalfeldt

Tonality and Transformation, Steven Rings

Audacious Euphony: Chromaticism and the Triad’s Second Nature, Richard Cohn

Mahler’s Symphonic Sonatas, Seth Monahan

Beating Time and Measuring Music in the Early Modern Era, Roger Mathew Grant

Pieces of Tradition: An Analysis of Contemporary Tonal Music, Daniel Harrison
Foundations of Musical Grammar

Lawrence M. Zbikowski
Dem Andenken meines Bruders:
Eugene Francis Zbikowski, 1953–2015
CONTENTS

Preface ix
Acknowledgments xiii

1. Origins 1
   Music and Language in Human Cultures 5
   Foundations of Musical Grammar 16
   Musical Grammar and Music Theory 25

2. Music and Analogy 26
   Analogy 28
   Perceptual Symbol Systems, Embodied Knowledge, and Analogy 32
   Sonic Analogs and Systems of Reference 38
   Sonic Analogs for Dynamic Processes 42
   Analogy and Imagination 47
   Conclusion 52

3. Music and Emotion 56
   Recent Research on Emotion 63
   Recent Research on Music and Emotion 70
   Music, Emotions, and Sonic Analogs 77
   Conclusion 88

4. Music and Gesture 95
   Gesture and Language 101
   The Infrastructure of Human Communication 105
   Music and Movement 114
   Music and Gesture 124
   Conclusion 127

5. Music and Dance 129
   Music and Dance in the Ancien Régime 133
   Of Waltzes and Waltz Music 147
   Conclusion: Music, Social Dance, and Musical Grammar 165

6. Music and Words 167
   “Über allen Gipfeln ist Ruh” 169
   Music, Words, and Song 195
   Conclusion 198
PREFACE

Thomas Busby opened his 1818 *A Grammar of Music*—one of the many such volumes produced in the late eighteenth and early nineteenth centuries—with this observation:

Music is a science having sounds for its element, and teaching their due division, succession, and combination. By the theory of *Music*, we are enabled to measure the intervals between sonorous impressions; to calculate upon their effects, as resulting from their succession and their union; and so to understand the principles of *melody* and *harmony*, as not only to be qualified to comprehend and appreciate the compositions of others, but to produce compositions of our own. (Busby 1818, vii)

He then, in one long but carefully worded paragraph, surveyed the history of music from antiquity through the end of the eighteenth century and, after a few more preliminaries, finally turned to his main topic. “Musical Grammar,” Busby wrote, “comprehends the knowledge necessary to the just arrangement and combination of musical sounds, and to the proper performance of musical compositions” (Busby 1818, 1).

Busby’s perspective is hardly unique, conforming in its basic outlines with countless musical textbooks and primers produced over the past two hundred years. The only thing that has changed is that “musical grammar,” as a term, has fallen into disuse. Today, were we to describe the discipline that “comprehends the knowledge necessary to the just arrangement and combination of musical sounds, and to the proper performance of musical compositions,” we would simply call it “music theory.”

In this volume I aim to return to the idea of musical grammar and to revivify it by drawing on recent research in cognitive science. Although I consider the result to fall well within the domain of music theory, it is not a very typical theory of music. One reason is that the understanding of what is involved with grammar that is developed here is shaped to a considerable extent by the work of a group of linguists who, over the past three decades, have been exploring how humans’ cognitive capacities shape the construction of language and the way it is used to communicate meaning. The approach they have developed has come to be known as cognitive linguistics. One of the foundations of cognitive linguistics is the notion that the basic elements of language consist of stored pairings of form and function. The result is what is often called a cognitive grammar, with “grammar” here understood to combine features of both syntax and semantics.
Applying this notion of grammar to music has a number of interesting consequences. Perhaps the most important is a fresh view of musical organization. While many music theorists are more or less comfortable with the notion of musical syntax, understanding it to be the basis for the “just arrangement and combination of musical sounds,” the question of musical meaning remains quite open. The idea that the basic grammatical elements of music combine features of both syntax and semantics thus invites us to look a bit further into how musical utterances are constructed and how they are used to communicate meaning.

This idea also offers an opportunity for a thoughtful comparison of the expressive resources offered by language and music. That these resources are, in some measure, distinct from one another seems evident from the different roles language and music have played in human cultures past and present. That said, this distinction has not always been clear in accounts of how language and music are organized to create meaningful utterances. It is often the case, for instance, that language is accorded a privileged status in the way we think about communication. Thus Leonard Bernstein, in his 1973 Norton lectures, proposed that “by studying in depth why we talk the way we do—by abstracting the logical principles of language—we may be in a position to discover how we communicate in a larger sense: through music, through the arts in general, and ultimately through all our societal behavior” (1976, 8). An approach of the sort I would like to pursue here, which puts the “how” of communication (its syntax) on an equal footing with the “why” of communication (its semantics), offers the possibility of thinking about grammars that are specific to each communicative medium: the point is not how much music is like language, or how much language is like music, but how each is organized for specific communicative purposes. The model of musical grammar that emerges is one in which the dynamic nature of musical utterances is brought into the foreground, and which in many cases involves interwoven syntactic layers that unfold simultaneously. It is a model markedly different from any of the models currently proposed for language, and one that is consonant with the resources music offers for human communication.

The commitment to a cognitive grammar requires setting out the cognitive capacities that support the productivity of the grammar—that is, giving some account, based on what is currently known about human cognitive processing, why utterances take the form they do and how they realize their intended function. It bears mention that much of the current work in cognitive science regards the brain and nervous system as interacting, at practically every turn, with the capacities of the body. Cognitive processing is thus not strictly limited to what happens within the brain but extends to the body and, in some accounts—such as that of the philosopher Andy Clark (2008, 2016)—out into the world. The idea of a cognitive grammar fits quite well with the perspective I set out in my earlier book, *Conceptualizing Music*, where I explored the cognitive capacities that make possible the understanding and production of music. This is hardly surprising, since *Conceptualizing Music* borrowed from the methodology of cognitive linguists to develop this perspective, although it stopped well short of offering anything other than a few scattered observations on musical grammar. Here,
building on and expanding that perspective, I develop those observations quite a bit further, both in the interests of discovering some of the central features of a cognitive grammar of music and with the hope that this may further our understanding of how the cognitive capacities of humans shape the understanding and production of music.

One of the central goals of Conceptualizing Music was to put the inquiries undertaken by music theorists in dialog with what we currently know about human cognitive processing. This book continues that venture by providing music theorists with some of the tools—by way of an account of the elements of a cognitive grammar of music—through which they might build systematic interpretations of individual works (or recordings, or performances) that fit with the findings of cognitive science as well as with theories about the role various modes of communication play in human cultures. To that extent, what I offer here conforms, in a general way, with Busby’s notion that musical grammar comprehends how musical sounds are arranged and combined. In at least one important aspect, however, what I propose here is markedly different from the model offered by Busby, for whom music theory was a science of measurement and calculation. While I would never wish to gainsay the power of mathematical thought, the approach I take here trusts less in the certainties of mathematics and more in the contingencies of the human: music is, for me, an expression of what it means to be human, and what order there is in it is a human order, one constantly transformed by the quiddities of the human enterprise.

The audience I had in mind while writing this book was one interested in an account of musical organization grounded in recent research in cognitive science that could support close readings of musical works. As I envisioned them, such readings could be put to a variety of uses, including hermeneutic interpretation, the preparation of performances, historical contextualization, and an exploration of music’s role in social exchanges, as well as for the study of music’s relationships with expressive genres like gesture, dance, and poetry. These readings might also be done for the pure pleasure of discovering some of the compositional strategies behind musical utterances that have, time and again, held us in thrall. While I have imagined that music theorists, historical musicologists, and ethnomusicologists might all like to pursue these sorts of close readings of musical works, I hope the space of discourse I have opened up is capacious enough that it would extend to anyone who is interested in the challenges music poses to our ideas about how humans organize their understanding of the world, as well as to those who would like to explore how nonlinguistic forms of expression might be organized.

Among the ideas that I develop in the following pages is the notion that, at the most fundamental level, musical utterances provide sonic analogs for dynamic processes. This notion fits quite well with a striking image offered by Elizabeth Margulis in her recent book on the role of repetition in musical organization. Reflecting on the impression created by music with which we have become familiar, Margulis remarked, “Each moment seems not like a bead strung along a necklace, resting next to dozens of other beads, but more like a drink just when it starts to be poured—the cascade of liquid is so much a part of the gesture as to seem to
It is this cascade of liquid that I wish to study in what follows, knowing that the best I can do with the linguistic resources available to me is the equivalent of setting a slow shutter speed when taking a photograph. Using a slow shutter speed not only offers a kind of truth about the relentless flow of fluid phenomena, it also saturates its surroundings in deeper colors. This, then, is the perspective I hope to develop in my exploration of the foundations of musical grammar.
Some projects take longer than others. This one started with a leave at the National Humanities Center in 2003–4, which was supported in part by the National Endowment for the Humanities. Although I had proposed to the Center a study that I thought might yield a couple of articles, I soon came to see how I might write a book that investigated musical grammar from the viewpoint provided by research in cognitive science. A leave in 2006–7 supported by the Division of Humanities of the University of Chicago allowed me to make further progress toward realizing this ambition, as did another leave in 2010–11, which was supported by a fellowship from the American Council of Learned Societies and a Fulbright Visiting Research Chair in Cognitive Music Theory at the Centre for Interdisciplinary Research in Music Media and Technology, which has its home in the Schulich School of Music of McGill University. My heartfelt thanks to all of these institutions, without which not only this volume but much of what I have written since my first book would not have been possible.

Over the years I was able to present preliminary versions of my research at a number of institutions, and at various conferences throughout the northern hemisphere. The list of these institutions and conferences is somewhat long, and I cannot begin to remember all of the colleagues whose questions, comments, and constructive resistance stimulated my thought. Nonetheless, I have come to appreciate that academic work simply cannot thrive without the opportunities provided by such exchanges, and to all who participated I owe my sincere thanks. There is, however, one gathering in particular that was instrumental to the development of my theory of musical grammar, and that was a small working conference on communicative strategies in music of the late eighteenth century that Danuta Mirka organized at Bad Sulzburg, outside Freiburg, Germany, in 2005. It was at that meeting—deep in the Black Forest and in the company of scholars like Kofi Agawu, Janet Schmalfeldt, Dean Sutcliffe, Bill Rothstein, Professor Mirka herself, and the late Wendy Allanbrook—that I was able to present the first working model of the approach to musical grammar set out in more detail here. It is fitting that the present volume ends by touching on the analysis of a Haydn quartet initially sketched for that conference.

My academic home for more than two decades has been the Department of Music of the University of Chicago. Interactions with colleagues, students, and visitors over that time have convinced me that there is simply no better place to ask and explore the kinds of questions that have captivated me ever since I began
to think seriously about how music was organized. To my colleagues Thomas Christensen and Steven Rings I owe a special thanks: without their continued confidence in the way music theory is presented and studied in the department—which they shared in the classroom, in faculty meetings, and at the pub—a book like this one would simply have never come to fruition.

Deep and sustained thanks goes to my editor at Oxford University Press, Suzanne Ryan. Early on, Ms. Ryan saw the value of this project and waited patiently for me to wrestle it into submission, and she has never stopped believing in what has been at times an unbelievable muddle of scholarship, enthusiasm, and broad-ranging theoretical inquiry. Thanks as well to Professor Rings, who encouraged me to submit the manuscript to Oxford Studies in Music Theory and who enthusiastically advocated for its inclusion in the series.

My family—Vicky Long, Anna Katia, and Andrei Nicolai—have long grown accustomed to my working on this project, and although they had some expectation it would eventually appear in final form they were considerate enough not to press the point. Without their support, understanding, and uncritical companionship, the work that I have done over the past decade would have been much the poorer. My brother, Gene Zbikowski, helped me both to clarify my translation of passages in French and German and to appreciate the nuances and ambiguities of these languages. Although I was able to share with him the completion of the manuscript, his sudden death robbed me of further opportunities to draw on his considerable knowledge of language and unfailing good humor about the scholarly back alleys into which I have been drawn. It is to his memory that this book is dedicated.

When dealing with historical sources written in languages other than English, I have generally followed one of two practices. Where there is no translation of the work, or where I think existing translations are not wholly satisfactory, I provide my own translation, with the original immediately following. Where satisfactory translations exist, I simply use these, although in a few cases I have modified the translation slightly (and note this in the citation).

Foundations of Musical Grammar
Of the many reflections on the relationship between music and language written during the eighteenth century, perhaps the most striking is one the author never saw fit to publish. Jean-Jacques Rousseau’s “Essay on the Origin of Languages” (c. 1763) is concerned, at its beginning and end, with the role language played in the formation of the social contract. Rousseau takes us back to a time when nomadic tribes from arid southern regions came together around wells that provided both nourishing water and a safe haven for social intercourse.

Young girls came to fetch water for the household, young men came to water their herds. There eyes accustomed to the same objects from childhood began to see sweeter ones. The heart was moved by these new objects, an unfamiliar attraction made it less savage, it felt the pleasure of not being alone. Imperceptibly water became more necessary, the livestock were thirsty more often; they arrived in haste and parted reluctantly. In this happy age when nothing marked the hours, nothing obliged them to be counted; time did not have any measure other than amusement and boredom. Beneath aged oaks, conquerors of years, an ardent youth gradually forgot its ferocity, gradually they tamed one another; through endeavoring to make themselves understood, they learned to explain themselves. There the first festivals took place, feet leaped with joy, eager gesture no longer sufficed, the voice accompanied it with passionate accents; mingled together, pleasure and desire made themselves felt at the same time. There, finally, was the true cradle of peoples, and from the pure crystal of the fountains came the first fires of love. (Rousseau [1763] 1998, 314)

As Rousseau would have it, ardent passion, rather than cold necessity, was the key to human society and to language. Various animals, he argued, have needs just as humans do, needs that can be satisfied without recourse to language. To be human, however, required more than this—it required recognizing, sharing, and attempting to alter the passions of others. “Fruit does not elude our grasp, one can feed on it without speaking, one stalks in silence the prey one wishes to devour; but in order to move a young heart, to repulse an unjust aggressor, nature dictates accents, cries, complaints” (Rousseau [1763] 1998, 294). And it is these utterances, produced to move the passions, that are the source of language.

Rousseau’s fanciful account of the origin of language echoes one he offered in the second Discourse, published in 1755, but it also aims to address a problem raised there. The key passage runs as follows: “For a long time inarticulate cries, many gestures, and some imitative noises must have composed the Universal language; by joining to this in each Country a few articulated and conventional sounds—the
institution of which, as I have already said, is not too easy to explain—there were particular languages, but crude imperfect ones, approximately like those which various Savage Nations still have today” (Rousseau [1755] 1992, 45). The solution offered in the “Essay” is twofold. First, language at its origin was more than a tool to facilitate the cooperation among humans (which was the view Rousseau took in the second Discourse)—it was the means to move the passions. As Downing Thomas has noted, la pitié, or compassion, was a crucial concept for Rousseau: “in compelling sympathy and identification, la pitié forms a bond of society and communication between two beings” (1995, 87). Through moving the passions, language made possible a truly human society. Second, primordial language was able to move the passions directly because it had almost nothing in common with modern languages. “Not only would all the turns of phrase in this language have to be in images, in feelings, and in figures of speech; but in its mechanical aspect it would have to answer to its first object, and to present to the sense as well as to the understanding the almost inevitable impressions of the passion that is sought to be communicated” (Rousseau [1763] 1998, 295). What, then, would this first language have been like? It would have been like music. Rousseau writes, “Around the fountains of which I have spoken, the first discourses were the first songs; the periodic and measured recurrences of rhythm, the melodious inflections of accents caused poetry and music to be born along with language; or rather, all this was nothing but language itself in those happy climates and those happy times when the only pressing needs that required another’s help were those to which the heart gave rise” (Rousseau [1763] 1998, 318).

At this point in the “Essay” Rousseau changes focus, turning away from the origin of language and toward an explanation of how the first songs could accomplish what various gestures, grunts, and imitative noises could not. The key was the efficacy of melody: “The sounds of a melody do not act on us solely as sounds, but as signs of our affections, of our feelings; it is in this way that they excite in us the emotions they express and the image of which we recognize in them” (Rousseau [1763] 1998, 323). Melody was thus more than a succession of pleasant noises or a shallow imitation of natural sounds—it was a representation of the essential features of the passions. More complex music—specifically, music that involved harmony—could not accomplish this, for such music destroyed the delicate shadings of melody and set speech and song in conflict with one another. On the one hand, this explanation addressed the crux of the problem raised in the second Discourse: language, at its origins, drew its power not from arbitrary correlations between specific sounds and specific meanings but from its embodiment of the passions through its melody-like qualities. On the other hand, Rousseau’s account allowed him to trace the path of degeneration that led from the idyllic world of the first languages to the Babel of the modern world. Along this path, music and language separated: music, encumbered by harmony, lost its ability to enthrall; language, sundered from its connections to the passions, lost its ability to persuade.

Since its publication in 1781, Rousseau’s account of the origin of human communication has occasioned a wealth of commentary. The novelty of the perspective and the fact that it was not published during Rousseau’s lifetime prompted a
number of writers to question whether it should be taken at all seriously. Among those who did take it seriously was Jacques Derrida, who used it as a centerpiece for the reflection on philosophical discourse he offered in *Of Grammatology* (1976). For Derrida, Rousseau’s essay offered an opportunity to lay bare a fundamental tension in Western thought, which was its reliance on representation. Rousseau’s proposal that language emerged only after humans had discovered how to move one another’s passions, and that this emergence initiated a decline into civilization, opened up the possibility that representation was not a necessary feature of human thought. Richard Rorty, in a commentary on the perspective on philosophical writing for which Derrida advocated, characterized that perspective as one in which “we must question simultaneously the notion that philosophy is an inquiry into representation, and the notion that ‘representation’ is a useful concept” (1977, 680). In his perceptive critique of Derrida’s reading of Rousseau, Paul de Man observed that Derrida achieved this perspective in part by eliding the role music plays in the “Essay.” Through a careful reading of Rousseau’s account of the effects that music was able to achieve, de Man set out the argument that music provided the first demonstration of a model of communication in which representation played no part, a model that then informed the Rousseauian account of language celebrated by Derrida (de Man 1983, 124–31). Although de Man’s interpretation is compelling, it is nonetheless compromised by the idea, shared with Derrida, that all forms of communication strive toward the state of language, in which “representation” was achieved by the correlation of the sign with a real or imagined state of the world. What both Derrida and de Man missed in their encounters with Rousseau’s “Essay” was the opportunity to consider whether musical representation might, in some way, be fundamentally different from the mode of representation employed by language.

The possibilities of that alternative are clearly set out in the contrast between painting and music that Rousseau draws midway through the essay. Painting, for Rousseau, is always about objects portrayed through line and color. Music, by contrast, is always about movement, which it can use to portray even stillness and silence.

Let nature as a whole be asleep, he who contemplates it sleeps not, and the musician’s art consists in substituting for the imperceptible image of the object that of the movements that its presence excites in the heart of the contemplator. Not only will it agitate the sea, fan the flames of a blaze, make streams run, rain fall, and torrents swell, but it will depict the horror of a frightful desert, darken the walls of an underground dungeon, calm a tempest, make the air tranquil and clear, and spread from the orchestra a renewed freshness over the groves. It will not represent these things directly, but will awaken the same feelings in the soul that are experienced in seeing them. (Rousseau [1763] 1998, 327)

In broad terms, the notion of representation Rousseau had in mind was based on theories of mimesis commonly used in the eighteenth century to explain how music achieved its effects. As but one example, in a treatise on poetry, painting, and music first published in 1719, the Abbé Dubos wrote, “Wherefore as the painter imitates the strokes and colors of nature, in like manner the musician
imitates the tones, accents, sighs, and inflexions of the voice; and in short all those sounds, by which nature herself expresses her sentiments and passions” (1748, 360–61). Dubos, for his part, was aware that music could go beyond imitating the human voice, especially when used for dramatic ends. The most typical case occurred when a composer needed to depict the forces or power of nature through purely instrumental means—what Dubos called “symphonies.” If such imitations were to be successful, however, the correlation between the musical and natural sounds—both in terms of the actual sonic materials and their context—had to be as exact as possible: “The truth of the imitation in symphonies consists in their resemblance with the sounds they are intended to imitate. There is truth in a symphony composed for the imitation of a tempest, when the modulation, harmony, and *rhythmus*, convey to our ear a sound like the blustering of the waves, which dash impetuous against one another, or break against the rocks” (Dubos 1748, 363–64). What neither Dubos, nor Rousseau, nor any of those who engaged with musical mimesis attempted to explain was the process through which one sound could come to stand for another. The famous tempest summoned by Marin Marais in his *Alcyone* (the 1706 opera that Dubos had in mind in his discussion of instrumental mimesis) does not, in truth, sound anything like an actual storm: storms rarely feature rushing scale passages, rapid repeated notes, or prolonged pedal points, all of which are central to Marais’s musical depiction. Indeed, it appears that ours is the only species that has the capacity to interpret sound sequences in this sort of an imaginative way (such that passagework in stringed instruments is heard to resemble the tumult of a tempest), extending even to the representation of phenomena that have no distinctive sonic component (such that sequences of musical sound can summon the terror of a frightful desert or evoke the gloom of an underground dungeon). That we can do so, I shall argue, is a testament to humans’ capacity for analogical thought. This capacity makes possible a form of reference quite different from that employed by language (and which I shall describe more fully in the following pages), one that is uniquely exploited by music and that is central to the account of musical grammar I shall develop in this volume.

It should be emphasized that, for Rousseau, properly formed sequences of musical sounds did not directly represent the agitations of the sea or the flames of a blaze but instead awoke internal feelings of the same sort that had been experienced in seeing them. This is in conformance with his ideas about the efficacy of melody and with his conception of the basic purpose of communication, which was to move the passions of another. Put in another (and somewhat more technical) way, the function of properly formed musical utterances was to effect a change in the internal feeling-states of a conspecific, for it was by this means the bonds of society could be forged. Although I should not want to make claims for Rousseau’s essay beyond what might have been appropriate within the context of eighteenth-century speculations on the origins of language, the notion that the form of a mode of communication has an intimate connection with its function is one that resonates with recent accounts of linguistic organization developed by cognitive linguists. These accounts are based on the assumption that the basic elements of language consist of stored pairings of form and function, and have led
to the development of what has been called cognitive grammar (with “grammar” here understood to combine features of both syntax and semantics). There are, of course, other ways to construe grammar, but the notion that it concerns both the form and function of a communicative medium turns out to be extremely useful for thinking about the ways language and music are organized to achieve communicative and expressive ends that, in many respects, are remarkably different from one another.

In what follows, I would like to explore how some of the assumptions basic to cognitive grammar have informed accounts of linguistic grammar, and how they might be applied to an account of musical grammar. I should note here that I offer only the briefest of overviews of previous work on cognitive grammars; interested readers should consult the work of Ronald Langacker (especially Langacker 2008) and William Croft (especially Croft and Cruse 2004) for more comprehensive introductions. Indeed, my aim in these opening pages is less to explicate linguistic grammar and more to explore how some of the basic presuppositions of the cognitive grammars developed by linguists could serve as a foundation for a principled account of musical organization. I shall thus be interested in setting out what I believe are basic and significant differences between linguistic and musical utterances, illustrated through analyses of a short recitative by Johann Sebastian Bach and a funeral dirge from Western Africa. These analyses will set the stage for a brief discussion of the basic features of musical grammar that will then be developed in the remaining chapters. Subsequent to an outline of topics covered in these chapters I shall offer a few concluding thoughts on the theory of musical organization I offer here and its place within the broader discipline of music theory.

Music and Language in Human Cultures

The Function and Structure of Language

Although we can only speculate on what the functions of early language might have been (just as we can only speculate on what forms realized these functions), we can get some sense of what language offered to our species by studying the change it makes in infants’ lives when they start to acquire it. It is a simple fact of human culture that most children begin producing and using language around the time of their first birthdays. What has not been simple has been explaining why this should be so. One proposal has been that this capacity is innate—that is, all children are born with a specific sort of knowledge about language structure (called Universal Grammar) and this, together with an inborn language-acquisition device (Chomsky 1965, 32–33), enables them to acquire their native language. The best explanation of how such a capacity would have appeared in our species (and in no others) is that it is the result of a fortuitous genetic mutation. This approach has not been satisfactory to everyone, on both theoretical and empirical grounds. As an alternative, a number of researchers have proposed that children acquire language as a consequence of behavioral developments that take
place during the first year of life, and that this process is supported by various more-general cognitive capacities that are applied to the acquisition of language. In particular, language is viewed as facilitating forms of social cognition that are crucial to the developing infant.

As those who have observed them know, human infants are intensely social creatures who closely monitor the behavior of and readily respond to others around them. For most of the first year of their life, infants’ interactions with people and objects in their world are dyadic—they will interact with a person, change focus and interact with an object, and then switch focus back to the person, but they do not typically engage with both a person and an object within the same referential frame. Around nine to twelve months of age a new set of behaviors begins to emerge, and infants start to coordinate object interactions with people interactions. The result is a referential triangle consisting of the infant, an adult, and some object or event. Within this referential frame both infant and adult jointly attend to the third element, and are aware not only of the other participant but also of their own role as a participant in the interaction (Baldwin 1995, 132–33). This blossoming of joint attention has long been recognized to be key to the acquisition of language, not the least because language provides the child with an important tool for managing joint attention (Bruner 1975; 1983, 70–77). Indeed, as Michael Tomasello has observed, the interactions that take place within joint attentional scenes reveal a major function of language, which is to manipulate the attention of another person (Tomasello 1999, 151).

Tomasello has gone on to propose that two learning mechanisms support children’s acquisition of language. The first mechanism is intention-reading, which includes the ability to share attention, to follow the attention and gesturing of other persons to distal objects, to actively direct the attention of others to distal objects, and to culturally learn the intentional actions of others. The second mechanism is pattern-finding, which includes the ability to form perceptual and conceptual categories (broadly understood as a kind of pattern), to create sensory-motor schemas derived from perceptual and proprioceptual information, and to make analogies across different domains (Tomasello 2003, 19–30). Although pattern-finding abilities are common among primates, intention-reading abilities are not. This suggests why, within naturally occurring populations, nonhuman primates do not typically move beyond dyadic interactions, and also why they do not develop language.

Establishing and exploiting joint attentional scenes, and using the referential symbols of language to manipulate the attention of others within such scenes, can give clear advantages to the developing child as she learns how to get along in her world. Presumably, these same abilities would have given her distant forebears evolutionary advantages as they learned how to survive in their world. According to this interpretation, language has its origins in the need to manipulate the attention of conspecifics within the context of a joint attentional scene; the structure of language, and indeed the systems of symbolic reference basic to it, follows from this.

One way to characterize the structure of language is to focus on what some linguists have called parts of speech (Croft 2001, chap. 2). An account of two of the most basic parts of speech that accords well with a functional perspective
was provided some ninety-five years ago by Edward Sapir, who wrote that in language

there must be something to talk about and something must be said about this subject of discourse once it is selected. The distinction is of such fundamental importance that the vast majority of languages have emphasized it by creating some sort of formal barrier between the two terms of the proposition. The subject of discourse is a noun. As the most common subject of discourse is either a person or a thing, the noun clusters about concrete concepts of that order. As the thing predicated of a subject is generally an activity in the widest sense of the word, a passage from one moment of existence to another, the form which has been set aside for the business of predicking, in other words, the verb, clusters about concepts of activity. No language wholly fails to distinguish noun and verb, though in particular cases the nature of the distinction may be an elusive one. (Sapir 1921, 126)

A noun thus provides a way to draw attention to the relevant element of a joint attentional scene—a subject for discourse—and a verb characterizes actions on or by this element. Parts of speech like these allow us to manipulate what another person pays attention to within a joint attentional scene.

More recently, Len Talmy proposed a comprehensive approach to semantics that recognizes parts of speech as belonging to one of two subsystems of meaning-bearing forms in language. The first of these is the open-class, or lexical, system. The second—the one to which parts of speech belong—is the closed-class, or grammatical, system. Open classes have a large number of members to which more are easily added. Closed classes, by contrast, have relatively few members and are more difficult to augment. These features shape the kinds of meaning associated with each class. Talmy observes that “the meanings that open-class forms can express are virtually unrestricted, whereas those of closed-class forms are highly constrained, both as to the conceptual category they can refer to and as to the particular member notions within any such category” (2003, 15). For instance, many languages have closed-class (grammatical) forms that combine with the root of a noun to indicate the number of its referent (in English, “s” is added to many nouns to change them from singular to plural), but no language has a closed-class form that indicates color. Put another way, closed-class forms are never used to indicate the color of a referent, although (in many languages) closed-class forms can give an approximate idea of number. If we wish to indicate the color of a referent or specify its number we must use open-class forms.

Consequent to the different semantic constraints they embody, these two types of classes have different functions in discourse. Talmy writes, “In the conceptual complex evoked by any portion of discourse, say, a sentence, the open-class forms contribute most of the content, whereas the closed-class forms determine most of the structure. Thus the inventory of conceptual categories and individual concepts that closed-class forms can ever express amounts to the fundamental conceptual structuring system used by language” (2003, 16). Although grammar has been regarded by some to be conceptually neutral, Talmy’s argument is that grammar shapes what we can conceive and how we conceive it. Grammar is thus basic to rather than distinct from semantics.
This perspective is one that is broadly shared within the disciplinary formation that has come to be known as cognitive linguistics, which is guided by the idea that language reflects the cognitive capacities of humans. As William Croft and Alan Cruse put it,

the cognitive processes that govern language use, in particular the construction and communication of meaning by language, are in principle the same as other cognitive abilities. That is, the organization and retrieval of linguistic knowledge is not significantly different from the organization and retrieval of other knowledge in the mind, and the cognitive abilities that we apply to speaking and understanding language are not significantly different from those applied to other cognitive tasks, such as visual perception, reasoning or motor activity. Language is a distinct human cognitive ability, to be sure. From a cognitive perspective, language is the real-time perception and production of a temporal sequence of discrete, structured symbolic units. This particular configuration of cognitive abilities is probably unique to language, but the component cognitive skills required are not. (Croft and Cruse 2004, 2)

The approach to linguistic knowledge that follows from this perspective—one in which language is viewed as a consequence of, rather than a precondition for, conceptualization—has given rise to what has come to be called construction grammar (Croft and Cruse 2004, chap. 10; Goldberg 1995, 2006). Constructions are defined as “stored pairings of form and function, including morphemes, words, idioms, partially lexically filled and fully general linguistic patterns” (Goldberg 2003, 219). Thus the word “run” (which is a member of Talmy’s open-class forms) when paired with VERB (a closed-class form) provides a way to characterize rapid ambulatory locomotion by some agent (“I had to run to catch the bus”). When paired with NOUN, however (another closed-class form), “run” provides a subject for discourse, whether that be a form of exercise (“I went for a run this afternoon”) or the score of a baseball game (“That run in the eighth inning won the game for us”). Again, the assumption is that these pairings of form and function obtain at all levels within the grammar of a language—as Adele Goldberg has remarked, “it’s constructions all the way down” (2006, 18).

The notions that the basic function of language is to manipulate the attention of another person within the context of a joint attentional scene, that this function is realized in its structure, and that the structure of language itself has semantic content represent but one possible configuration of ideas through which we might craft an account of the relationship between the development of language in our species and the way language is structured and organized. Most important for my purposes, however, is the general approach to human communication offered by this configuration, one which might be applied to music as it has been applied to language. How might we characterize the basic function of music in human cultures such that we could develop an understanding of ways music has been organized to realize this function? Over the course of human history there have, of course, been any number of answers to the first part of this question; the one I would like to offer takes as its point of departure the different forms of reference employed by language and music.
Forms of Reference and the Function of Music

Language provides a tool for moving beyond basic dyadic interactions because it exploits, to a degree seen in no other communicative medium, the mechanism of symbolic reference. Symbolic reference involves various symbolic tokens—sounds, hand gestures, ink marks, or the like—that are systematically correlated with various referents. If I wish to refer to any one of the woody perennial plants with a single elongate main stem having few branches on its lower part that I see outside my window I could utter a sound, use the appropriate gesture from American Sign Language, or simply write “tree.” None of these symbolic tokens need share any properties with the thing to which they refer: there does not, for example, have to be any relationship between the token for “tree” in American Sign Language and the shape of a tree. The fact that there is such a relationship in American Sign Language does not affect the way the token functions as a symbol, although it may help a user of the language to learn the connection between the token and its referent. (For further discussion of this last point, see Perniss, Thompson, and Vigliocco 2010, and Emmorey 2014.)

The inherent flexibility of symbolic reference makes it a powerful and versatile tool for communication. As the biologist Terry Deacon has observed in his work on language evolution, it is also a tool that is remarkably complex beneath its seemingly simple surface. Deacon, building on some of the key insights of Charles Sanders Peirce’s semiotic theory, shows that symbolic tokens can function only as part of a system of symbols. Within this system, symbols are connected not only to the things to which they refer but also to each other (Deacon 1997, chap. 3; 2003; 2006). The management of these manifold interconnections places considerable demands on cognitive resources, and Deacon argues that it is for this reason that no other species has been able to make use of symbolic reference with anything like the speed, flexibility, or sophistication typical of humans. I would also suggest that symbolic reference makes an important contribution to sustaining and expanding the joint attentional scenes that mark humans’ linguistic interactions, for this sort of reference makes it possible to refer to things that are not physically present within the current attentional scene.

Music does not, except in rare instances, make use of symbolic reference. Instead, it makes use of what I shall call analogical reference, and which I shall discuss in more detail in chapter 2. Analogical reference obtains when a token shares structural features with some other entity or phenomenon. To the extent that these shared features are noted, the token can be said to refer to the other entity or phenomenon. For example, musicians have long used two-note musical figures that fall in pitch, and that proceed from a part of the rhythmic framework that carries emphasis to one that does not, as analogs for a sigh (Monelle 2000, 66–73). In truth, actual unvoiced sighs do not sound much like this (there is simply an exhalation of breath, without any discernible pitch content or requirement for a rhythmic framework), but this does not much matter: the purpose of the musical figure is to provide not a simulacrum of the sound of a sigh but an easily apprehended analog for the drop of the shoulders and release of physical tension that are typical of sighing.
In most cases, the tokens of analogical reference do not connect with one another (although they may succeed or be juxtaposed with one another), and systems of analogical reference rarely achieve the sophistication and flexibility that typify systems of symbolic representation. (An exception may be the “home signs” created by the deaf children of hearing parents, which have been studied by Susan Goldin-Meadow and her colleagues [2003a, chap. 14; 2003b].) When comparing these systems of reference, however, it is important to keep two things in mind. First, each system has its advantages and disadvantages. Because practically anything can serve as a symbolic token, such tokens are fairly easy to produce; the dense systems of reference of which such tokens are a part, however, require considerable effort to establish and learn. By comparison, analogical tokens require rather more effort to produce (to ensure that they have the structural features they share with their referents), but make far fewer demands on interpretation. There is also a contrast between the ways in which each system of reference tends to be used. Symbolic tokens are very useful for picking out objects and events, and characterizing relationships between them. Analogical tokens are less useful in this regard, but they are very good at representing complex spatial relationships and summoning dynamic processes that unfold over time. The second thing that should be kept in mind about these systems of reference is that, while language most typically relies on symbolic reference and music on analogical reference, the employment of either system need not be exclusive. Language can quite readily make use of analogical reference (as onomatopoeia and, more generally, work on sound symbolism [Hinton, Nichols, and Ohala 1994; Anderson 1998] suggests), and one could argue that the use of musical topoi by Wolfgang Amadeus Mozart and others (Ratner 1980; Allanbrook 1983; Agawu 1991; Mirka 2014) constitutes a form of symbolic reference.

A better sense of the resources offered by each system of reference, as well as insight into the different roles played by language and music, can be had by exploring what happens when words and music are used within the same utterance. Example 1.1 provides the score for the vocal part of the fourth movement recitative of Johann Sebastian Bach’s cantata for solo voice, “Ich habe genug” (“It is enough,” BWV 82), written for the Feast of the Purification of the Virgin Mary. The focus of the text for the cantata, by an unknown author, is not on Mary’s prescribed ritual of purification but on the fulfillment of a prophecy involving a man named Simeon, who lived in Jerusalem at this time. As described in Luke 2:25–32, it had been prophesied to Simeon that he would not die until he had seen the Savior. When Jesus was brought out of the temple after the ceremony, Simeon took the child into his arms and said, “Lord, now lettest thou thy servant depart in peace, according to thy word; for mine eyes have seen thy salvation.” Each movement of Bach’s cantata explores a different aspect of this scene. The fourth movement recitative concentrates on Simeon’s impatient anticipation of the fulfillment of the prophecy, with the knowledge that the arrival of the Savior will allow him to join God in the afterlife.

The first line of text, which will be my focus in what follows, finds an impatient Simeon speaking directly to God and asking for the peace he has been denied. Some of the urgency of his request is apparent from the two exclamations and cluster of monosyllabic words, but the bulk of the meaning comes through the tools provided by symbolic reference, which construct a joint attentional scene within which an emphatic question is being asked. In his musical setting Bach fleshes out the scene
considerably, not only shaping the delivery of Simeon’s words but adding analogical tokens that suggest Simeon’s complex emotional state. Bach’s rhythmic notation (which, following the performance practice of recitative, must be understood to provide relatively approximate indications of durational relationships) suggests that the delivery of the opening two syllables of the line (“Mein Gott!,” which spans three eighth notes and an eighth rest) should take about as long as the delivery of the concluding six syllables (“wenn kömmt das schöne: Nun!,” which spans four eighth notes and a sixteenth-rest). The latter portion of the line thus bursts forth in a rush, a rapidity of utterance typically associated with urgency. Simeon’s emotional state is further summoned through pitch: in the initial invocation of God the voice falls a diminished fifth, a melodic gesture that makes an apt analog for a despairing sigh. The line continues with an ascent, but one that overshoots its goal (the A♮4 and B♮4 would typically proceed to C5) to end on what will turn out to be the highest note of this brief movement, E♭5. Combined with the rapid declamation of the words, the melody for the conclusion of the first line creates an analog for a charged, even anxious, emotional state markedly different from that suggested by the setting of “Mein Gott!”

Throughout the vocal part for the recitative the words set and elaborate the scene, and the music supplies sequences of sonic events that serve as analogs for various dynamic processes. (These are described in more detail in a further analysis of this recitative in chapter 3.) Among the dynamic processes that music seems particularly apt for analogizing are those associated with emotions, and my analysis certainly exploits this correlation. Note, however, that the dynamic processes for which music furnishes sonic analogs could also involve various movements of the body: Simeon’s shoulders drooping, his chin falling as he cries “Mein Gott!”; the old man drawing himself up and casting his eyes toward heaven with “wenn kömmt das schöne: Nun!” There is, of course, a quite close connection between emotional states and physical dispositions—there is, for instance, evidence that an emotional state can be induced simply by assuming the physical disposition with which it is typically associated (Ekman 1992b)—and a sonic analog need not be viewed as

Example 1.1 Vocal part for the fourth movement recitative of Johann Sebastian Bach’s cantata “Ich habe genug” (BWV 82). Text translation: “My God! When will the lovely ‘now!’ come: / When I will journey into peace / and into the cool soil of earth, / And there, near You, rest in Your lap? / My farewells are made, / World, good night!” (Bach 2003)
making exclusive reference to either one or the other. The important point is that analogical reference can accommodate both possibilities: sonic analogs can refer equally well to the dynamic processes of emotions and of physical actions.

An instructive example of a sonic analog correlated with physical actions is provided by one of the Akpafu funeral dirges discussed by Kofi Agawu a number of years ago. The Akpafu are an ethnic group who live in the Volta region of Ghana, and at the time of Agawu’s study numbered about four thousand (Agawu 1988, 76). Most Akpafu dirges are relatively free-flowing—the singing is shaped more by the delivery of the words than by metric structure, and the dirge can include speaking as well as singing. At some points in the funeral service the singing may be accompanied by rapid back-and-forth pacing, but this movement does not typically conform to any metric pattern. In the case of the dirge “Mi ledza so” (“Get yourselves ready”), however, performance practice includes a rhythmized mourning walk shaped by the metric structure of the music (Agawu 1988, 96). As shown in example 1.2, which is adapted from an example provided by Agawu, the singers’ footsteps correlate with the duple meter of the dirge.

Example 1.2 The opening section of the Akpafu funeral dirge “Mi ledza so” with correlated footsteps. Text translation: “Get yourselves ready, death does not ask a person before he takes him!” (Agawu 1988, 103)
A straightforward account of the relationship between this movement and the music would see the recurring pattern of musical materials basic to the meter as a sonic analog for the singers’ footsteps, but the melody of “Mi ledza so” complicates this interpretation. Although quite a few of the footfalls coincide with a note of the melody, there are a number of instances where the footfalls and the melody are independent of one another. In measures 4 and 5, for instance, the singer sustains her F4 through beat 2 of measure 4 and beat 1 of measure 5. Similar sustained notes occur in measures 8–9 and measures 12–13, and in measure 16 the singer’s repeated C4s actually create a triplet cross-rhythm against the duple pattern of the footfalls. The melody does not, then, provide a thoroughgoing sonic analog for the physical actions of the mourners. What it does provide, however, is the basis for constructing a metric frame through regularly occurring melodic figures—in particular, what are notated in example 1.2 as triplet figures involving an eighth note and a quarter note—whose beginnings are separated by a consistent durational length. For instance, starting in the second half of measure 5 and continuing through measure 7 the notated durations of the melody are triplet eighth + quarter; even eighths; triplet eighth + quarter; triplet eighth + quarter; and triplet quarter + eighth. Each of these figures has the duration of a regular (that is, non-tripleted) quarter note. When the note that begins measure 8 lasts longer than this
it is fairly easy to project (in the sense developed in Hasty 1997) where the next quarter note (and the one after that) would have begun, even without support from the melody. The metric frame set up by this succession of durational units (and projections from such units), reinforced by the large-scale patterns created by the words for the dirge, creates a sonic analog for the footfalls of the mourners. The analog is supported by a network of correlations between the musical materials and the physical act of pacing: the sound, sight, and feeling of pacing all find analogs in the regular durational patterns set out by the music.

While this interpretation of the relationship between the music and the mourners’ rhythmized walk follows from Agawu’s account of “Mi ledza so,” it could be argued that it gives too much priority to the music of the dirge, and that the focus should instead be on the bodily movements of the mourners. It is these movements, then, rather than whatever patterns might be discerned in the melody or projections from its durations, that are the basis for our sense of metric regularity. Deciding which should be granted priority—the regularized pacing of the mourners, or its musical representation—is perhaps less important than acknowledging the intimate relationship between the bodily movements associated with the ritual and the organization of music. This relationship has significance not only for the study of musical grammar but also for understanding the place of music in human culture, in that the ritualized actions of the Akpafu funeral can be seen as one of the ways in which social memory is embodied or, using Paul Connerton’s term, incorporated (1989, chap. 3). Viewed from this perspective, physical actions like ritual pacing—learned through observation and practice, and performed at any number of previous ceremonies—enact the commemoration basic to an Akpafu funeral. Music too has its part in incorporating commemoration, but in two slightly different ways. First, incorporation must be understood to extend to the physical actions of the mourners as they sing the dirge, actions that range from the basic mechanics of vocal production to such stereotypical gestures as might accompany the singing. Second, the sonic analogs basic to the dirge can provide representations of bodily movements associated with mourning even when such movements are totally absent in performance. While this latter sort of incorporation may seem to be fairly abstract (since it need not correlate with any actual motion, either by the mourners or by the listener), recent research has suggested that embodied knowledge of this type is key to the creation and maintenance of conceptual knowledge (Barsalou 1999, 2008; Gibbs 2006).

In summary, symbolic reference makes possible a species of communication distinguished by rich contextual scenes. The opening words of Bach’s recitative create a small theater that includes an attendant and personal deity (to whom the “My God!” is addressed) and an absent agent (actually, an event in the form of a temporal moment—“the lovely ‘now’”) whose arrival is anticipated. Similarly, the words of the Akpafu dirge set a scene in which an audience is warned of the presence of an agent (actually, the biological process of death) who will take action without observing the niceties of social discourse. Analogical reference, by contrast, makes possible a species of communication distinguished by images—both seen and felt—of signal immediacy. The melody in the first few measures of Bach’s recitative summons a variety of emotional states, including despair and anxious anticipation,
as well as the physical dispositions and actions that might accompany such states. Similarly, the melody of “Mi ledza so” summons—perhaps even regulates—a rhythmized mourning walk and, given the comparatively rapid tempo of the dirge, an agitated emotional state appropriate to an almost frantic sense of loss.

The different communicative resources offered by these two forms of reference, and the near-exclusive employment of one by language and the other by music, suggest that the basic function of music is fundamentally different from that of language. Where language provides the means to manipulate the attention of conspecifics within the context of a joint attentional scene, music provides the means to create a context for the shared experience of dynamic processes. Although such contexts for shared experience can play a number of roles in human culture, one of the most important is within commemorative ceremonies. The liturgy of which Bach’s cantata was a part and the Akpafu funeral rites that include “Mi ledza so” are examples of such commemorative ceremonies. Within these, music provides a way to incorporate commemoration through a deliberate shaping of the experience of the participants.

My proposal, then, is that the basic function of music in human culture is to represent through patterned sound various dynamic processes that are common in human experience. Chief among these dynamic processes are those associated with the emotions and the movements of bodies—including our own—through space. Music does this through sonic analogs for dynamic processes which, on the account I shall develop here, are central to musical grammar.

Alternative Accounts

I should note, in passing, that there have recently been a number of accounts of the relationship between linguistic and musical organization that represent an alternative to what I offer here, including monographs by Joseph Swain (1997) and Aniruddh Patel (2008), and essays included in collections edited by Patrick Rebuschat, Martin Rohrmeier, John A. Hawkins, and Ian Cross (2012), by Nicholas Bannan (2012), and by Michael Arbib (2013). In general, these accounts view our basic understanding of both linguistic and musical structure to be relatively settled, with the former collecting around a model based on Noam Chomsky’s work, and the latter collecting around the musical practice of late eighteenth-century Europe. Again, it is not my purpose here to weigh in on the nature and organization of linguistic structure, that being well beyond my expertise. I do hope, however, to unsettle at least a bit our understanding of musical structure. I shall aim to do so through a careful exploration of two basic questions: why it is that humans have developed the expressive means offered by patterned nonlinguistic sound (the “function” of musical utterances), and how patterned nonlinguistic sound is organized to achieve these expressive ends (the “form” musical utterances take). I shall not, in truth, stray very far from the musical practices of late eighteenth-century Europe, but I believe that the foundation for musical grammar I shall sketch will provide new insights not only into the effectiveness of those practices, but also into many other musical practices past and present.
Foundations of Musical Grammar

Basic Features of Musical Grammar

The approach to musical grammar that will take shape over the following chapters takes as its basic model the grammars developed by cognitive linguists over the past thirty years, in particular those of Ronald Langacker (1987, 1992) and William Croft (2001). As such, it is conceived of as a species of construction grammar, in which all grammatical elements, from the smallest to the largest, are combinations of form and function. The overall function of the grammatical elements of music is to provide sonic analogs for dynamic processes; the forms grammatical elements take reflect the means through which sounds are organized to create these analogs. Construction grammars are based on the way communicative media are used rather than on formalized preconceptual designs. With respect to linguistic grammar, Langacker has observed that “putting together novel expressions is something that speakers do, not grammars. It is a problem-solving activity that demands a constructive effort and occurs when linguistic convention is put to use in specific circumstances” (1987, 65). In the practices of many musical cultures, the problem-solving activity of constructing utterances is understood to be the province of a relatively restricted subsection of the population (within the Western tradition, these individuals are called “composers”), but this activity can also be seen in the songs a child sings to her doll (when these are not simply those that her mother has sung to her), in the performances of highly skilled improvisers, and in the musical utterances of many of those who fall in between.

Given the different functions of language and music in human cultures, and an approach to grammar in which grammatical elements combine form and function, it stands to reason that the grammar of language and the grammar of music will be quite different from one another. On the account I should like to develop here, musical grammar has three features that distinguish it from linguistic grammar. First, the basic elements of musical grammar are holistic structures that combine and temporally order various sonic events to create analogs for dynamic processes. While it is possible to isolate and analyze individual sonic events—something that those interested in developing systematic theoretical accounts of musical organization have done since antiquity, and which has yielded constructs like “interval,” “note,” and “chord”—I take the view that musical practice centers on the production of combinations of sonic events to create analogs for dynamic processes rather than on the application of abstract grammatical rules to concatenate sound-event primitives. Second, the basic elements of musical grammar are organized through syntactic processes that are themselves analogs for dynamic processes. For example, measures 15–16 of “Mi ledza so” bring to a conclusion the first portion of the dirge, marked through the use of a rhythmic figure (notated as triplet quarter notes) used nowhere else in the piece. This unique rhythmic figure together with the recurrence of C4 (on which previous phrases have often concluded) exemplify the compositional strategy of cadence, which serves as a sonic analog for the process of arrival (further marked in the text by the exclamation “Bo ade!,” which might be loosely translated as “folks”—that is, “death does not ask a
person before he takes him, folks!”—and which, as a phatic sign, emphasizes the sharing of attitudes and emotions that is basic to the dirge). Third, musical utterances typically involve multiple syntactic layers. A glimpse of this can be seen in the performance practice of “Mi ledza so”: the sound of the footfalls of the rhythmized mourning walk represent a syntactic layer distinct from, but coordinated with, that of the melody of the dirge. To construe the sound of these footfalls as “musical” is, of course, something of a stretch in that the production of sound does not appear to be the principal motivation for the rhythmized mourning walk. That said, the combination of these two separate layers of sound demonstrates something that often happens in a wide range of musical practices, in which multiple syntactic strands are woven into a single expressive utterance.

**Chapter Outline**

Adopting the approach taken in my previous work, this volume uses research on the cognitive capacities of humans to ground an account of musical organization. The overall perspective, then, is decidedly interdisciplinary: while the main issues engaged all fall within the domain of the humanities (and, more specifically, musicology), the framework for exploring these issues draws upon work from empirical science. That said, the weight given to each of these disciplinary focuses shifts over the course of the book. Where chapters 2 and 3, for instance, are firmly grounded in empirical research, chapter 4 presents a somewhat more mixed approach: empirical research still informs the argument, but more emphasis is placed on using this research to build out my account of musical grammar. In chapters 5 and 6 I endeavor to expand this account through the interpretation of both musical practice and musical works, and, in doing so, move away from close discussion of the empirical bases for musical grammar.

**Chapter 2: Music and Analogy.** Chapter 2 begins with an aspect of human cognition that is of central importance to the account of musical grammar I wish to develop: humans’ capacity for analogical thought. As I suggested in my brief discussion of eighteenth-century ideas about mimesis, on the best evidence we currently have, the capacity for analogical thought demonstrated by humans exceeds that of any other species and is key to the distinctive character of human thought. In the first portion of chapter 2 I review recent research on analogy and explain how analogical thought shapes our understanding of music. Although I am by no means the first musicologist to show the part analogy plays in thinking about, and in, music—prior work includes studies by Marianne Kielian-Gilbert (1990), Zohar Eitan and Roni Granot (2007), and Amatzia Bar-Yosef (2007)—I believe mine to be the first account that draws on the extensive research done on analogical processes over the past three decades and applies it to music. Although research on analogy sheds new light on the cognitive capacities of humans, as a whole it tends to adopt a somewhat disembodied perspective: making analogies seems to be something that machines could do just as well as humans. As a corrective, I draw on Lawrence Barsalou’s theory of the perceptual basis for conceptual knowledge (first set out in detail in Barsalou 1999) to explain how
embodied experience informs analogical thought, especially that associated with music. The perspective on analogy that I develop informs my engagement with and expansion of Peirce's concept of iconicity, which I use as the basis for my notion of analogical reference. My adaptation of Peirce's theories sets the stage for a more rigorous definition of sonic analogs for dynamic processes, and also leads toward a broader view of how meaning can be constructed through sequences of musical sound.

As should be apparent, the approach I take to grounding, and building out, an account of musical grammar is somewhat novel. That said, I am not the first to suggest that something like sonic analogs for dynamic processes are key to musical communication: in recent years at least three scholars, each of whom was confronting a rather different challenge to our ideas about musical understanding, have proposed similar constructs. The first of these was Richard Middleton who, midway through his pathbreaking *Studying Popular Music*, sought a characterization for the schematic structures (which he called “forms”) around which much popular music was organized. He proposed thinking of them as what he called “gestures,” by which he meant “‘cognitive shapes’ (with affective and motor connections) which lie behind musical structures and ‘explain’ different transformations of the same idea” (1990, 201). In subsequent work Middleton made clear that he thought of these “gestures” as highly dynamic structures that were manifested in a broad range of music, that potentially comprised a wide array of musical features, and—perhaps most importantly for the perspective I wish to develop here—that connected accounts of music’s formal structures with the way it comes to have meaning (Middleton 1993).

Philip Tagg developed a rather more specific correlate for what I call a sonic analog in the course of fashioning an account of the role music plays within multimedia contexts. Tagg called this construct an anaphone, and he conceived of it as a musical sound sequence that was modeled on paramusical sound, movement, or sense of touch (1992, 372–73). In later work he generalized the concept a bit further and connected it with Peirce’s notion of an icon: “If analogy means another way of saying the same thing, anaphone just means using an existing model outside music to produce musical sounds resembling that model. An anaphone is basically an iconic sign type and can be sonic and/or kinetic and/or tactile” (Tagg 2005, 21). This concept then allowed Tagg to describe what he called “gestural interconversion,” which relied on anaphonic connections between music and other phenomena. More generally, gestural interconversion meant “the two-way transfer via a communality of gesture between, on the one hand, particular emotions or feelings which are both subjective and internal, and, on the other hand, particular external objects (animate or inanimate) in the material world” (Tagg 2005, 22). Although the notion of gesture used by Tagg remained at a relatively intuitive level—his examples range from actual physical gestures to imagined visual tracings—the concept of an anaphone together with the idea of gestural interconversion made it possible for him to describe how sequences of musical sound could, within multimedia contexts, become deeply connected with everything from naturally occurring sounds (such as the revving of a motorcycle engine), to visual images, to subjective emotions.
Perhaps the least formalized but closest correlate for a sonic analog was suggested by the psychologist John Sloboda. In an essay on musical meaning first published in 1998, Sloboda noted the failure of highly structuralist accounts of musical organization to capture what was meaningful in music, and advocated for an embodied account of how we understand music. To activate the idea of such an account, he offered the following scenario:

Imagine, for instance, that I come to a swinging door in a corridor. I push it, and it doesn’t yield. I push again, a little harder. Still it doesn’t quite yield, but I sense that a third push will do it. I push for a third time, harder still, and the door suddenly swings open, propelling me through to the next corridor. This is just one of many concrete examples of a dynamic process which involves repeated application of increasing force and a sudden release of whatever was holding the obstacle in place. (Sloboda 2005, 167)

Sloboda then brought this dynamic process into correspondence with the melody set out in the opening three measures of the first movement of Mozart’s Symphony in G Minor (K. 550): the melody begins with three repetitions of the succession E♭5–D5–D5 (“pushing on the stuck door”) that finally release to a high B♭5 (“the door swinging open”). The result of this analysis is an almost visceral sense of the effect of Mozart’s melody, one made more vivid by the way the rapid succession of melodic pitches compresses the image of pushing against a stuck door. Sloboda concluded this portion of his essay with the suggestion that one of the things that must be done if we are to explain how we understand music is to analyze and describe a wide range of analogies that are available to musicians, a proposal that anticipates one of the principal aims of this book.

While Middleton, Tagg, and Sloboda each pointed to the importance of analogical thinking for our understanding of music and proposed basic musical structures that were to some extent dynamic, none of them developed further the theoretical framework for the various correlates for sonic analogs that they proposed. This is hardly surprising: analogical constructs served as heuristics within, rather than the focus of, their accounts of musical thought. It bears mention, however, that each of them was concerned with two larger issues that will be explored in the course of the discussion of music and analogy undertaken in chapter 2: musical meaning, and the way embodied experience shapes thought. I shall defer comments on the latter until chapter 2, but with respect to the former it will be helpful if I offer a few words on the approach to musical meaning I take in this volume.

Explorations of musical meaning often involve the application of semiotic theories to music, something quite clear in Tagg’s formulation of the anaphone and more fully developed in the work of scholars like Kofi Agawu (1991, 2009), Naomi Cumming (2000), Robert Hatten (1994, 2004), David Lidov (2005), Raymond Monelle (2000, 2006), and Eero Tarasti (1994, 2002, 2012). The approach to musical meaning I shall elaborate here—one that is broadly consonant with the ideas of musical meaning deployed by Middleton and Sloboda—is of a rather more basic sort and focuses on the function of a musical utterance within a given cultural
context. Consider, for instance, the different kinds of meaning Richard Rodgers and Oscar Hammerstein’s “Oh, What a Beautiful Morning” would have in various cultural contexts: it would have one kind of meaning as part of the musical *Oklahoma!* (for which it was written), another if it were played as a waltz at a wedding reception, another if a recording of it were played when a caller was put on hold, and yet another if the incipit of the tune were used as a ringtone. Within each context the function of the tune would be different (ranging from its role within the story of *Oklahoma!* to music for dancing, to sounds to fill the emptiness of a technologically-induced limbo, to a means of getting one’s attention) and so, I should like to argue, would be its meaning. I would be the first to admit that this is a rather rudimentary notion of meaning, one that would not seem to lead naturally to the kind of highly detailed analyses offered by most music semioticians. That said, it has four distinct advantages for the project at hand. First, it incorporates cultural context from the outset, rather than adding it in as an afterthought. Second, placing the function of a musical utterance front and center conforms with the overall framework assumed by cognitive grammars, in which grammatical units are form-function pairs. Third, this notion of meaning begins with sequences of musical sound rather than with semiotic theories originally developed to account for language. And fourth, it is open to the influence of embodied experience on meaning, something for which Mark Turner argued more than twenty years ago (1994). In the case of the example I just offered, for instance, even if some of those in attendance at the wedding reception chose not to dance to “Oh, What a Beautiful Morning” they would understand that this particular sequence of sounds was intended to serve as a template onto which dancers can map their bodily movements. In situations like this, then, part of the meaning of “Oh, What a Beautiful Morning” derives from knowledge about coordinating bodily movement with patterned sound.

Chapter 3: Music and Emotion. In chapter 3 the theoretical framework developed in chapter 2 is used to explore the relationship between music and emotion. The closeness of this relationship has been remarked upon since antiquity—indeed, the power of music has often been attributed to its capacity to shape emotions—but exactly why this should be so has been somewhat less clear: music is, after all, but one of a number of cultural and social phenomena to which humans have strong emotional responses. In the first portion of the chapter I offer a summary of recent research on emotion, an exercise that provides a basis for evaluating claims about music’s capacity to shape emotions and suggests reasons why correlations between music and the emotions have been so enduring. In the second portion of the chapter I use the perspective provided by this summary to consider research on music and emotion that has been done over the past couple of decades. This consideration will lead to the proposal that one of the significant reasons that music has been linked so securely to the emotions is that music can provide sonic analogs for some of the most salient aspects of emotion processes (a proposal that echoes the account of the relationship between music and the emotions developed by Susanne Langer in the mid-twentieth century). As a way to illustrate this proposal, I shall take a close look at the movement that sets the stage for Simeon’s impassioned outburst in Bach’s “Ich habe genug,” the lovely aria “Schlummert ein”
Origins

T21 ("Slumber now"). My analysis of this aria will provide opportunities for a fuller introduction to the notions of syntactic processes and syntactic layers (along with the way syntactic layers interact with one another), and it will lead toward observations about the ways music is used to shape emotional responses within liturgical settings of the kind that motivated and framed Bach’s cantata.

Chapter 4: Music and Gesture. As emerged in Middleton’s and Tagg’s accounts of the bases of musical organization, the notion of a gesture is often used to refer to holistic sonic units that are analogous to the sort of spontaneous physical movement that often accompanies speech. Such a perspective led, for instance, to Robert Hatten’s detailed exploration of correlations between physical movements and the way Mozart, Beethoven, and Schubert organized the musical materials of their instrumental compositions (2004). In chapter 4 I pursue the notion of a relationship between music and gesture further, building on recent work by a number of researchers which suggests that the spontaneous movements that accompany speech are not simply an accidental accompaniment to the act of speaking but are instead evidence of thought processes that are in some measure independent from speech. The expanded view of communication that results (encompassing both language and gesture) and evidence that this sort of communication is unique to humans (despite the fact that other primates use gestures as part of their interactions with one another) will lead me to an exploration of the infrastructure for human communication, an infrastructure that I will suggest supports not only language and gesture but music as well. In the course of this exploration I shall be interested in how music and gesture are similar to one another, as well as how they are different (and not simply because one involves sound and the other physical movement). I will explore similarities between these two expressive modes through an analysis of the movements Fred Astaire made as he sang “The Way You Look Tonight” in the 1936 movie musical Swing Time, and differences through an analysis of an elegant gestural choreography that Charlie Chaplin performed to Johannes Brahms’s Hungarian Dance No. 5 in a scene from the 1941 film The Great Dictator. My discussion of the Hungarian Dance will offer a further illustration of syntactic processes and syntactic layers, leading toward the notion of referential frameworks, which serve as perceptual anchors for syntactic processes.

Chapter 5: Music and Dance. With the introduction of referential frameworks in chapter 4, the preliminary model of musical grammar that I develop in this volume will be largely complete. In chapter 5 I shall draw on this model to explore ways music and movement correlate with one another in two dance practices from Western Europe, drawn from the period before and after the French revolution. The first of these comes from France’s ancien régime and is exemplified by a bourrée initially used in the prologue to Jean-Baptiste Lully’s Achille et Polixène. Thanks to dance notations developed in the late seventeenth century we have relatively detailed knowledge of the steps dancers would perform to the music. I use this notation to offer a close analysis of the movements associated with a contemporary choreography of the bourrée and the way these movements correlate with the sonic analogs set out by the music. This analysis not only offers insights into the specific grammar of the bourrée but also into how dance practice contributed to the construction of social relations in the court of Louis XIV (something discussed
in more depth in Franko 1993, chap. 5, and McClary 2012, chap. 8). The second dance practice is associated with the waltz of early nineteenth-century Austria and Germany. In the wake of the enormous changes that swept Europe subsequent to the French revolution and Napoleonic wars, the waltz—as a dance that involved both intimate contact between and a rapid circulation of partners—claimed a prominent place in the social landscape. As a way to understand what the waltz offered to early nineteenth-century Europe I will again offer close analyses of the relationships between movement and music, but given the simplicity of the waltz step—which, in its most basic form, was a single pattern repeated indefinitely—my focus will shift to the ways composers adapted their music to the different social contexts within which the dance was performed. To this end I shall consider waltzes by Joseph Lanner, a composer and conductor of some fame within the dance culture of early nineteenth-century Vienna, and Franz Schubert. While Lanner’s waltzes were clearly intended for the capacious and ornate ballrooms of Vienna’s suburbs, Schubert’s are more varied and suggest the different roles waltz music could play.

Important to the perspective offered by chapter 5 is the notion that music makes a significant contribution to the social interactions of humans. This is hardly a new idea, being one for which Tia DeNora (2000) and others have advocated for a number of years, and yet the form it takes here is slightly different. Building on the presuppositions of musical construction grammar, in which grammatical elements are form-function pairs, my argument is that the social function of a musical utterance is realized in part through the way its musical materials are organized—that is, through its grammar. Thus a waltz written to succeed within—and, indeed, to help construct—an intimate social environment would not be one fitted to the boisterous activity of a crowded ballroom. Musical grammar is, accordingly, not simply about how to arrange sonic materials to create a “grammatically correct” utterance but about how these materials are arranged to realize a relatively specific social and cultural function, something demonstrated by the way the music of social dances is fitted to the steps and context of those dances. Put another way, the meaning of a given waltz or bourrée is in part constructed through its grammar, which facilitates (even if it does not absolutely determine) how it functions in social contexts.

Chapter 6: Music and Words. As noted earlier, according to the approach to grammar I take here, music and words offer different—indeed, often incommensurate—resources for the construction of meaning. Given, however, that both can be produced through the same physical modality—the voice—it is not surprising that they are often combined. In chapter 6 I explore such combinations and what they can tell us about the ways music and words construct meaning, both independently and together. Combinations of words and music encompass a potentially vast repertoire, which could include incantatory ritual practices of ancient origin, music-theatrical works of the early modern period, and different species of rap that developed in the late twentieth century, but in this chapter I focus on a single genre and a relatively isolated historical moment: the German Lied in the early nineteenth century. The reasons for this are three. First, as Yonatan Malin has recently shown (2010, chap. 1) composers of Lieder typically
Origins

23

gave close consideration to the poetry they set with the aim of finding a musical setting adequate to the demands of the text. Although Lied composers might still fall back on formulae or otherwise fail to fully exploit the expressive resources of music or words, their expectation was that a properly composed song offered more than did music or words alone. Second, the opening years of the nineteenth century were a period of rapid stylistic change in the Lied. Some of this had to do with an increased commercial market for Lieder (a genre ideally suited to performance in the homes of the emerging bourgeoisie), and some of it had to do with a growing sense that musical utterances could equal or exceed what could be expressed through language. And, third, the analyses I present here connect in a productive way with those I offered in previous work (Zbikowski 1999; 2002, chap. 6). One portion of that productivity derives from a more detailed account of the ways music shapes the delivery of words, another portion from restricting my study to three settings of one of the most well-known of German lyric poems: Johann Wolfgang von Goethe’s “Über allen Gipfeln ist Ruh.” The first setting of those I consider was done in 1814 by Goethe’s friend and confidante Carl Friedrich Zelter; the second was done around 1816 by Carl Loewe, a composer who later became famous for his ballads; and the third was completed sometime before 1824 by Franz Schubert. Through drawing on the model of musical grammar developed in this volume, and in particular on descriptions of the interaction of syntactic layers, I am able to describe how each composer’s music actively shapes—in a fully dynamic way—our understanding of Goethe’s text.

From a broader perspective, the analyses included in chapter 6 could be seen to realize, if only partially, one of the ambitions David Lewin described in the preface to his Studies in Music with Text. Lewin noted that the essays included in the volume—which were written over a period of some thirty-five years and dealt with a repertoire that spanned the eighteenth through twentieth centuries—embodied a consistent attitude toward what he called “texted European-American music.” Reflecting on this attitude, he observed, “I seem to be looking for ways in which music and text, in this repertory, enact each other. I like the term ‘enact’ better than ‘read’ or ‘allegorize,’ because the term emphasizes gestural—often even bodily—aspects of the interrelation” (Lewin 2006, xii). Although my approach to the analysis of relationships between music and words is rather different than that taken by Lewin (not least because I adopt a more explicitly cognitive approach), the notion that the basic elements of musical grammar are sonic analogs for dynamic processes and that these analogs are intimately bound up with embodied experience makes it possible to capture aspects of the idea of enactment with which Lewin was concerned. The mutual enactment of music and text can thus be seen to be a consequence of the combination of two different types of conceptual knowledge—one associated with music, the other with words—into the single expressive medium that is commonly called “song.”

Chapter 7: Questions, Answers, Questions. By way of conclusion, chapter 7 returns to speculations about the origins of music, and what such speculations might tell us about the role of music in human cultures and thus about how patterned nonlinguistic sounds come to be organized into the forms we call “music.” A case could be made that, in drafting the “Essay,” Rousseau was not particularly
interested in the origins of music as such but instead aimed to address the problem of the origins of language that had been raised in the second *Discourse*. By situating *la pitié* at the center of human communication, he provided a basis for the bonds of compassion that he believed were essential to social organization. And with the idea that a special form of human communication—a unified form of music and language—could serve as a vehicle for *la pitié*, he placed at the disposal of his first peoples the means to establish social order. A case could also be made, however, that the “Essay” was intended to be a contribution to the speculative anthropology practiced by a number of Enlightenment thinkers, a point of view for which Downing Thomas has argued (1995, chap. 4). There are, for instance, quite a few close correspondences between the primordial language described by Rousseau and the *langage d’action* proposed by Etienne Bonnot de Condillac in his *Essai sur l’origine des connoissances humaines* (Condillac 1746). Both modes of communication were assumed to provide the missing link between the way animals communicate and the way humans communicate; both ultimately gave rise to music and language; and both were part of origin myths meant to address anxieties about how, in the absence of religious doctrine, humans could be distinguished from animals. Condillac and Rousseau each believed that through moving the passions of others, and doing so through stable systems of signification, humans were able to separate themselves from animals (Thomas 1995, 56). Modern music and modern language were evidence of the success of this separation and also, perhaps, of its fragility.

It bears mention that when there was a renewed interest in the origins of music toward the end of the twentieth century, the issues—and the anxieties they belied—were somewhat different. Advances in neurophysiology, along with developing resources for brain imaging and the emerging disciplines of cognitive science, led a number of scholars to speculate about the role of musical behavior in human evolution. Some—including, most famously, Stephen Pinker (1997, 528–38)—saw little practical use for music; others strove vigorously to fit music into the evolutionary picture, believing it to be a distinctive aspect of the human species. (For a review of these approaches, see Cross 2007.) Underlying much of the speculation on musical origins was a relatively new anxiety about the value of music: in an age of increased commodification and waning interest in “the classics” many believed that music was itself under siege (J. Johnson 2002), a concern evident in a number of books aimed at explaining, to a general audience, the importance of music to human thought (Levitin 2006; Ball 2010; Lehmann 2010; Schulkin 2013).

As an instrumentality to situate speculations on the origins of music from the Enlightenment and of more recent vintage, chapter 7 begins midway between the two with Charles Darwin’s thoughts about why music developed in the human species. Although Darwin’s perspective was shaped at a time when musical practice was not thought to be in danger, there was nonetheless the question of what advantage, if any, musical behavior gave to humans in their struggle for survival. My summary of Darwin’s thoughts—which led him to conclusions about the origins of music and language remarkably similar to those of Rousseau—will provide a framework for a review of the arguments set out in this volume and offer an
opportunity for reflections on theoretical and empirical work that might build on the account of the foundations of musical grammar I offer here.

**Musical Grammar and Music Theory**

My aim in this volume is to develop a theoretical account of musical organization informed by work on construction grammars and shaped by an argument about the role of music in human cultures. The results of these efforts will be a description of what I believe to be the essential features of a cognitive grammar of music. That said—and as should already be apparent—the account I shall give of musical organization does not look much like the music theory that can be seen in journals like *Music Theory Spectrum*, *Music Theory Online*, or the *Journal of Music Theory*, or, for that matter, in the other volumes in this series. The reasons for this are two. First, my intent is to capture, with as much fidelity as possible, the cognitive and material resources humans have recruited to create the expressive medium conventionally called “music.” To do so I will need to take a few steps back from familiar theoretical constructs—things like “intervals,” “notes,” and “chords,” not to mention more complex notions like “harmony” or “tonality”—to get a clearer view of why such constructs have proven to be useful tools for describing the results of certain kinds of musical behavior (results often called “works” or “compositions” within the Western tradition). The second reason for the unconventional cast of the theory I develop here is that much of what I offer is preliminary. Although I have endeavored to be as clear as possible in my characterization of the elements of musical grammar, and to draw these together periodically to provide as consistent a view of musical organization as possible, much remains to be discovered. My hope is that the following chapters will provide an apt framework for this discovery and will motivate further thought about the nature and structure of musical organization.
Two

Music and Analogy

Example 2.1 gives the opening eighteen measures of a brief composition for guitar solo by the Argentinian teacher and performer Julio Salvador Sagreras (1879–1942). The work, while thoroughly idiomatic for the guitar, has features that make it sound a bit odd, as if it were poised midway between a study and a showpiece. On the didactic side, there are the nearly obsessive focus on repeated notes; somewhat idiosyncratic fingerings (the opening concert B2 and E3 of the melody are much more easily played on adjacent strings than on a single string, as specified by Sagreras); and repetitions of blocks of material (such as the reprise of measures 9–10 in measures 11–12). On the showpiece side, there are the breathtakingly fast tempo indication (which, if followed, would yield a relentless eight notes per second); the rapid shifts of the left hand (the ascending and descending shifts in measures 13–14 require the performer to move her left hand—with pinpoint accuracy—22 centimeters in less than an eighth of a second); and the flashy, if technically straightforward, use of campanellas technique in measures 4–7 (in which open, or unstopped, strings—here, those used for B3, G3, and E4—sound against higher-pitched notes on stopped strings). These features play out within a somewhat fragmented compositional style: the repeated notes and scale of measures 1–3 lead directly to the campanellas of measures 4–7, which are followed by the brief harmonic passage of measures 8–12 that redirects tonal focus toward B, and which is itself followed by a return to the repeated-note motif of measures 13–16. The result of the whole is a somewhat odd instrumental work, one that seems to be trying to make up for its lack of compositional coherence or didactic consistency with sheer virtuosic intensity.

As is often the case with musical works that have unusual features, there is a story behind Sagreras’s composition, one told principally by the title and subtitle for the work: “El Colibri, Imitación al vuelo del picaflor” (“The Hummingbird, Imitation of the flight of the hummingbird”; the redundancy in the title and subtitle reflect South American names for the hummingbird with different etymologies: “colibri” derives from indigenous languages, whereas “picaflor” derives from Spanish). Once the title is known, features that seemed odd quickly organize themselves into a coherent image. The rapid repeated notes, beginning in a low register, evoke the surprisingly visceral sound the hummingbird makes in flight; the sudden shifts along the length of the string (which produce brief but memorable portamentos) summon the darting movements of the bird; the larger blocks of musical material (measures 1–3, measures 4–7, measures 8–12, and measures 13–16) map out its path in search of flowers on which to feed; and the accented non-chord tones of measures 9–12 capture the small motions it makes
as it dips to drink. What Sagreras has given us, then, is a rather detailed sonic analog for a group of conjoined and highly dynamic processes associated with the flight of the hummingbird. But why should these sounds serve as analogs for processes that are, for the most part, visual? Or, moving away from the linguistic prompts offered by the title, why is it easy for most humans to make a connection between Sagreras’s music and the experience of encountering a hummingbird in flight when a similar sort of connection is simply beyond the cognitive capacity of other animals?

These questions bring us to the principal topic of this chapter—and one of central importance to this volume—which is humans’ capacity for analogical thought. As I noted in chapter 1, it is not so much that the topic of analogy has been ignored in musical scholarship, it is that it has not been incorporated into a broader account of the cognitive capacities that shape musical understanding. Accordingly, the first section that follows offers a review of recent findings from cognitive science on analogy, with an emphasis on its contributions to the distinctiveness of human intelligence. One thing that this review will make apparent is the rather disembodied approach that has been taken to analogy: while analogy might be central to human thought processes, the guiding assumption of much research is that making analogies is something that machines can do equally well.

In the second section that follows, I endeavor to provide a corrective—one that I would argue is essential to understanding musical thought—by using Lawrence Barsalou’s theory of perceptual symbol systems to explain how analogical thought can be grounded in embodied experience. Barsalou’s theory is also key to explaining why musical imitations of natural sounds (of the sort that were important for eighteenth-century theories of musical mimesis) range from fair approximations to insubstantial sonic specters: such imitations need capture only aspects of the sounds or phenomena to which they refer, the rest being supplied by memory and imagination. The third section takes up the idea of analogical reference introduced in chapter 1, building on and expanding C. S. Peirce’s concept of iconicity. This adaptation of Peirce’s theories sets the stage for the more rigorous account of sonic analogs for dynamic processes undertaken in the fourth section, one that draws on musical examples from Sagreras’s “El Colibri” and puts them in dialog with research on cognitive processes related to categorization and memory. The penultimate section explores the connection between imaginative thought fueled by analogical processes and the way concepts from disparate domains can be blended together such that a sequence of musical sounds is able to support and elaborate the miniature narrative we might build around the notion of a hummingbird in flight. By way of conclusion, the last section summarizes the resources that sonic analogs for dynamic processes offer to music, describes features of some of the most basic elements of musical grammar, and considers the role that analogy plays in the construction of meaning.

**Analogy**

Most discussions of analogy begin with similarity, because it is the similarity of one situation to another that is the point of departure for any analogy. Similarity judgments, which are allied with processes of categorization, are a basic tool for reasoning. For instance, the first-century BCE writer Marcus Vitruvius Pollio shaped his account of how theaters should be sited and designed by noting similarities between the way sound is propagated and the way waves pass through water. Turning to the matter of the best location for a theater, Vitruvius writes,

> It is also important to note carefully that the site itself not deaden sound; it should be the type in which the voice may travel with the utmost clarity. This can be accomplished if a site is selected where resonances are not impeded. The voice is a flowing breath of air, and perceptible to the hearing by its touch. It moves by the endless formation of circles, just as endlessly expanding circles of waves are made in standing water if a stone is thrown into it. These travel outward from the center as far as they can, until some local constriction stands in the way, or some other obstacle that prevents the waves from completing their patterns. As soon as these obstacles interfere, the first waves bounce back and upset the patterns. In the same way the voice makes circular motions; however, on the surface of water the circles move horizontally, while the voice at once advances horizontally and mounts upward, step by step. For the voice, therefore, just as for the pattern of
waves in water, so long as no obstacle interferes with the first wave, it will not upset the second wave or any of those that follow; all of them will reach the ears of the spectators without echoing, those in the lowermost seats as well as those in the highest. (Vitruvius Pollio 1999, 66)

Vitruvius builds an analogy between air and water from the simple observation that both can flow from one place to another and thereby “touch” remote objects. Observing that water can accomplish this through waves emanating from a single source, spreading in a circular fashion, he reasons that air must do something similar. A number of inferences then follow: that waves of sound will continue to spread until they meet with some obstacle; that when the waves encounter such an obstacle they will reflect back and disturb the spreading pattern of waves; and that echoes are a consequence of this sort of reflection. Vitruvius also extends his inferences beyond the model provided by water, proposing that waves of sound emanate both horizontally and vertically, and that vertical waves spread in the same way that horizontal ones do.

Inferences of the sort generated by Vitruvius point to what many researchers regard as the defining characteristic of analogy: the mapping of systematic structural relationships between a source domain (for Vitruvius, the water contained in a pond) and a target domain (the air contained by the site for a theater) (Gentner 1983; Gentner and Kurtz 2006; Holyoak and Thagard 1995, chap. 2; Holyoak 2005). In such a mapping, elements are mapped to elements, relations to relations, and the correspondences between elements and relations within each domain are preserved (Gentner and Markman 1997, 47). In Vitruvius’s analogy, water is mapped onto air, the tossed stone onto the impulse of the voice, and the spreading waves of water onto the spreading waves of air. Most important, the relationships between the given medium, the physical action on it, and the result are preserved. Analogy is not simply about correlating elements from one domain with elements in another domain, but about mapping relationships between these domains. It is thus often described as concerned with relations among relations (or “second-order” relations): Vitruvius’s analogy correlates the relationship of spreading circular wave to water with the relationship of spreading circular wave to air. One consequence of this mapping is that the notion of a wave gets turned into an abstraction, one that applies equally well to both water and air (and that would, in the nineteenth century, be applied to electromagnetic radiation).

It is worth noting that Vitruvius’s analogy is shaped by his goals. There are any number of similarities between water and air—both can be put in a closed container and moved from place to place; both are necessary for life; both remain unscathed if poked with a stick; both almost certainly belong to the Emperor—but Vitruvius focuses on just those features and relations that are relevant to his discussion of the acoustic properties of theaters. The alignment of features and structure that typifies analogy is thus constrained by contextual goals that are distinct from the analogical process proper (Holyoak and Thagard 1995, chap. 1; Medin, Goldstone, and Gentner 1993).

Making analogies is something that is virtually effortless for humans. Motivated by this fact, Douglas Hofstadter has argued that analogy, as the means by which
concepts are assembled and connected to one another, is at the very core of human cognition (Hofstadter 2001; see also Hofstadter and Sander 2013). At the very least, there is considerable overlap between judgments of similarity, making analogies, and processes of categorization, all of which contribute to the distinctiveness of human intelligence (Medin, Goldstone, and Gentner 1993; Glucksberg and Keysar 1990). Perhaps more striking is that the capacity for analogy is apparently unique to our species. Although other species are able to make some very sophisticated similarity judgments—there is research suggesting that chimpanzees can understand the second-order relations basic to analogy (especially for spatial reasoning) and that bottlenosed dolphins can perform sophisticated body-mapping analogies (Call and Tomasello 2005; Herman 2002)—current evidence indicates that no other species comes close to making or using analogies with the facility and speed of humans (Gentner 2003; Holyoak and Thagard 1995, chap. 3; Oden, Thompson, and Premack 2001). And this capacity is available from a very early age: children as young as ten months are able to solve problems by analogy (Chen, Sanchez, and Campbell 1997), and by the age of three years analogical abilities are quite robust (Goswami 1992, 2001; Gentner 2003).

The ability to map systematic structural relationships between disparate domains bears witness to a capacity for abstract thought—for thinking about relations between relations—of enormous flexibility and wide application. Analogy has been recognized as a key factor in human creativity (Hofstadter and the Fluid Analogies Research Group 1995; Donald 1999, 143; Fauconnier and Turner 2002, chap. 1) and has been linked to the conceptual flights of fancy and processes of meaning construction created through metaphor and metonymy (Gentner et al. 2001; Glucksberg and Keysar 1990; Glucksberg, McGlone, and Manfredi 1997; Holyoak and Thagard 1995, chap. 9; Holyoak 2005), a connection I explore further toward the end of this chapter. As I have argued in previous work (Zbikowski 2002, chap. 2), the conceptual domains involved in such mappings need not be restricted to those involving language, an argument supported by the capacity for analogy demonstrated by primates and prelinguistic children. (It should be noted, however, that there is general agreement that language provides crucial support for the second-order relations typical of analogy; see Gentner 2003 and Call and Tomasello 2005.) The short answer to the question of how we can connect the musical events of Sagreras’s “El Colibri” and the flight of a hummingbird, then, is that we as humans can scarce do otherwise: analogy is involved in our most mundane observations as well as our most profound.

There is, however, a somewhat longer answer to this question that engages more fully with the highly dynamic process that is the subject of this analogical mapping, and with the use of music to make the analogy. Let’s begin with the hummingbird itself. The bird is a member of a remarkable species, one found only in the New World and notable for the diminutive size of most of its members. The species is distinguished by its unique mode of flight, which allows it to hover and to fly backward and thereby feed on the nectar of flowers. It was just this that Nehemiah Grew noted in a 1693 communication to Philosophical Transactions, generally acknowledged to be the first detailed description of the hummingbird: “They feed by thrusting their Bill and Tongue into the blossoms
of Trees, and so suck the sweet Juice of Honey from them; and when he sucks he sits not, but bears up his Body with a hovering Motion of his Wings” (1693, 761). These characteristic movements together with the sound the bird makes in flight (which Grew, for some unaccountable reason, was not able to hear) make up a unified image—with “image” here extended to include both visual and sonic phenomena—that we might label “the flight of the hummingbird.” Sagreras’s composition offers temporally distributed and parametrically varied musical materials to correspond with specific aspects of the bird’s flight: the continuous succession of sixteenth notes correlate with the constant motion of the hummingbird and the rapidity of its flight; the sequences of repeated notes correlate with the distinctive sound created by the beating of its wings as well as with the active stasis of its hovering; and the accented nonchord tones correlate with its dipping movements as it feeds. All of these elements, brought together by their rapid succession and supported by other musical materials that reinforce the sense of a unified structure (such as the brief scalar passage of measure 2), create a sonic analog for the flight of the hummingbird.

The numerous correlations between the musical materials of Sagreras’s little piece and the characteristic features of the hummingbird’s flight, together with humans’ natural propensity for analogical thought, suggest that drawing an analogy between the two is thoroughly natural. Yet it is by no means certain that a listener, without being prompted by Sagreras’s title and subtitle, would make an immediate connection between the piece and the flight of the bird. What is needed is knowledge of the relational frame basic to the analogy. As research with young children has shown (Goswami 2001), relational knowledge of this sort is crucial to making analogies: a listener who is familiar with the conventions of programmatic musical works (such as Antonio Vivaldi’s “Four Seasons” concerti or Hector Berlioz’s Symphonie fantastique) will know that musical events that are exceptional or remarkable often correlate with some sort of extramusical narrative. (It bears mention that most eighteenth-century accounts of musical mimesis are illustrated by musical examples that are connected to an extramusical narrative.) Confronted with the concatenation of unusual features that make up Sagreras’s piece, this same listener would most likely be predisposed to look for some sort of extramusical phenomena with which these features might be correlated. But without such knowledge, or prompts from other domains (such as visual images of a hummingbird in flight), Sagreras’s sonic analog will most likely fail. Sonic analogs always function within some larger context, which provides the relational frame for connecting sounds with concepts from other domains.

What is striking is that, despite the importance of contextual knowledge to the interpretation of sonic analogs, our typical experience of music is one marked by its immediacy. Such immediacy is oftentimes an illusion—even our most basic perceptions have proven to be heavily mediated by the various brain functions activated by sound, smell, or light—but in the case of music the illusion seems to be borne out by the cognitive transparency of many sonic analogs and by music’s independence from language. Sagreras’s “El Colibri” simply is an imitation of the flight of the hummingbird. Confronted with the success of this imitation, language
can do little more than act the part of the shamed child and stand dumbstruck, twisting its toe in the dirt.

A way to explain music’s apparent immediacy, its independence from language, and its potential for providing sonic analogs for dynamic processes—and one I shall pursue in the following—is to regard musical concepts as grounded in bodily experience. The idea of a deep and durable connection between the body and the mind is now at least three decades old (having received thoughtful formulations in M. Johnson 1987 and Varela, Thompson, and Rosch 1991), and empirical evidence supporting this idea has been accumulating steadily for the past two. What has, to a certain extent, been lacking has been a robust theoretical account of the way perceptual information—gathered as the body monitors itself and interacts with the world—is transformed into conceptual knowledge. This is, to my mind, supplied convincingly by Lawrence Barsalou’s theory of perceptual symbol systems (Barsalou 1999). In what follows I shall provide a brief review of Barsalou’s theory and show how it can be used to explain music’s immediacy and independence from language, as well as the basis for sonic analogs for dynamic processes.

### Perceptual Symbol Systems, Embodied Knowledge, and Analogy

**Barsalou’s Perceptual Symbol Systems Theory**

One of the principal motivations behind Barsalou’s theory of perceptual symbol systems was to find a way to explain how perceptual information shapes the cognitive representations that occupy our conscious thought. Prevailing theory held that, as part of the process of human thought, perceptual information was transduced into symbolic structures that were functionally amodal. The symbolic structures derived from audition, for instance, were not conceived to be different in kind from those derived from vision. The disjunction between perceptual mode and symbolic structure was seen as a strength of amodal symbol systems, since it linked such systems more securely to the arbitrary symbol systems typical of language. The account of cognitive representation that developed from this theory was powerful in its flexibility but hampered by a number of problems, not the least of which was a lack of empirical verification for the process of transduction or for the existence of amodal symbols. The alternative that Barsalou offered took as its point of departure the idea that perceptual information could give rise to structures with all of the computational features of amodal symbols but without sacrificing modal specificity.

Barsalou’s approach builds on work done in the neurosciences over the past twenty-five years, which has shown that the perception of a physical entity engages a number of coordinated feature detectors in sensorimotor areas that are relevant to a given perceptual mode (Barsalou 2005, 398). During the visual processing of a hummingbird, for instance, some neurons will fire for the edges of the bird’s figure,
others for the surfaces of its body, color, orientation, and direction of movement. Example 2.2a provides a schematic diagram of this diversity of visual processing. Similar distributions of activation would occur in other modalities, and they would be represented in feature maps specific to those modalities. These might represent the sound of the bird, the way it feels in the hand (both in terms of the touch of its feathers and its near weightlessness), and introspective states summoned on encountering the bird (such as the thrill of discovering it or trepidation at the sight of its sharp bill). Example 2.2a also shows that when a pattern becomes active in a feature map, conjunctive neurons in an association area capture the pattern’s features for later use (Barsalou 2005, 399). This feature of Barsalou’s account of the processing of perceptual information adopts the theory of convergence zones proposed by Antonio Damasio. For Damasio, as for Barsalou, perceptual information is first recorded in a fragmentary fashion. The neural records of these fragments are then brought together through the distributed neural structure of the convergence zone, which Damasio describes as “an amodal record of the combinatorial arrangements that bound the fragment records as they occurred in experience. There are convergence zones of different orders; for example, those that bind features into entities, and those that bind entities into events or sets of events, but all register combinations of components in terms of coincidence or sequence, in space and time” (1989, 26). Although convergence zones are not linked to any specific modality, their basic components consist in all cases of information gathered from perception.

Example 2.2 Illustration of the storage (a) and simulation (b) of sensorimotor information (adapted from Barsalou 2005, Figure 15.1; image of ruby-throated hummingbird by James McClelland, from Johnsgard 1997, Plate 13)
As illustrated in example 2.2b, the conjunctive neurons in the association area also support a sensory-motor reenactment of the original activation pattern. Barsalou sketches the process as follows: “Once a set of conjunctive neurons in a convergence zone captures an activation pattern in a feature map, the conjunctive neurons can later reactivate the pattern in the absence of bottom-up sensory stimulation. While remembering a perceived object, for example, conjunctive neurons reenact the sensorimotor states that were active while encoding it” (2005, 399). Such reenactments are, of necessity, only partial and will be tailored to the agent’s current context of action (Barsalou 2003). According to Barsalou’s theory, the concepts that populate our conscious thought rely on activations of this sort—that is, they have their origin in perceptual information.

Barsalou called the fragmentary records of neural activation captured in a feature map perceptual symbols, and he argued that cognitive operations that make use of perceptual symbols can do everything that amodal symbol systems can do: they can represent types and tokens, produce categorical inferences, combine the symbols to produce hierarchical propositions, and yield abstract concepts (Barsalou 1999). The key to this productivity was a cognitively distributed system that Barsalou called a simulator, which, through the neural reenactment of sensorimotor states, gave rise to concepts. The process involves successive iterations of the storing of sensorimotor information diagrammed in example 2.2a, which occur as further instances of the original stimulus are encountered. Each successive encounter—with different hummingbirds, or with the same bird under different circumstances—will activate similar states in the feature maps. Similar activations of the feature maps will be captured by similar populations of conjunctive neurons in the association areas, and over time an integrated, multimodal sensorimotor representation of the category will develop. For the category HUMMINGBIRD, visual information about the appearance of the birds is integrated with information about distinctive aspects of their flight, auditory information about how they sound, and introspective information associated with encounters with the birds. This creates a distributed system throughout the brain’s association and modality-specific areas that establishes the conceptual content for the category (Barsalou 2005, 400). This system, through the reenactment of the sensorimotor states associated with HUMMINGBIRD, makes possible a simulation of the features of a hummingbird even when no bird is present; in consequence, Barsalou calls it a simulator. Again, the reenactments of the activation of a pattern in a feature map are only partial, and so each simulation will involve only a small subset of the content of a simulator. (A potential source of confusion in this perspective is the relationship between categories and concepts. In conformance with a perspective that guides much work in cognitive science, Barsalou regards categories and concepts to be functionally equivalent. See also Barsalou 1993; Barsalou et al. 1993; Hampton and Dubois 1993; Hofstadter and Sander 2013; Mahon and Caramazza 2009; Medin, Goldstone, and Gentner 1993; and Murphy and Medin 1985.)

Simulators reflect our capacity for selective attention: when we focus on a particular bird or flower we are of necessity not focusing on other things in the environment. We can, of course, attend to specific properties of objects as well
as the objects themselves, giving rise to simulators for the TAIL or BILL of a bird or the STEM of a flower. And we can attend to relations between properties and objects: the relationship between the tail of the hummingbird and its bill is captured by the simulator for BELOW; the relationship between the blossom of a flower and its stem is captured by the simulator for ABOVE. Simulators for relationships typically involve objects or properties, and they have an implied focus—while the blossom of a flower is ABOVE its stem, the stem of the flower is BELOW the blossom.

In a like fashion, simulators for events—which are, after all, another aspect of our environment on which we can focus our attention—typically require the presence of object simulators, as well as property and relation simulators, and reflect a particular focus. An event such as THE FLIGHT OF A HUMMINGBIRD will involve simulators for the bird, the characteristic properties of its flight, and its relationships to other objects in the environment, all strung together in a sequence and typically focused on the bird (rather than, say, on the flowers it flies between). Simulating this event—that is, recalling it in memory—will involve not only partial simulations of the objects, properties, and relations involved in the event, but also a simulation of the distinctive temporal sequence that comprises all of the other simulations.

In his work on the relationship between abstractions and perceptual symbol systems, Barsalou emphasized that the activation of a simulator is a dynamic process. Simulations—whether of objects, properties, or relations—are created in real time in response to an agent’s current context of action (Barsalou 2005, 414). Simulations of events add another level of dynamism: simulating THE FLIGHT OF A HUMMINGBIRD will involve online simulations of all of the objects, properties, and relations basic to the event together with a simulation of the temporal sequence of the event itself. Again, such simulations are always partial—only a few key aspects of the event may be activated in the course of the simulation—but to the extent that such an event is conceived as an event (rather than an object) its simulation will proceed by successive stages disposed over some span of time.

**Perceptual Symbol Systems Theory and Sonic Analogs**

Two aspects of Barsalou’s theory of perceptual symbol systems are particularly important for our understanding of sonic analogs. First, simulators operate independently of the symbolic system of language. Neither the word “hummingbird” nor any part of the dense network of symbols within which this word is embedded are necessary to establish the simulator for the category HUMMINGBIRD. This approach to categorical knowledge is similar to that which I developed for music in earlier work (Zbikowski 2002, chap. 1). For instance, a simulator for the category of conjoined musical events associated with the term “perfect authentic cadence” could be established simply through multiple encounters with exemplars of the category. It should be emphasized that such a simulator would not only include auditory information, but also extend to sensorimotor information
about the feeling of performing these events (either as a soloist playing a harmonic instrument or as a member of an ensemble), introspective states associated with such cadences, and physical responses to hearing them. Second, the configuration of properties and relations encapsulated by the simulator for a category may, under certain circumstances, be applied to a different category, giving rise to analogy (Barsalou 2005, 422). For instance, among early Spanish explorers of the New World, the hummingbird was known as the “paxaro [pájaro] mosquito”—that is, the mosquito bird (Fernández de Oviedo y Valdés [1547] 1851–55, 1: 444). There seems little doubt the explorers appreciated that the creature they had encountered was a bird rather than an insect, but its small size together with the characteristic sound and trajectory of its flight made mapping the configuration of properties and relations for the category MOSQUITO onto the category BIRD an apt way to identify the animal. In a like fashion, the configuration of properties and relations associated with the event-category THE FLIGHT OF A HUMMINGBIRD could be applied to a category of conjoined musical events, such as that exemplified by Sagreras’s “El Colibri.”

By its very nature, Barsalou’s theory lends itself to a highly dynamic view of concepts: successive encounters with members of a category will result in slightly different activations of perceptual feature maps, and these in turn may modify the populations of conjunctive neurons in association areas. The categories through which we structure our understanding are, in consequence, dynamic structures that are modified as further exemplars are added, and the simulation of which changes with the current context of action. Musical categories—the notion of which is developed further below—are not fundamentally different but have two distinctive characteristics. First, the bulk of the perceptual information that forms the basis for musical categories—sounds, physical sensations, and information from proprioception and introspection—has nothing to do with vision (except, of course, in relatively rare cases of music-image synesthesia). Although concrete objects, in the form of musical instruments and other sounding bodies, are important as a source and locus for this information, these objects tend to be regarded as distinct from musical phenomena. Musical categories are, in consequence, rather insubstantial structures, being both invisible and lacking a concrete intersubjective manifestation. Musical objects, to the extent that such exist, are convenient fictions that facilitate discourse about music but, without further interpretation, do little to reveal its inner workings. Second, musical categories typically involve temporally distributed events: before and after are intrinsic parts of musical experience. Because the phenomena that make up these temporally unfolding events are themselves ephemeral and insubstantial, conceptualizing music involves dealing with materials that mirror the volatility of thought itself.

I began this chapter by suggesting how an event (the flight of a hummingbird) could be correlated with a particular succession of musical materials (Sagreras’s “El Colibri”). While mapping from music to some natural phenomenon is a notion not without seductive power for the musician, it must be acknowledged that the more typical mapping is from some natural phenomenon to music. Let me suggest two reasons why this should be so. First, although the performance of music is a thoroughly embodied affair, listening to music may be markedly less so (even
if we allow for the sort of mimetic coproduction proposed in Cox 2016). For the listener, the conceptual structures of music can appear to lack any concomitant material structure. Ed Hutchins, in his work on material anchors, has shown that in such cases our tendency is to find a material structure that can be correlated with conceptual structure (Hutchins 2005). Mapping from natural phenomena to music offers one way to provide music with material anchors and thereby ground its apparently insubstantial concepts in objects and events from the concrete world. Second, it is often (although not necessarily) the case that our experience of natural phenomena is perceptually rich and multimodal. Sonic analogs for such experiences will inevitably simplify them but will also offer an opportunity to reveal the configuration of properties and relations that makes the experiences distinctive. Mappings between natural phenomena and music can thus make those phenomena, in a manner of speaking, more definite. In this way, sonic analogs can be thought of as capturing deep aspects of our experience.

Summary

It should be emphasized that Barsalou’s theory of perceptual symbol systems is a theory: it is a comprehensive account of structures and processes within human cognition that makes possible predictions on how perceptual information is handled and how it contributes to the substance of our thought. It is one of a number of current theoretical approaches that seek to explain the part embodied experience plays in human cognition and, in particular, how it shapes conceptual knowledge (for a review, see Gibbs 2006, chap. 4; see also Barsalou 2008 and the essays collected in Pecher and Zwaan 2005). Human cognition is enormously complex, and what we know about it changes daily. Hence, even the most fully worked-out theory, if it aims to account for cognitive processes of any extent, must be regarded as provisional. That said, empirical evidence that supports Barsalou’s theory is gradually accumulating. This includes studies of the situational content of abstract concepts (Barsalou and Wiemer-Hastings 2005); the process of simulating properties (Solomon and Barsalou 2004); the areas of the brain involved in verifying properties (Kan et al. 2003); and the way property attributes from different perceptual modalities are processed (Pecher, Zeelenberg, and Barsalou 2004). What is perhaps most important for my purposes, however, is that the theory offers a way to explain how the perceptual information basic to musical experience contributes to the concepts through which we organize our musical thought. As such, Barsalou’s theory can be seen as a way to augment Mark Johnson’s theory of image schemas (M. Johnson 1987), which I and a number of music scholars have used as the basis for an embodied account of musical organization (Zbikowski 2008b, 510–12).

From the perspective of Barsalou’s theory of perceptual symbol systems, the sonic analogs of music can be seen as a form of conceptual knowledge that is derived from perceptual information, and that is connected to other kinds of conceptual knowledge through shared configurations of properties and relations. Together with the account of analogy I developed above, the framework offered
by Barsalou’s theory also provides a basis for fleshing out the notion of analogi-
cal reference introduced in chapter 1. In what follows I would like to return to
the topic of analogical reference by way of the semiotic theory of the nineteenth-
century polymath Charles Saunders Peirce. Although analogical reference was of
only limited importance for Peirce, the overall framework provided by his theory
is useful for understanding the relationship between analogical and symbolic ref-
erence, and it can offer insights into how analogical reference both makes possible
and constrains musical expression.

**Sonic Analogs and Systems of Reference**

**Peircean Semiotics and Analogical Reference**

It was in the work of Peirce that the study of systems of reference was gathered under
the rubric of semiotics (an approach anticipated by John Locke’s threefold division
of human knowledge at the end of his *Essay Concerning Human Understanding*
[(1689) 1965, 2: 309]). Peirce’s ideas have been central to the study of signs and
meaning, but applying them to the specific issues with which I am engaged—
analogical reference in general, and its manifestations in music in particular—
presents two problems. First, although Peirce had a deep love of system, he was
not always systematic in his writing. As Umberto Eco observed, “Peirce was com-
pelled continuously to rediscuss and revise his ideas; he felt a sort of psychologi-
cal pleasure in challenging and re-defining his own formulas; it is rather difficult
to find two separate passages on the same topic in which he does not contradict
and re-propose what he has previously said” (1976a, 1457). What rigor there is in
Peirce’s system of semiotics is often an artifact of interpretation rather than part of
the legacy of his thought. Second, Peirce’s focus was almost invariably on objects
and relations, for which language provides the most sophisticated system of refer-
ence. He only rarely commented on issues related to the representation of dynamic
processes of the sort with which music is typically concerned, and so extension into
that domain must proceed with caution.

One way to think of Peirce’s study of signs is as an exploration of the origin
and nature of the thoughts that are connected with some aspect of experience.
Peirce described this connection through a set of nested triadic relationships,
which begin with an object (the relevant aspect of experience), a sign which stands
for this object, and the thought-structure created in someone’s mind by this sign.
Peirce offered at least one formulation of these relationships, along with more for-
mal terms for the elements involved, that is refreshing in its clarity: “A sign, or
*representamen*, is something which stands to somebody for something in some
respect or capacity. It addresses somebody, that is, creates in the mind of that per-
son an equivalent sign, or perhaps a more developed sign. That sign which it cre-
ates I call the *interpretant* of the first sign. The sign stands for something, its *object.*
It stands for that object, not in all respects, but in reference to a sort of idea, which
I have sometimes called the *ground* of the representamen” (Peirce 1955, 99).
As suggested by the perspective on analogical reference I developed in chapter 1, my reading of Peirce is informed by Terry Deacon’s work on language evolution. This is especially so in the case of Peirce’s second set of triadic relationships, which concerned the forms the sign could take: as icon, index, or symbol. Deacon notes that what was important for Peirce was the relationship between the characteristics of the sign token and those of the physical object represented. Deacon summarized these relationships as follows: “icons are mediated by a similarity between sign and object, indices are mediated by some physical or temporal connection between sign and object, and symbols are mediated by some formal or merely agreed-upon link irrespective of any physical characteristics of either sign or object” (1997, 70). It must be emphasized that this is a summary of basic distinctions Peirce made rather than a comprehensive account of his theory. In his characterization of the icon, for instance, Peirce emphasized its almost purely phenomenological status: “An Icon is a Representamen whose Representative Quality is a Firstness of it as a First. That is, a quality that it has qua thing renders it fit to be a representamen” (1960, 2: 276). (Peirce described a first as “that whose being is simply in itself, not referring to anything nor lying behind anything” [Peirce 1960, 1: 356]. “Firstness” is the essential quality of a first.) Peirce then offered a way to ground this elusive concept: “a sign may be iconic, that is, may represent its object mainly by its similarity, no matter what its mode of being. If a substantive is wanted, an iconic representamen may be termed a hypoicon. Any material image, as a painting, is largely conventional in its mode of representation; but in itself, without legend or label it may be called a hypoicon” (Peirce 1960, 1: 276). More properly, then, the type of sign Deacon calls an icon is a hypoicon. In this and other cases, however, Deacon’s basic perspective on Peircean semiotics will be sufficient for the account of analogical reference that I want to develop.

Peirce’s semiotic theory gave Deacon a framework for describing how our species developed language. The simple ability to use signs, broadly understood, was not enough in itself, since there is clear evidence that other species are able to make use of icons and indices. What was crucial was being able to use signs—and in particular indices—to refer not simply to objects but to other signs. By this means it was possible to build up the dense network of interconnected symbols on which language is based (Deacon 1997, chap. 3; 2003).

If, as suggested by Philip Tagg’s notion of an anaphone (discussed in chapter 1), what I call a sonic analog is akin to Peirce’s notion of an icon and if, from Deacon’s perspective, other species can make use of icons, why is it that other species have not developed music? The key is analogical reference, a form of reference that is part of Peirce’s fuller account of the icon (or, more accurately, the hypoicon). As suggested by Deacon’s summary, icons represent their objects by being like them. For Peirce, this likeness may take one of three forms: “Those [hypoicons] which partake of simple qualities, or First Firstnesses, are images; those which represent the relations, mainly dyadic, or so regarded, of the parts of one thing by analogous relations in their own parts, are diagrams; those which represent the representative character of a representamen by representing a parallelism in something else, are metaphors” (Peirce 1960, 1: 277). Peirce did not elaborate this division further, but based on his overall approach it seems fair to say that the image was,
in its essential respects, indistinguishable from its object. In contrast, diagrams preserve structural relationships with their objects (but not, perhaps, their surface features), where metaphors offer a looser but still discernible connection between the icon and its object. As noted earlier, humans appear to be the only species with a robust capacity for analogy: while other species may be able to make use of the form of icon that Peirce called an image, they will not, independent of human intervention, be able to understand icons that are diagrams or metaphors.

To develop more fully the notion of a sonic analog, I would like to expand Peirce’s notion of an icon in two ways. First, I want to propose that the “object” of an icon may be a dynamic process. Second, although Peirce appeared to regard the iconic image, diagram, and metaphor as discrete categories, I want to suggest that they can be thought of as situated along a continuum of signs that range from those with a great deal of fidelity to the object to those that preserve only a few selected features of the object. Example 2.3 provides a sketch of this continuum, offering one interpretation of how different sonic events might be situated along it. Sound effects that attempt to represent with as much fidelity as possible some actual sonic event (such as the sound of a helicopter) are regarded as a kind of image. Particularly effective examples (among which I would include the imitation of birdsong) may deceive a listener into thinking that the sound was produced not by a human but by its typical source (be that a helicopter or a bird). Sound symbols, which include onomatopoeic words and ad hoc sound effects interjected into the stream of speech, fall between the image and the diagram. Just where they would fall will depend in part on the extent to which they replicate the essential features of some target sound event, and in part on whether the dynamic event they aim to represent carries with it any sound at all. (One example of the latter is a rapidly descending whistling sound, used to represent a quick and typically precipitous—but soundless—descent of an object.) The sonic analogs of music seem closest to Peirce’s diagram or metaphor: although they may have correlations with a real or imagined sonic event (such as a bird call, or a sudden loud sound), they more typically analogize dynamic processes whose attributes are predominantly nonsonic.

Mine is by no means the first adaptation (or appropriation) of Peircean semiotics for music. Most scholars who have applied semiotic theories to music—including those I mentioned in chapter 1—take as their point of departure some version of Peirce’s theories. My approach is different in that it is decidedly

---

**Example 2.3** Illustration of a continuum of icon types drawn from Peirce’s tripartite division of hypoicons, including various specific types of sonic events

<table>
<thead>
<tr>
<th>sound effects</th>
<th>sound symbols</th>
<th>sonic analogs</th>
</tr>
</thead>
<tbody>
<tr>
<td>image</td>
<td>all essential aspects of object preserved</td>
<td></td>
</tr>
<tr>
<td>diagram</td>
<td>only some aspects of object preserved</td>
<td></td>
</tr>
<tr>
<td>metaphor</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---
limited—my interest in Peirce, as far as this volume goes, pretty much begins and ends with his basic ideas about reference and iconic signs—and my focus is on musical signs that stand for dynamic processes. Again, my aim in this volume is to develop a foundation for a cognitive grammar of music, not to set out a fully worked-out semiotics of music. My assumption is that such a semiotics will involve musical signs that stand for things other than dynamic processes, although I believe any fully developed semiotics of music must come to terms with the different ways music and language construct meaning, a difference that reflects the highly dynamic character of musical expression.

**Summary**

I view the crucial factor in reference to be the notion of “standing for” highlighted in Peirce’s fundamental trichotomy: for a sign to be a sign it has to stand for something to somebody, and the manifestation of this “standing for” is a mental construct (Peirce’s interpretant). With analogical reference, the connection of a sign with an object is facilitated by structural similarities between the two. As with all analogies, the discovery of these structural similarities reflects the goals of the person drawing the analogy, and it requires knowledge of the relational frame basic to the analogy. Analogical reference thus entails the establishment of an analogical relationship between two things, and for one of the things (the sign) to stand for the other (the object) for some person.

As I noted in chapter 1, both symbolic reference and analogical reference have their advantages and disadvantages. Because practically anything can serve as a symbolic token, such tokens are fairly easy to produce; the dense systems of reference of which such tokens are a part, however, require considerable effort to establish and learn. By comparison, analogical tokens require rather more effort to produce (to ensure that they have the structural features they share with the objects with which they are correlated), but they make far fewer demands on interpretation. With respect to dynamic processes, analogical tokens also have the potential to represent, through an ordered temporal sequence, the parametric changes that typify such processes.

Within the framework of Barsalou’s theory of perceptual symbol systems, analogy arises when the configuration of properties and relations encapsulated by the simulator for a category is applied to a different category. For example, the simulator for the category of sounds made by a horse’s hooves on pavement could be applied to the sounds made when I say “clip-clop, clip-clop.” I would like to propose that this process may also work in reverse. Although it is doubtful that you would mistake the sounds I make for those made by a real horse ambling down the street, hearing me say “clip-clop, clip-clop” will have two consequences. First, feature maps quite similar to those connected with hearing the sounds made by a horse walking will be activated, with the result that a similar group of conjunctive neurons in an association area will also be activated. Shared neural connections between these two groups of conjunctive neurons may provide additional support for analogical reference. Second, if you recognize the sounds I make as like those
made by a horse walking, you will be prompted to remember those specific horsey sounds, and in remembering them reenact the sensorimotor states that were active when you encoded them. In both cases, the analogical reference of “clip-clop, clip-clop” will be one that is embodied. If this sort of reverse process is in fact activated by analogs for dynamic processes—and I take this to be a question open to empirical investigation—it suggests one explanation for the immediacy and effectiveness of analogical reference.

Sonic Analogs for Dynamic Processes

Analogizing Dynamic Processes

The dynamic processes I have mentioned or discussed thus far include the sound made by a horse walking on pavement, the precipitous if soundless descent of some object, the sound of a bird or helicopter, the flight of a hummingbird, and (in chapter 1) the ritual pacing done by the Akpafu when they mourn, and the emotions and actions that might accompany the declamation of “Mein Gott! wenn kommt das schöne: Nun!” I proposed that through sonic analogs we can make analogical reference to each of these processes. (Note that in some cases—such as that of a horse walking on pavement—the dynamic process yields a sound to which other sounds, such as “clip-clop, clip-clop,” make analogical reference.)

I would now like to offer a more formal, if provisional, definition of a dynamic process, as well as consider various ways we can analogize such processes. The definition will lend a bit of precision to the notion of what counts as a dynamic process, and the consideration will situate musical practice within the range of means through which such processes can be analogized.

I take a dynamic process to be a coherent sequence of phenomena that is distributed over time and typified by parametric modulation or change. The notion of a coherent sequence of phenomena assumes that these phenomena follow one another and can be collected together into a single category, such as THE SOUNDS MADE BY A HORSE WHEN WALKING. Given this understanding, specifying the distribution of this sequence over time would appear to be slightly redundant, since things that follow one another will of necessity be distributed over time. Nonetheless, I would like to foreground our sense that the temporal frame for dynamic processes is a comprehensible one: there are, for instance, any number of coherent sequences of geological phenomena that are distributed over time, but given the limits of human perception our appreciation of the temporal frame occupied by such phenomena is decidedly limited. The requirement of parametric modulation or change suggests that a sort of figure-ground distinction is basic to dynamic processes: while certain features of the phenomena comprised by such processes undergo change, others do not, and this contrast helps us to appreciate the sequence of phenomena as a dynamic process. Given this definition, the minimal requirements for any analog for a dynamic process are that the analog provide a correlate for the parametric modulation or change characteristic of the
process and represent its sequence of phenomena within a temporal frame that is proportional to that of the original.

As my examples have shown, one way to analogize dynamic processes is through sounds, which can range from sound effects to the sonic analogs of music. Because language relies primarily on symbolic reference, most of the sounds that occur in the speech stream (with the exception of onomatopoeic words and ad hoc sound effects) are nonanalogical. There is some evidence, however, that the speech stream itself may take on an analogical function and thereby represent dynamic processes, either through the organization of grammatical elements (Haiman 1985; Langacker 2001, 2005) or through prosody (Shintel, Nusbaum, and Okrent 2006; Shintel and Nusbaum 2007; Hedger, Nusbaum, and Hoeckner 2013). Successions of visual images offer another resource for analogizing dynamic processes, either through arranging the images in a series or through animation (Arnheim 1974, chaps. 8 and 9; Pedone, Hummel, and Holyoak 2001). And gesture (which I shall consider in greater depth in chapter 4) provides yet another means of creating analogs for dynamic processes. An example can be found in the rapid shifts of the left hand that occur in the performance of Sagreras’s “El Colibri.” These shifts provide the listener with a ready visual image of the movement of the hummingbird and also directly involve the performer in these characteristic gestures: for a brief moment, the guitarist’s hand becomes the hummingbird flitting between flowers.

**Sonic Analogs, Musical Grammar, and Cognitive Capacities**

The provisional definition of a dynamic process that I have offered, together with research on analogy, makes possible a somewhat more detailed account of how the basic elements of musical grammar are configured. As I observed in chapter 1, one of the distinctive features of musical grammar is that its basic elements are holistic structures that combine and temporally order sonic events to create analogs for dynamic processes. As an example, consider measures 9–12 from Sagreras’s “El Colibri,” which I earlier suggested could be correlated with the small movements the hummingbird makes as it dips to drink. First, a bit of musical context: this short passage occurs within a move toward the dominant initiated by the campanellas passage of measures 4–7, a move signaled by the introduction of A♯ in measure 6 and secured by the first-inversion F♯ seventh chord of measure 8. Measures 9 and 10 are built around a cadential 6–5–4–3 figure in B major, shown in schematic form in example 2.4, which is repeated in measures 11–12. This repetition is followed by the octave leaps of measures 13–16, which toss the pitch class B back and forth between four octaves and which lead to the reprise of measure 1 (and return to tonic) in measure 17. In terms of the harmonic design of the piece, then, measures 9–12 are part of a holding pattern in which the music, having arrived on the dominant, pauses for a moment before returning to tonic.

The accented passing-tone figures of measures 9–12 (which embellish the basic framework of ex. 2.4) are novel within the sequence of musical events up to this point (although they recall some of the energy of the quick portamento of measure
Foundations of Musical Grammar

3), as is the repetition of a two-measure block of material. Given this, a case can be made that, even though these four measures last only about two seconds at tempo, they stand out from the whole as a distinctive collection of musical materials. The rapidity of the change of pitches (from E5 to D♯5, from D♯5 to C♯5, and from C♯5 to B5) and the comparatively longer duration of each of the arrival pitches (as they are repeated three or seven times) correlate well with sudden, jerky changes of movement followed by temporary stasis; the arrival on a harmony that functions as an objective for the preceding musical events correlates well with a goal-directed physical action, albeit one that is immediately repeated. In brief, measures 9–12 represent a holistic structure that combines and temporally orders sonic events to create an analog for the dynamic process of a hummingbird dipping to drink from flowers. From the perspective I should like to develop here, this structure represents a basic element of musical grammar. Two points bear emphasis here: first (and as shown by my analysis of the passage), it is certainly possible to isolate and analyze individual sonic events that are encompassed by this holistic structure (and, in the course of such an analysis, making use of terms quite familiar to music theorists). The function of the structure is, however, a product of the combination of these events into a single coherent whole. Second, and related to this, the particular combination of events realized in measures 9–12 represents only one way to create a sonic analog for the dynamic process of a hummingbird dipping to drink from flowers. Other combinations of sonic events, perhaps of a quite different cast than those heard in measures 9–12, could achieve the same function.

The perspective that has guided my analysis of this passage is one that will develop more fully in the following pages, but it takes as one of its framing assumptions the idea that musical grammar reflects the cognitive capacities of human beings. The capacity for making and understanding complex analogies is key to musical grammar, but so are the processes of categorization through which humans organize their understanding of the world (discussed in the next section) and the memory functions that support our retention of these categories (discussed in the section that follows after).

Processes of Categorization. In their general features, holistic units such as that represented by measures 9–12 of Sagreras’s “El Colibri” exemplify two aspects of the processes of categorization through which humans organize their understanding of the world. The first of these is a preference for a level of categorization that is at the middle of a taxonomy—what has come to be called the basic level (Rosch 1999, 191–96; Zbikowski 2002, 31–36). The importance of basic-level categories is reflected in the shape taken by our everyday conversation: were I to

Example 2.4 Schematic representation of the harmonic basis of measures 9–10 of Julio S. Sagreras’s “El Colibri”
refer to the creature who demands to go out each morning to patrol our backyard, I would probably describe him not as “the diminutive gray and white cat” or “the domestic quadruped” but simply as “the cat.” The first description provides highly specific information (picking this cat out from a host of other cats), the second description provides very generic information (classing the animal with any number of domestic quadrupeds), while the third description—which, in this context, constitutes a basic-level category—makes a compromise between these two that adds greatly to the efficiency of communication. In a similar fashion, measures 9–12 represent a compromise between information that is highly specific (of a sort hinted at in my careful description of aspects of these measures) and information that is highly generic (as would be illustrated by a characterization of these measures as comprising a steady succession of musical events that are divided between two sounding voices), and in this way is very much like a basic-level category.

The second aspect of categorization concerns what are called typicality effects: although it is common to think of categories as having crisp boundaries determined by a set of necessary and sufficient conditions through which category membership is determined, the categories that humans actually use in daily life have a graded membership, with individual members characterized as being more or less typical of the category as a whole (Rosch 1999, 196–201; Zbikowski 2002, 36–49).

An example of something like typicality effects can be seen in Sagreras’s reuse of the material of measures 9–12 a bit later in the piece. As shown in example 2.5, after a complete reprise of the music of measures 1–8 a modified version of measures 9–10 appears in measures 25–26, D♯5 having been replaced by D♮5. This is followed by a version modified yet further through the introduction, in the latter half of measure 28, of an inverted form of the accented passing-tone figure. As a last development of the figure (at least in this section of “El Colibri”), the material of measure 28—with the opening major second now expanded to a minor third—serves as the basis for the stepwise-descending sequence of measures 29–31 that leads to the arrival on a first-inversion C major chord in measure 32. Were we to think of measures 9–12 as the most typical members of a musical category (a notion reinforced by their unmodified reappearance a bit later in this short piece), measures 25–28 would be a slightly less typical form, and the sequential passage of measures 29–32 less typical still. As I observed in previous work, this contrast between materials that typify a given category of musical events and those that are less typical offers composers a means to shape the overall dynamic progress of a work and is, in consequence, important for large-scale musical syntax (Zbikowski 2002, chap. 4).

What I have provided here offers only a brief glimpse of the role of processes of categorization in human understanding (a somewhat expanded view is offered in chapters 1 and 4 of Zbikowski 2002), but it should give a sense of the importance of such processes to the establishment of the basic elements of musical grammar (through musical structures similar to basic-level categories) and to the organization of musical syntax (through the ways musical structures that are similar to categories that show typicality effects are arranged over the course of a musical utterance).

Working and Long-Term Memory. Current research indicates that two interrelated components of human memory shape our understanding of communicative
The first is what is commonly called working memory, which is understood to be a limited-capacity, temporary storage system that provides support for complex human thought (Baddeley 2007, 6–7; see also the discussion of musical memory in Snyder 2000). Although specifying the limits of working-memory capacity has generated lively discussions (as shown by the responses to Cowan 2001; see also Cowan 2005 and Baddeley 2007, chaps. 10 and 11) there is converging evidence that it extends to approximately four “chunks” of information, a span that comfortably fits the materials of measures 9–12 from Sagreras’s “El Colibri.” Provisionally, then, we might conjecture that the information organized by the basic elements of musical grammar fits within the normal capacity of working memory, which would in turn contribute to the usefulness of such elements for the assembly and comprehension of more complex series of musical materials (a compositional process in evidence in measures 25–32).

The second component of human memory important for understanding communicative media is what is commonly called long-term memory, which is understood to reflect relatively durable changes to synaptic structure. Where working

Example 2.5 Measures 17–33 of Julio S. Sagreras’s “El Colibri.” © 1954 Ricordi Americana, Buenos Aires.
memory stores (and only briefly) a limited number of chunks of information, long-term memory stores more complex relational structures that make possible the interpretation of the information held in working memory. In studies of the role of memory in language comprehension, the current understanding is that the more complex relational structures that are stored in long-term memory are related to syntax (Baddeley 2007, chap. 8). Although it gets ahead of our story just a bit (as I shall devote more attention to syntactic processes in chapter 3), with respect to Sagreras’s “El Colibri” it seems likely that knowledge about how musical materials are organized to create larger structures would help with the interpretation of measures 9–12, in that the momentary pause created by repeating the material of measures 9–10 in measures 11–12 correlates with an arrival in a contrasting key area (here, the key of the dominant), a common event in works created within the tradition of compositional strategies to which Sagreras belonged. Knowledge about the part played by contrasting key areas in tonally organized music seems a likely candidate for the sort of information that would be stored in long-term memory and that would be called upon not only to make sense of measures 9–12 but also to understand the reprise of this material—but now in B minor rather than B major—in measures 25–28.

**Summary.** As with my account of how processes of categorization may shape our understanding of music, there is much more that could be said about the relationship between the resources provided by human memory systems and the organization of musical materials. When taken together with research on analogy, however, what I have sketched here offers a basic framework for an exploration of the ways musical grammar reflects the cognitive capacities through which humans structure their understanding of the world.

**Analogy and Imagination**

The perspective on the role of analogy in understanding sequences of musical events sketched in this chapter provides a way to answer some of the questions with which I began—namely, why sounds can serve as analogs for processes that are predominantly visual, and why such connections appear to be limited to humans. My analysis of measures 9–12 of Sagreras’s “El Colibri” made explicit the connection between sonic analogs and musical grammar, and it also suggested why thinking of musical mimesis as a strictly mechanical phenomenon or a manifestation of culturally conditioned associations fails to capture the effectiveness of sonic analogs and the resources they offer for structuring musical discourse. My further consideration of measures 25–32 of “El Colibri” offered a glimpse of ways the creative imagination of the composer—exploiting listeners’ capacities for categorization and their ability to retain information in memory—could transform relatively limited musical resources into the basis for more sustained utterances. The realm of the imagination opened up by musical utterances can also be augmented through the recursive application of analogical mappings and through the principled exploration of additional analogical relationships suggested by the
first- and second-order relations basic to a given analogical mapping, offering opportunities to blend musical concepts with those from nonmusical domains.

**Induced Schemas and Conceptual Blending**

In their review of an extended series of experiments they conducted on the use of analogy to solve problems, Mary Gick and Keith Holyoak noted that the success of participants relied on the discovery of an abstract core idea, or schema, shared between the source domain (which offered the means to solve the problem) and the target domain (to which the problem belonged) (Gick and Holyoak 1983). As Gick and Holyoak saw it, the schema that was induced through this process reflected a partial mapping between the source and target domains, which then provided the means by which the mapping could be extended to build new knowledge.

An extension of this perspective on the process of making analogies—and one important for understanding the part analogy can play in more complex thought processes—came about as part of research by Gilles Fauconnier and Mark Turner on ways that analogical and metaphorical mappings contribute to the construction of meaning. Key to their approach was the notion of a mental space, which they contrasted with the notion of a domain typically used in studies of analogy: “We use the term mental space in contrast to the term conceptual domain. A mental space is a (relatively small) conceptual packet built up for purposes of local understanding and action. Mental spaces are constructed whenever we think and talk. They are interconnected, and they can be modified as discourse unfolds” (Turner and Fauconnier 1995, 184). Turner and Fauconnier would thus interpret Gick and Holyoak’s model as involving a network of three interconnected mental spaces: one associated with the source domain, one associated with the target domain, and one associated with the induced schema that linked the two. In certain cases, this network of three spaces would give rise to a fourth mental space in which concepts from the mental spaces associated with the source and target domains were combined to yield new possibilities for the construction of meaning, a phenomenon that Turner and Fauconnier called conceptual blending (Fauconnier and Turner 2002, chap. 3; Zbikowski 2002, 77–82).

**Conceptual Blends and Sagreras’s “El Colibri”**

As a way to illustrate the process of conceptual blending, let us return to the analogy set up by Sagreras’s “El Colibri.” On Turner and Fauconnier’s interpretation, and as shown in the conceptual integration network (CIN) diagrammed in example 2.6, Sagreras’s music would prompt the construction of one mental space (the “El Colibri” space), and the title of the work (“El Colibri, Imitación al vuelo del picaflor”) would prompt the construction of another (the hummingbird space). The “El Colibri” space would include concepts prompted by the distinctive features of the piece: streams of repeated notes, discrete blocks of musical materials, accented non-chord tones, and a continuous succession of pitches. The
hummingbird space would include concepts associated with the distinctive features of the bird: its darting to and fro, the dipping movements it makes as it feeds, its active stasis as it hovers over a flower, and, of course, the characteristic sound of its flight. As is typical for analogical mappings, these mental spaces could then be correlated with one another, perhaps through reading the title of the work on a concert program and then hearing the work performed. This correlation, indicated by a solid double-headed arrow in example 2.6, would activate a third mental space, the generic space at the top of the diagram; the correlated spaces can thus be viewed as input spaces for the CIN. The generic space, which plays a similar role in analogical mappings to that of Gick and Holyoak’s induced schema, captures the abstract core idea shared between the two correlated spaces, which is of highly dynamic processes that are both noteworthy and immediately evident. This network of three spaces then supports the construction of a fourth space (the conceptual blend of example 2.6) in which concepts from the “El Colibri” space and the hummingbird space are blended together to create new concepts not found in

**Example 2.6** Conceptual integration network for Julio S. Sagreras’s “El Colibri”
any of the other mental spaces in the CIN, including the idea that highly dynamic physical processes create saturated soundscapes or, more succinctly, that energetic agents are noisy agents.

The dashed arrows linking the generic space to the input spaces, and the input spaces to the blended space, indicate the directions in which structure is projected: from the generic space to the input spaces, and from the input spaces to the blended space. The arrows are double-headed because, under certain circumstances, structure may also be projected from the blended space back into the input spaces, and from the input spaces back into the generic space. The notion that agents associated with considerable energetic forces are agents that also produce considerable sound could thus inform the way we hear a composition without any extra-musical program. For example, Matthew Riley describes minor-mode symphonies from eighteenth-century Vienna whose last movement is characterized by relentless energy and fast surface rhythms as works with a stormy finale, the force and energy of a tempest being a way to account for the features of the music (Riley 2014, 120–39). As I noted in earlier work, the double-headed arrows in diagrams like example 2.6 also serve as a reminder of the limitation of diagrams of CINs: mental spaces are dynamic structures, as are the CINs that are built from them (Zbikowski 2002, 80). What example 2.6 represents is a sort of analytical snapshot of a particular network, framed with the intent of capturing its essential features but making no claim to exhausting the possibilities for description. Hints about how a CIN and its spaces may develop can be gleaned from the diagram, but a full account would require a series of such snapshots.

Turner and Fauconnier proposed that the emergent structure unique to the blended space reflected three constitutive and governing principles: composition, completion, and elaboration (Fauconnier and Turner 2002, 345). Composition puts together elements from the input spaces to create new entities in the blended space—in the case of the conceptual blend developed around Sagreras’s “El Colibri,” this yields a sequence of visual images so tightly correlated with a sequence of sounds that the two seem inseparable (and which thus exemplify what Nicholas Cook has called an “Instance of MultiMedia”; Cook 1998, 98). Completion extends the image suggested by the initial mapping from the input spaces, drawing on our background knowledge of the circumstances summoned by the analogical mapping. For instance, we know that dynamic processes often change in substantive ways over time, sometimes quite suddenly, making it difficult to predict their long-range course (something borne out by the conclusion of Sagreras’s “El Colibri,” when the steady succession of pitches is suddenly interrupted by silence, followed by an arpeggiated fully diminished seventh chord that describes a precipitous ascent into the guitar’s highest register). Elaboration is an operation more extensive than completion; it develops the structure of the blended space by building upon the network of analogical mappings basic to the blend. In effect, the input spaces decrease in importance and the focus is directed toward the rich imaginary
possibilities of the blended space. In the case of the conceptual blend associated with Sagreras’s “El Colibri,” we are led toward a realm of the imagination in which a wide range of highly dynamic physical phenomena can be represented through sequences of patterned sound. The possibilities of such imaginative play are vividly illustrated by the soundtracks for cartoons, in which image and sound are often in close correspondence with one another (Curtis 1992, 201–2); as we shall see over the course of the following chapters, however, these possibilities can also be realized in more subtle ways.

Three aspects of the process of conceptual blending as it applies to music bear further comment. First, the potential for music to participate in conceptual blends suggests that the musical materials that set up the “El Colibri” space are properly conceptual (a notion formalized at the end of this chapter). Although these materials can be described with words (as they are in example 2.6), that is simply an artifact of analysis—music need not rely on language for the formation or stabilization of concepts. Second, the concepts that emerge in the blended space provide a sense of how music shapes our understanding: that is, within the blend physical phenomena are not simply accompanied by sound, but represented by (nonlinguistic) sound. Third, the theory of conceptual blending assumes that all of the mental spaces of a CIN share a basic topology, meaning that similar elements and relations can be found in each of the mental spaces (a similarity reflected in the abstract structure offered by the generic space; cf. Fauconnier and Turner 2002, 325–28). This introduces a tension not immediately evident in my discussion of the conceptual blend associated with Sagreras’s “El Colibri” but that will become more evident in my analysis of nineteenth-century Lieder in chapter 6. As I have noted, language is a useful tool for picking out objects and events, and characterizing relationships between them; in consequence, the concepts associated with language will tend to involve discrete elements and relations between them. This is not to say language cannot activate an image (which is certainly the intent behind the “darting path of flight” in the hummingbird space of example 2.6) but its strength lies in the system of symbolic reference that undergirds words. Music, by contrast, is organized around sonic analogs for dynamic processes; its concepts are thus different in kind from those of language. This difference between linguistic and musical concepts puts pressure on the notion that all of the mental spaces in a CIN share a basic topology, a matter to which I shall return in my discussion of music and words in chapter 6.

I should emphasize that analogical mappings do not necessarily lead to conceptual blends—as research on conceptual blending has shown, the integration of concepts from different domains is constrained by a number of factors that shape the processes through which new meaning is created (Fauconnier and Turner 2002, chap. 16; see also Fauconnier 2001). That said, the theory of conceptual blending will prove a useful tool for exploring how musical materials can combine with physical gestures (in chapter 4) as well as with words (in chapter 6) to build up rich contexts for the construction of meaning.
Conclusion

Analogy, Analogical Reference, and Dynamic Processes

One of the guiding assumptions of this chapter was that listeners can, with little effort, link the sounds of Sagreras’s “El Colibri” with the dynamic processes associated with a hummingbird in flight. As has been noted, however, without the aid of the title or similar cues, it could well be that a listener would fail to make this connection—she might instead be prompted to think of an insect trapped under a glass and frantically searching for escape, or of a Greek waiter rushing from table to table in a packed restaurant (dynamic images offered by acquaintances who have listened to “El Colibri” without knowing its title). That said, what is important for my argument is less the fidelity of such images with Sagreras’s compositional inspiration and more the notion that sound sequences can prompt these sorts of mental constructs in the first place. That this should be so reflects humans’ capacity for analogy and (building on Barsalou’s theory of perceptual symbol systems) the ways conceptual knowledge that is derived from perceptual information can be connected to other kinds of conceptual knowledge through shared configurations of properties and relations. Indeed, the rich tradition of text painting and musical imagery in evidence across musical cultures—ranging from portrayals of waves, clouds, storms, and horses among Western musicians (Monelle 2000, 17), to a summoning of the flight of the lark by Afghani musicians (Baily 1997, 56), to an evocation of a rushing stream by a Tuvan musician (Levin and Süzükei 2006, 26–29)—points toward a signal resource for human expression through which sound sequences analogically refer to dynamic processes.

Two aspects of analogical reference are especially relevant for the development of a cognitive grammar of music. First, music is not the only communicative medium to exploit analogical reference: as noted, linguistic communication can exploit analogical reference (through onomatopoeia and prosody) and, as subsequent chapters will prove, so can gesture and dance. Second (and building on Barsalou’s perceptual symbol systems theory), analogical reference can, in many instances, be grounded in embodied experience: the compositional strategy of cadence, for instance, provides a sonic analog for the process of arriving at a goal (no matter how temporary), a process we reenact each time we sit down in a chair or roll into our bed. Such grounding supports the apparent immediacy of analogical reference and can also constrain—in productive ways—its application.

Elements of Musical Grammar: Basic Elements

The basic elements of musical grammar involve nonlinguistic sound sequences that unfold in time. As such, they will be constrained by the human cognitive processes concerned with temporal processing, in particular those associated with working memory. These basic elements will, in consequence, be of limited scope, perhaps corresponding to the four chunks of information currently thought to
exhaust the capacity of working memory. Such basic elements can, of course, be connected with other basic elements through syntactic processes; according to present theories of the way human memory systems are organized, the resulting complex relational structures would be stored in long-term memory.

According to the theory I develop in this volume, the basic elements of musical grammar are analogous to basic-level categories. This construal has three important consequences. First, it makes a direct connection between the principles through which musical utterances are organized and the basic cognitive processes through which humans organize their understanding of the world. Second, it provides, through the notion of categories that show typicality effects, a way to describe relationships between grammatical elements that are similar, but not identical, to one another. As I have shown in previous work (Zbikowski 2002, chap. 4), such relationships can be thought of as a way to organize musical syntax. Third—and of significant importance for the view of music I would like to set out—construing the basic elements of musical grammar as analogous to categories provides a way to explain how music can be properly conceptual. Although the notion of a concept has, in the past, been firmly tied to language, the various examples of analogical reference I have given in this chapter suggest that it is possible to have concepts independent of language.

In my earlier work I proposed that concepts have three defining characteristics: (a) they are cognitive constructs stable enough to be stored in memory—within the theoretical framework I developed there, such constructs are strongly associated with processes of categorization; (b) concepts are one of the key resources upon which organisms draw to guide present and future actions (understanding “actions” here to include the simulation of dynamic processes analogically related to such constructs); and (c) concepts of one sort (say, musical or related to physical movements) can be related to concepts of another sort (say, those associated with visual or linguistic domains); relationships of this kind are, of course, basic to the process of analogy (Zbikowski 2002, 61). It follows, then, that concepts need not be firmly tied to language and are indeed a feature of many nonlinguistic domains of knowledge.

Support for this perspective comes from Jean Mandler’s account of how preverbal infants build conceptual knowledge. Although it was long thought that preverbal infants were not able to have concepts, Mandler developed a theoretical framework—built on an impressive array of empirical research—that could be used to explain how infants’ demonstrated capacity for attending to events was recruited for the formation of concepts (Mandler 2004, 2008, 2010, 2012). A key element of this framework was a mechanism that redescribed, in a general and simplified way, perceptual information, which Mandler called perceptual meaning analysis (PMA). As she put it, “PMA is an attentional mechanism dedicated to simplifying spatiotemporal information. It is activated by attention to objects, especially when they move, thus emphasizing the paths that objects take through space” (2012, 426). The spatiotemporal information produced by PMA took the form of image schemas, mentioned earlier in connection with Mark Johnson’s contributions to our understanding of the part embodied experience played in
the formation of knowledge. As theorized by Johnson and others (Hampe and Grady 2005), an image schema captures a repeated pattern of bodily experience; through doing so it provides the basis for conceptual thought. According to Mandler's account, the image schemas key to the development of concepts by preverbal infants were predominantly concerned with movement through space—that is, they concerned coherent sequences distributed over time and typified by parametric modulation or change, or what I call dynamic processes. With respect to what can be gleaned of the thought processes of five- and six-month-old infants, Mandler observed, “Aside from particular interest in what people do, object motion in general attracts attention and attention activates PMA, which in turn produces a schematic summary of an object’s path in relation to other objects that is the beginning of understanding what an object is doing” (2012, 428). The first concepts that emerge, then, concern things that are in motion; attention to dynamic processes is thus central to the formation of conceptual knowledge.

For present purposes, the most important aspect of Mandler’s work is the evidence she provides that preverbal children are able to have concepts, an insight that makes the idea of nonverbal musical concepts somewhat less strange. That said, the notion that attention to dynamic processes is important for concept formation is striking, for it suggests how music’s sonic analogs may contribute to refining concepts related to objects and relations.

**Analogy, Mimesis, and Metaphor**

As noted in chapter 1, during the eighteenth century one of the dominant models for musical representation was mimesis, according to which various art forms achieved their effects by imitating the sights and sounds of nature. As should now be apparent, to the extent that mimesis obtains it relies on a cognitive capacity of central importance to human cognition: analogy. From the perspective provided by Barsalou’s perceptual symbol systems theory, we can also find support for Rousseau’s contention that sequences of musical sound did not directly represent the agitations of the sea or the flames of a blaze but instead awoke internal feelings of the same sort that had been experienced in seeing them. In Barsalou’s terms, these feelings would be simulations: reenactments of the sensorimotor states active upon the sight of the sea or the blaze. And, according to the theory I have developed here, such simulations are a direct consequence of structural correlations between the attributes of those sound sequences and the salient features of turbulent waves or flickering flames.

In previous work I organized much of my discussion of mappings between two domains around the notion of metaphor (Zbikowski 2002, chap. 2; 2008b). My shift to a focus on analogy here reflects a desire to dig a bit deeper into the mechanics of cross-domain mapping. To my mind, analogy is the result of cognitive processes that are more streamlined than those concerned with metaphor. The contrast between the two is evident in the oft-noted directionality of metaphors: it means one thing to say that the elephant is like a ballerina, and quite another to say that the ballerina is like an elephant. Both metaphors rely, of course, on an
analogy between a dancer and a pachyderm that establishes correlations between the relevant elements and relations associated with the concepts activated by those terms. As such, analogy is an essential part of the substrate of cognitive processes that are fundamental to the kind of meaning construction that is associated with metaphor (and, for that matter, metonymy and synecdoche). Perhaps most important for the perspective on musical grammar developed in this volume, analogy provides an ideal framework for understanding how sequences of patterned sound can, independent from language, create meaning and thus shape humans’ cultural interactions.
CHAPTER Three

Music and Emotion

What we know of the third-century Syrian philosopher Iamblichus is precious little, deriving chiefly from those of his writings that have been passed down through the years. Among these is his On the Pythagorean Way of Life, which gathers together what late antiquity knew about the Greek philosopher to present a model of a reflective and productive life. Iamblichus was particularly interested in the almost magical powers attributed to Pythagoras, for he believed they testified to the systematicity and comprehensiveness of Hellenic thought. A dramatic example of these powers was provided by Pythagoras’s use of music to transform behavior.

Also among Pythagoras’ deeds, it is told that once with a spondaic tune played by a piper he quelled the rage of a drunken Tauromenion youth, who made a riotous assault at night on his mistress at the gate of a rival, intending to burn the house down around her. For he was inflamed and excited by the Phrygian mode of music for the pipe, which Pythagoras then stopped as quickly as possible (he himself happened to be studying astronomy at the dead of night). And (Pythagoras) suggested to the piper a change to the spondaic, by means of which the youth, immediately restrained, went home in an orderly manner, though a little before he was not in the least patient or responsive to Pythagoras’ attempt at admonition, but even madly cursed Pythagoras’ intervention. (Iamblichus 1991, 135)

Iamblichus was hardly the first to tell this story—similar versions were related by Quintilian in the first century (1920, 1.10.32) and Sextus Empiricus in the second (1986, 131)—but he was the first to situate the tale of the Tauromenion youth within a more extensive account of Pythagoras’s works. The next chapter in On the Pythagorean Way of Life, for instance, retells the account of Pythagoras’s discovery of harmonic ratios as it was related by Nichomachus of Gerasa in the second century, an account that linked music’s pleasing effects with a group of tightly interlocked ratios between simple numbers. The power of music was thus placed side by side with its systematic organization, both attesting to the comprehensive body of thought that Pythagoras personified.

During the fourth century, at the direction of the Emperor Julian, Iamblichus’s writings came to be broadly distributed for the training of pagan clergy, but his name was most often known to later generations as the source of stories like that of the Tauromenion youth. It was, after all, Iamblichus’s version of Pythagoras’s miraculous deed—and in particular the central part played by the Phrygian mode—that was incorporated into Boethius’s enormously influential sixth-century treatise De institutione musica (1989, 5). And so we find the sixteenth-century Florentine
Music and Emotion

scholar Girolamo Mei, in a letter from May 1572, refer to the story in the course of describing to Vincenzo Galilei how ancient music affected listeners:

Because the mode was the Phrygian tonos, as the writers tell concerning this incident, and since its nature is inciting, as was the instrument on which it was played (for the pipe was held by the ancients to be such an instrument)—these together were the reason why the youth, already infuriated by the blood that was heated up around his heart, was ignited and excited even more and in a way inflamed by that exciting-sounding air. When the meter and rhythm of that measure were changed to the spondaic, which was in its general nature even and slow, with a graveness of meter and rhythm, the vigor of the excitement of that tonos was abated. (Palisca 1989, 71)

Stories like that of the Tauromenion youth, which were brought to light through the humanist revival of the fifteenth and sixteenth centuries, offered to Mei and other scholars evidence that ancient music was significantly different from the multipart madrigals of their time. Where ancient music could both inflame and cool the passions, Mei observed that “the music of our time does not work any of these miracles, since it conveys to the soul of the listener at one time diverse and contrary signs of affections as it mixes indistinctly together airs and tonoi that are completely dissimilar and of natures contrary to each other. Since each of these things has, naturally, its own quality and force, capable of stirring and moving through its resemblance appropriate affections, this music cannot generally by itself arouse any affections” (Palisca 1989, 61). The conclusion that Galilei and others drew from Mei’s researches was that modern music would have to change if it were to regain the power of ancient music. In pamphlets, treatises, and discussions among the Florentine Camerata they advocated for a new style of music that would move the passions directly. From this intellectual foment there emerged various forms of dramatic music and new life for the idea that music could have a focused and immediate effect on human emotions, an idea that has informed work on music and emotion through the present day.

The idea that music has a focused and immediate effect on the emotions is a credible one even if, as Susan McClary has argued (2004, chap. 6), it has done much to restrict our notions of what kinds of music are worthy of sustained attention. And yet it is an idea not without its complications. Consider, for instance, the aria “Schlummert ein,” which Anna Magdalena Bach transcribed early in 1727. The aria is taken from her husband’s cantata for solo voice, “Ich habe genug” (BWV 82; the opening of the recitative that follows this aria was discussed in chapter 1), and it appears in her notebook set in G major for high voice with an unfigured bass; the first eighteen measures, with a realization of the bass that draws on the music of the cantata, is given in example 3.1. The words, shaped by a melody that combines flowing lines in the upper register with repeated or held notes in the lower register over a steadily pulsing bass, have suggested to many that the aria is intended as a lullaby. As such, the function of the music of these measures is fairly straightforward—it is intended to calm and soothe a child into peaceful rest.
Foundations of Musical Grammar

Bach’s aria thus seems to be a perfect, if slightly less colorful, example of the power of music illustrated by Iamblichus’s story, calming the passions through gentle melodies and deliberate rhythms. A quick glance beyond the surface, however, suggests that the emotional range of the aria exceeds that typically accorded lullabies. In measure 13, for instance, Bach momentarily darkens the melody through a diminished seventh that points toward A minor. Whatever the uncertainties suggested by this darkening, they are set to one side by the orderly and measured arrival on tonic in measure 15. The held note that sustains this arrival continues through measure 16, but at the last moment İ collapses to ♮7, and the music pauses in wonder. There are any number of ways to interpret this gesture,
but at the very least Bach suggests that there is unfinished business, which is finally brought to a conclusion by the forthright cadence of measures 17–18.

While the diminished seventh leap of measure 13 and the pause on ♯7 in measure 15 are, in themselves, small disruptions of the placid surface of the aria, they are part of a larger strategy of phrase expansion that gives considerable weight to the final cadence of measures 17–18. In the first phrase (measures 1–4) Bach sets out his text in a quite straightforward way. The second phrase begins just a bit early, interrupting the half cadence of measure 4 with a gesture that is both persuasive and reflective: the two statements of “schlummert ein” that stretch across measures 5–6 almost seem like an aside, the long syncopated durations and repeated notes unlike anything heard in the aria up to this point. The impression
that these measures are a suasive interjection is strengthened by the return to the music of the opening phrase in measure 7, which both contrasts with and seems a consequence of the meditations of measures 5–6. The long held note that begins the third phrase in measure 11 acts as a return to these meditations, which have as their sequel the more poignant gestures of measures 13 and 16 and the ultimate resolution provided by measures 17–18.

The subtle intensity of these measures might be dismissed as a consequence of compositional artifice, one that colors but does not substantially affect its function as a lullaby, but such an approach would do a disservice to the meaning of the aria. This meaning is, of course, borne not only by the music of the aria but also by its words. The complete text fills out the context for the soothing music of measures 1–18:

| mm. 1–18, 31–48 | Schlummert ein, ihr matten Augen, | Slumber now, you weary eyes, Fallet sanft und selig zu. | Close softly and pleasantly! |
| mm. 19–30 | Welt, ich bleibe night mehr hier, Hab ich doch kein Teil an dir, Das der Seele könnte taugen. | World, I shall not remain here any longer, I own no part of you That could matter to my soul. |
| mm. 49–67 | Hier muß ich das Elend bauen, Aber dort, dort werd ich schauen Süßen Friede, stille Ruh. | Here I must build up misery, But there, there I will see Sweet peace, quiet rest. |

Bach sets the second sentence of the text with sharply contrasting music in measures 19–30, using irregular rhythms and tonal peregrinations to capture the protagonist’s rejection of earthly life. This fraught music calls into question the peaceful environment established by measures 1–18, and it leads (in measure 30) to a cadence on the relatively remote key of B minor. This is followed immediately by a reprise of the music of measures 1–18, which, with one exception, returns intact. The exception is that the voice pauses after little more than a measure and, as shown in example 3.2, remains tacet until returning to its meditations on “schlummert ein” at the beginning of the second phrase. The effect is a magical one, summoning both the point when a restless infant, gentled by lullaby, begins to fall asleep and the image of a singer rendered mute by the course of her thoughts. Further contrasting music in measures 49–67 sets the third sentence of the text, which is then followed by a da capo return to the words and music of measures 1–18. The contrasting music that sets the second and third sentences of the text does nothing to diminish the calming effect of measures 1–18, but it does deepen our understanding of their function: the reward of the sleep that they summon is not simply rest, but the promise of awakening in
God's embrace. The lullaby, if we still wish to call it that, is not for a restive child but for the singer's impatient soul.

The range of emotions encompassed by Bach's “Schlummert ein” and the way these emotions are shaped over the course of the aria suggest that our reactions to music are not nearly as immediate as implied by Iamblichus's story, nor as focused as supposed by members of the Florentine Camerata. As often as not, our emotional reactions to music develop over time, affected not only by the succession of musical materials but also by relationships among these materials, by the influence of other communicative media with which they may be coordinated,
and by the larger situations of which our encounters with music are a part. It also bears mention that music is not unique among communicative media in shaping or arousing emotions. That said, the insubstantial character of sonic analogs makes the connection between music and emotions appear to be immediate: both seem to have their proper home within our internal experience, rather than in the world without.

One of the assumptions that guides this chapter is that music can indeed shape the emotions of those who attend to it, and that this is one of its principal functions in human cultures. What is of particular interest, from a construction grammar perspective, is how musical materials are arranged to achieve this end. To provide a framework within which to address this question, the first section that follows offers a review of recent research on human emotions that will give a sense of how emotional responses have been theorized and studied. Two things that will soon become apparent are that emotional responses are highly dynamic (and, indeed, using one term for these responses—“emotions”—is quickly revealed to be inadequate) and distinguished by their predominantly internal, subjective character. The second section then uses the perspective provided by this review to evaluate theoretical and empirical work on music and emotion that has been done over the past few decades, with particular attention to evidence for the claim that music can directly induce emotional responses (as Iamblichus would have us believe, and as many more recent researchers would maintain). In the third, and longest, section I shall develop the proposal that music’s sonic analogs serve to simulate (in Barsalou’s sense), rather than directly induce, emotional responses. Much of my argument here will be made through musical analysis—of the recitative by Simeon that follows “Schlummert ein” (discussed briefly in chapter 1) and then of the aria itself—built on an expanded account of musical syntax. As noted in chapter 1, according to the approach I develop in this volume, the basic elements of musical grammar are organized through syntactic processes that are themselves analogs for dynamic processes. My analyses of the movements from Bach’s “Ich habe genug” will, with respect to the “form” portion of the form-function pairs basic to musical grammar, provide illustrations of syntactic processes as well as of interactions among multiple syntactic layers and, with respect to the “function” portion of grammatical pairs, suggest how such forms can realize the function of representing aspects of emotion processes. By way of conclusion, the final section steps back from the specifics of these two movements to consider how, within liturgical settings of the kind that motivated and framed Bach’s cantata, music is used to shape the emotional responses of the congregation to realize the larger goals of communal worship.

It bears mention that my approach to the relationship between music and the emotions is, in certain respects, similar to that of Susanne Langer, who many years ago proposed that musical structures reflect the morphology of emotions (Langer 1957, 238; see also Zbikowski 2010, 37–38, 42–43). My approach is different, however, in two important ways: first, I assume that, while musical structures can reflect the morphology of emotions, they can also represent the
morphology of various other kinds of dynamic processes. Second, where Langer’s proposal was part of a larger argument about the nature of knowledge and symbolic structures, mine is framed around recent research from the mind and brain sciences, with the aim of better understanding the basis and goals of musical organization—that is, what relationships between music and the emotions can tell us about musical grammar.

Recent Research on Emotion

*Emotion and Cognition, circa 1980*

In a wide-ranging paper published in 1980, the social psychologist Robert Zajonc advanced a hypothesis that many of his peers regarded as far-fetched: Zajonc proposed that under certain circumstances affective dispositions could be established prior to and separately from cognitive processes. As evidence for this hypothesis, Zajonc marshaled a variety of empirical evidence, including an experiment he and William Kunst-Wilson had performed showing that subjects developed preferences for stimuli of which they had no conscious recognition (Kunst-Wilson and Zajonc 1980). In the first part of that experiment, Zajonc and Kunst-Wilson presented their subjects with ten images of randomly constructed polygons. The images were each shown five times, in random order, and for a duration of one millisecond, well under the threshold for recognition. In the second part of the experiment, subjects were asked to compare each of the polygons to which they had been exposed with similar, but novel, polygons and indicate whether they recognized one of the polygons and which they liked better. Although subjects’ judgments about which polygon they had been previously shown never rose above the level of chance, they showed a significant preference for the polygons to which they had been already exposed. Zajonc suggested that this showed that subjects had developed an affective disposition toward the polygons to which they had previously been exposed even though they could not remember having seen them before.

Zajonc’s hypothesis was certainly provocative, but it was also framed in terms of a model of cognitive processing that other psychologists were beginning to call into question. According to that model, perceptual information was transformed into amodal symbols that served as the basis for thought processes that led to the evaluations reflected in affectual dispositions. (Barsalou’s perceptual symbol systems theory, discussed in chapter 2, is a response to this same basic model of cognitive processing.) Because emotions emerged rather far downstream in this model, it could not readily account for the experimental results gathered by Zajonc. For his part, Zajonc wanted to move toward a model that would allow affect into the game much earlier: “What I want to argue is that the form of experience that we came to call feeling accompanies all cognitions, that it arises early in the process of registration and retrieval, albeit weakly and vaguely, and that it derives from a parallel, separate, and partly independent system in the organism” (1980, 154).
While other psychologists were aware of inadequacies in the information-processing model of cognition, they were not quite ready to accept the model proposed by Zajonc: the notion that there could be emotions prior to cognition seemed to be little more than nonsense. Prominent among Zajonc’s critics was Richard Lazarus, a clinical psychologist who had long argued for—and demonstrated through his research—the role of appraisal in emotions. As he made clear in his 1982 response to Zajonc, Lazarus viewed cognitive appraisal as central to emotion: “Its [i.e., emotion’s] role is to mediate the relationship between the person and the environment. The appraisal process gives rise to a particular emotion with greater or lesser intensity depending on how the relationship is evaluated with respect to the person’s well-being” (1982, 1019). For Lazarus, then, there was no emotion without some sort of appraisal. Lazarus argued that Zajonc, having framed the issue in terms of a grossly inadequate model of cognitive processing (which reduced everything to amodal symbols), had no way to account for appraisal. And without appraisal, there was no way to explain how humans created meaning. As Lazarus saw it, “Humans are meaning-oriented, meaning-creating creatures who constantly evaluate events from the perspective of their well-being and react emotionally to some of these evaluations” (1982, 1020). Emotions, rather than being prior to a cognitive process like evaluation, were instead a manifestation of this process.

Although Lazarus was able to make quite clear weaknesses in Zajonc’s argument, the view of the relationship between cognition and emotion he offered (both in his initial response and in a further rebuttal to Zajonc in Lazarus 1984) was ultimately rather diffuse. Three problems hampered both Zajonc and Lazarus in setting out their respective positions. First, there was the matter of whether the affectual dispositions on which Zajonc focused should be regarded as emotions proper or as rather underspecified responses related to but not definitive of emotion states. (The issue was complicated by Zajonc’s willingness, as the debate went on, to allow the term “emotion” to be applied to what he had previously called “affect”; see Zajonc 1984.) Second, it was unclear to what extent the phenomena with which Zajonc and Lazarus were concerned were accessible to consciousness, or what part consciousness played in emotions. Although both acknowledged the problem in their colloquy of 1984, neither offered a way out of the impasse. Finally, both Zajonc and Lazarus approached the relationship between cognition and emotion from the perspective of contemporary psychology. Although perfectly in keeping with their training and research, this perspective did not allow them to take full advantage of burgeoning work in neuropsychology and neuroanatomy, nor to make more than passing suggestions how such work might inform their debate.

Research on Emotion and Cognition Subsequent to the 1980s Debate

The years that followed these exchanges bore witness to a significant expansion in work on emotion, which provided researchers with ways to address the problems
that clouded the debate between Zajonc and Lazarus, and also ramified that
debate in interesting and unexpected ways. For instance, there is a substantial
body of research which suggests that human emotional life is organized around a
limited number of basic or primary emotions that are not a product of appraisal.
Although the size and membership of this category varies, the list offered by Paul
Ekman may be taken as typical, and includes anger, fear, sadness, enjoyment,
disgust, and surprise (Ekman 1992a; see also Izard 1971, chap. 6; LeDoux 1996,
chap. 5, Damasio 1999, chap. 2). One of the principal motivations for the notion
of basic emotions came from research by Ekman and others on the facial expres-
sions associated with emotions. This research demonstrated both that there were
specific expressions associated with particular emotion states and that there was
significant cross-cultural agreement on how such expressions should be labeled
(Ekman 1992b). These findings conformed with ideas set out by Charles Darwin
in a classic study (1872), in which emotions were viewed as evolutionary adap-
tations shared by members of a species. The various emotions that correlated
with distinctive facial expressions were thus thought of as products of innate sys-
tems that were in some measure independent from social or cultural learning.
Although the complexity and adaptability of human emotions suggest that innate
systems are but one part of a larger picture, such systems support the notion of
evolutionary continuity, a notion that connects human emotions with those of
other animals (Panksepp 2005a, 2005b). Basic emotions, then, were conceived
of as part of an organism’s response to changes in its immediate environment, and
as such were rapid and thoroughly interconnected with its physiology; they were
thus very much like the affectual predispositions discussed by Zajonc. Although
appraisal was still considered to be part of the process through which emotions
were induced, researchers distinguished between the extremely rapid appraisal of
stimuli that gave rise to basic emotions and a more extended process of appraisal
in which emotion processes were subjected to slow, deliberate, and conscious
evaluation (Ekman 1992a, 187), a view that was substantially adopted by Lazarus
in later work (1991, chap. 5; for an overview of more recent work on appraisal,
see Scherer 2001).

The exploration of basic emotions can be seen as part of a more comprehensive
reconsideration of the evolution of the human cognitive system, a topic that cogni-
tive science had, for the most part, neglected. In the course of this reconsideration,
a number of neuroscientists turned their attention to an aspect of human cognition
long thought to be distinctive of the species and one also implicated in the expe-
rience of emotions: consciousness. There emerged various accounts of the brain
systems involved in consciousness, which typically viewed human consciousness
as a special case of a cognitive phenomenon broadly distributed among vertebrates
(Edelman 1989, 2003, 2006; Crick and Koch 1998; Donald 2001; for a review, see
Morin 2006). Most of these studies did not address the topic of emotions directly,
although they recognized overlap between the brain systems associated with con-
sciousness and those associated with emotions. An alternative was provided by
Antonio Damasio, whose research was grounded in clinical experience with a
variety of neural disorders and who used humans’ capacity for emotions as one
of the foundations for his account of consciousness (Damasio 1994, 1999, 2003,
Building on a detailed knowledge of neuroanatomy, evidence from neuroimaging, and lesion studies, Damasio made a compelling argument for the role of emotion in human thought processes. For Damasio, emotions are complicated collections of chemical and neural responses that form a pattern; their function is to create (either directly or indirectly) circumstances that are advantageous to the organism (1999, 51). He contrasts emotions with feelings, which he regards as neural images of the sensory patterns associated with emotions. (For similar and contrasting views on the relationship between emotions and feelings, see LeDoux 2015 and the essays collected in Manstead, Frijda, and Fischer 2004.) The neural images associated with feelings are incorporated into a representation of the state of the physical structure of the organism in its many dimensions (what Damasio calls the “proto-self”), and subsequently they become available to consciousness (Damasio 1999, 283). By this process, the full efficacy of emotions (as biological responses to the environment) becomes available to the organism: “Consciousness allows feelings to be known and thus promotes the impact of emotion internally, allows emotion to permeate the thought process through the agency of feeling” (1999, 56). Although Damasio’s approach does not resolve all of the issues raised by connections between emotions and consciousness (not the least because there is still spirited debate on the extent to which consciousness is informed by emotions; see, for instance, Rolls 2000, Panksepp 2005a, and LeDoux 2015) it does provide the tools that Zajonc and Lazarus found wanting when they confronted the part consciousness played in emotions.

Key to Damasio’s work were refinements in the understanding of neuroanatomy and neurophysiology of emotions. Although work in this area had begun in earnest around the time of Zajonc and Lazarus’s initial exchange (Panksepp 1982; see also Panksepp 1998) it was only through focused studies of the sort conducted by Joseph LeDoux and his associates that specific brain circuits associated with specific emotion responses began to be identified (LeDoux 1996, 2000). As researchers have come to appreciate, the neurophysiology of emotions is enormously complicated: emotions make use of numerous intertwined systems within the brain and central nervous system, involve complex biochemical processes, and bear witness at every turn to the thoroughly dynamic nature of the human brain and the multifarious ways it interacts with the body (Lane and Nadel 2000; Damasio 2003; Colombetti 2014). Indeed, if research on emotion has revealed nothing else it is that the dualism of mind and body prevalent in Western thought for centuries (and, as discussed in D. Clarke 2003, somewhat inaccurately attributed to Descartes) cannot be sustained in any principled inquiry into human cognitive processes.

Central Issues in Emotion Research

The issues engaged by contemporary research on emotions are various, and few can be regarded as settled. As Klaus Scherer suggested in a review of psychological models of emotion, however, research has converged around a few key points. To begin with, emotions are most typically regarded as episodic and processual: some
event, which may be external or internal to the organism, triggers a noticeable change in the way an organism functions (Scherer 2000, 137–38). This change in function is thought to have a number of components, including what has been called the “response triad” of emotion: physiological arousal (for instance, changes in respiration and temperature sensation), motor expression (as evidenced by the facial expressions studied by Ekman and others), and subjective feeling. Drawing together these ideas, Scherer offers this definition: “emotions are episodes of coordinated changes in several components (including at least neurophysiological activation, motor expression, and subjective feeling but possibly also action tendencies and cognitive processes) in response to external or internal events of major significance to the organism” (2000, 138–39).

Something assumed by Scherer and other researchers on emotion, but not explicit in his definition, is that emotional responses are always valenced: an emotional response is either positive or negative, never neutral (Ortony and Turner 1990). Making this assumption explicit highlights one thing left out of the definition offered in Scherer’s review, which is the adaptive function of emotion responses: in that definition, emotional responses simply lead to change, but such responses are better thought of as having the preservation of the organism as their goal (something evident in the component process model summarized in Scherer 2009). Damasio, for his part, proposed that the function of emotions is to achieve homeostasis, with homeostasis understood not simply as a neutral state midway between life and death but as an active state of well-being (Damasio 2003, 35; 2010, 51–53). And while a rudimentary componential approach is productive for thinking about the different systems involved with emotions, it should be emphasized that the component most directly linked to consciousness—subjective feeling—is deeply intertwined with the embodied processes of neurophysiology and motor expression. As but one demonstration of this, research by Ekman and his colleagues (which I noted in chapter 1) showed that when subjects simply adopted the facial expression specific to one of the basic emotions (not by being asked to pose the emotion but by following muscle-by-muscle instructions to create the expression on their own faces) they reported experiencing the emotion associated with the expression (Ekman, Levenson, and Friesen 1983; Ekman 1992b; for further discussion of relationships between bodily experience and emotion, see Colombetti 2014, 163–70).

Another point made by Scherer in his review was that the term “emotion” was often applied rather indiscriminately. That is, differences in the duration of emotion episodes and the interaction among various response components are such that to use one term for the entire range of responses is, at best, misleading. Drawing on the framework provided by his episodic and componential account of emotion, Scherer suggested that it would be better to distinguish between emotions and other affective phenomena on the basis of what he called design features. These features offer a way to distinguish between different sorts of affective responses on the basis of a number of parameters, including the intensity and duration of the response and its significance for the organisms’ interaction with its environment. Under this analysis, “emotions” are intense phenomena of relatively short duration in which the changes among response components
are tightly synchronized, and in which the organism is focused on the triggering event and its consequences for behavior. “Moods,” by contrast, are somewhat less intense phenomena with a noticeably longer duration but with less synchronization among response components, and in which the organism is less focused on the triggering event and its consequences for behavior (Scherer 2000, 140–41; slightly expanded arrangements are offered in Scherer and Zentner 2001 and Scherer 2004). The goal of Scherer’s taxonomy of affective phenomena was both to focus research and to understand the limits of emotion theories: the preferences upon which Zajonc’s 1980 study focused are thus regarded by Scherer as affective phenomena different in kind and function from the emotions with which Lazarus was concerned in his 1982 response. Put another way, research on emotion processes over the past thirty-five years has demonstrated the range and complexity of affectual responses. If researchers were to make any headway through this terrain, Scherer believed they needed to specify in greater detail the phenomena they aimed to study.

Social Function, Culture, and Emotion

Although the bulk of the research on emotion has focused on the responses of individual subjects, most who work on emotion assume that one of the functions of these responses is to help humans solve the challenges of social living. By far the most prevalent approach takes the position that emotions evolved to guide behavior in social groups (Ekman 1992a; Lazarus 1991, chap. 2; Plutchik 1980), and that the interactions that occur within such groups refine the patterns of response associated with basic emotions (a process that yields what Ekman 1992b called “display rules”). A rather different approach takes the social function of emotions as their defining attribute—that is, what we call “emotion” is always framed relative to some social context—and that the ensemble of responses associated with any given basic emotion is simply too undifferentiated to account for the way emotions regulate social interactions (Averill 1980). This perspective has been important for those interested in the cultural interpretation of emotion, not least because it suggests that culture plays a prominent role in the construal of emotion.

Dacher Keltner and Jonathan Haidt proposed that these two approaches could be reconciled by assigning them to different stages within the development of social structures. Basic emotions—which Keltner and Haidt called primordial emotions—belong to an early stage of this development, evolving as solutions to the problems of physical survival, reproduction, and group governance (Keltner and Haidt 2001, 195). The various behavioral manifestations of primordial emotions, including but not limited to facial expressions, then provided a means to structure social interaction in one of three basic ways: first, to evoke complementary emotional responses in others; second, to provide information about the mental states, dispositions, and intentions of the individual manifesting the emotion; and, third, to serve as incentives or deterrents for others’ social behavior (Keltner and Haidt 1999, 511; Keltner and Haidt 2001, 199–200; see also Fischer and Manstead 2008). Primordial emotions are subsequently shaped by various forms of social
Music and Emotion

interaction, as well as by concepts of the self, morality, and social order, to yield what Keltner and Haidt call elaborated emotions. As a product of a later stage in the development of social structures, elaborated emotions vary across cultures, cannot be experienced by infants, and are sustained through various social and cultural practices (Keltner and Haidt 2001, 200). Although primordial emotions may well connect us with our distant forbears, in our day-to-day interactions with others we are more typically concerned with the fine-tuned and subtly graded responses that mark elaborated emotions.

Keltner and Haidt’s proposal offers a way to resolve the tension between the construal of basic (or primordial) emotions as focused on the survival of the individual and social emotions as concerned with the survival of the individual as part of a social group of conspecifics (a tension evident in Damasio 1999, chap. 2; 2003, chap. 2): from their perspective, basic emotions are not strictly about the individual, but also about the behavior of the individual in groups. Keltner and Haidt’s framework does not, however, provide a way to analyze where basic or primordial emotions leave off and elaborated emotions begin. The challenge of locating the boundary is made all the more difficult by the rapidity of evaluative processes: emotional responses that seem automatic often prove to be shaped and conditioned by cultural knowledge, and are thus anything but innate (Mesquita 2003). In view of this, a number of researchers interested in the cultural interpretation of emotion have argued that a unitary conception of basic emotions is simply inadequate for cross-cultural comparisons (Lutz and White 1986; Mesquita and Frijda 1992; Mesquita and Markus 2004; Shweder 2004). In comparing a third-century CE Indian account of the organization of emotions given in the “Rasādhyyā” of the Nātyaśāstra with a list of basic emotions given by Paul Ekman (1984), for instance, Richard Shweder and his associates found only three emotions in common: anger, fear, and sorrow (Shweder et al. 2008). Differences between the two lists of emotions—and, for that matter, even the characterizations of anger, fear, and sorrow—suggested that it might be difficult if not impossible to develop a list of basic emotions applicable to all cultures. Shweder and his associates proposed that it would be more productive to analyze the various components of the emotional responses that help individuals negotiate their behavior in cultural contexts. The components on which they focused include the somatic and affective experiences associated with a given response, the environmental determinants and consequent appraisals of the significance of these events, and the normative social appraisals and communicative functions of the response. Analyses guided by this approach provide a basis for cross-cultural comparisons of emotional responses (Horton 2006), revealing what is similar and what is different in responses that might, from a North American perspective, be characterized with the same basic emotion term.

Research on the social function and cultural interpretation of emotional responses is part of a much larger debate about what counts as an emotion, and the extent to which emotional responses are shaped by culture and, in at least some cases, constitutive of culture. Although this debate can only be noted here, it is well worth keeping in mind as we turn to music, for music is nothing if not a product of human cultures. Such cultural products may certainly exploit the
responses associated with basic or primordial emotions, but the interpretation of these responses—both by individuals and within cultural groups—must be seen as manifestations of the function of emotions within social and cultural settings.

**Emotion and Cognition**

Although it may still be convenient for some researchers to distinguish between emotion and cognition (as Zajonc attempted to do in his 1980 article), most evidence suggests that if there is a line between the two it is a very fuzzy one. Part of this evidence comes from work in neurophysiology, which has shown that the neural systems through which emotions operate are thoroughly integrated into brain structure as a whole. While it does appear that many of the brain areas involved with emotions developed at an earlier point in the evolution of the mammalian brain, in humans these areas are richly interconnected with brain structures that are specific to our species: it is difficult, if not impossible, to isolate the areas involved with emotions from those involved with higher-order cognitive processes. The social function and cultural interpretation of emotions further complicate our understanding of relationships and boundaries between emotion and cognition: while the notion of basic or primordial emotions is one that offers a number of theoretical advantages (not least because it can be used to explain similarities in emotional responses across different species), we really only know human emotions in their socially elaborated forms. Put a slightly different way (and adopting terms proposed by Batja Mesquita), while there is good reason to see the physiological and psychological mechanisms behind emotional responses as part of our evolutionary endowment, these mechanisms represent only a potential for emotion. By contrast, our experience of emotion is always with emotion enacted—that is, with the practice of emotion (Mesquita 2003; Mesquita and Markus 2004). And so, while a broader understanding of emotion demands that we look beyond (or beneath) higher-order cognitive processes and to the biological functions of affectual responses, we must keep in mind that it is higher-order cognitive processes that shape, direct, and ultimately make coherent the emotions that are so central to our experience.

**Recent Research on Music and Emotion**

**Meyer’s Critique of Research on Music and Emotion**

From the vantage point available to him in 1956, Leonard Meyer had the opportunity to survey nearly half a century of research on music and emotion. Although this research seemed to support traditional thinking on the close relationship between music and emotion, Meyer viewed it as ultimately inconclusive, not the least because it was unclear whether music actually induced emotions or whether emotion states were simply ascribed to music.
Any discussion of the emotional response to music is faced at the very outset with the fact that very little is known about this response and its relation to the stimulus. Evidence that it exists at all is based largely upon the introspective reports of listeners and the testimony of composers, performers, and critics. Other evidence of the existence of emotional responses to music is based upon the behavior of performers and audiences and upon the physiological changes that accompany musical perception. Although the volume and intercultural character of this evidence compels us to believe that an emotional response to music does take place, it tells us almost nothing about the nature of the response or about the causal connection between the musical stimulus and the affective response it evokes in listeners. (Meyer 1956, 6)

For Meyer, the key to explaining the causal connection between musical stimulus and affective response was found in psychological theories that held that emotions or affects are aroused when a tendency to respond is arrested or inhibited (1956, 14). When listeners acquainted with a musical style encounter a passage of music in that style, they will have expectations for what will follow the passage. The salient part of the listener’s response to the music begins with this sense of expectation. When expectations of this sort are realized, there are few emotional consequences, but when they are denied—that is, when the listener’s response is arrested or inhibited—profound emotional experiences result. As Meyer saw it, explaining the relationship between music and emotion needed to begin not with the problematical responses of experimental subjects but with a psychologically informed analysis of musical structure.

For a good thirty years after Meyer wrote, research on music and emotion languished. John Sloboda, reflecting on some of the challenges that faced research on the emotional aspects of music, suggested that Meyer’s critique of empirical work caused researchers to focus on studies of musical structure rather than on behavioral responses to music (Sloboda 1991, 110). It should also be said that, from the perspective of research in psychology, the model for emotional response offered by Meyer was somewhat limited. As David Huron has shown (2006), expectation accounts for only one small—if quite important—part of humans’ emotional responses to music. Finally, shortly after the time Meyer wrote, there was a turn toward highly formalized study in music theory, often framed by a strongly positivistic methodology. This approach to musical organization stood apart from the richly humanistic perspective that framed Meyer’s work, having little to say about emotional responses to music.

Recent Research on Music and Emotion

Whether due to the critical distance afforded by the passage of time, fresh energy from a new conception of how research on music psychology and music cognition might be conducted, or a broadened conception of the aims of music theory, work on music and emotion resumed in earnest during the 1990s. Since then hundreds, if not thousands, of studies on the topic have been published. And yet the situation
is not markedly improved from the time when Meyer wrote—as Patrik Juslin and Daniel Västfjäll have remarked, contemporary research on music and emotion presents “a confusing picture with conflicting views on almost every topic in the field” (Juslin and Västfjäll 2008, 559). Juslin and Västfjäll, for their part, argued that the reason that more progress has not been made is that researchers have failed to focus on the psychological mechanisms that underlie listeners’ emotional responses to music. Before considering a few of the issues raised by their argument, I should first like to turn to two methodological problems that I believe are an even more significant hindrance to understanding the emotions induced by listening to music: the assessment of physiological responses to music, and the relationship between musical and basic emotions.

Assessing Physiological Responses to Music. In a study published in 1991 John Sloboda sought to confront the challenge presented by Meyer’s critique of empirical research on musical emotions by exploring the frequency and nature of listeners’ physical responses to music. His principal evidence came from a questionnaire that asked respondents about various physical phenomena they experienced while listening to music during the previous five years. He found that nine out of ten reported experiencing shivers down the spine at some point during this period (Sloboda 1991, 112), which, for Sloboda, demonstrated that music could directly induce a quantifiable physical response that is independent of appraisal.

A few years later Ivan Nykliček and his associates set out to explore this sort of phenomenon through a set of experiments that included direct measurements of physiological responses while subjects listened to music. In the first stage, they had a group of subjects listen to twenty-five musical excerpts and rate, on a five-point scale, how well each excerpt expressed one of twelve emotions (Nykliček, Thayer, and Van Doornen 1997). The experimenters then selected twelve excerpts with the most distinctive ratings and put them into four groups of three excerpts each, based on a factor analysis of level of arousal and emotional valence. According to this arrangement, each group of excerpts was taken to represent one of four basic emotions: happiness, sadness, serenity, and agitation. In the second stage of the experiment, Nykliček and his associates recorded data from sixteen cardiovascular and respiratory variables while a second group of subjects listened to the excerpts. In their analysis of this data they found that, while it was not possible to distinguish among the groups of excerpts on the basis of individual physiological variables, it was possible to draw distinctions between the groups on the basis of the profile of physiological responses taken as a whole: the musical excerpts chosen to represent happiness, for instance, induced a different set of physiological responses than the excerpts chosen to represent serenity, sadness, or agitation.

Results similar to those of Nykliček and his associates were reported by Carol Krumhansl in an exploratory study carried out around this same time (Krumhansl 1997). In Krumhansl’s experiments, subjects listened to six excerpts associated with three different emotions (fear, sadness, and happiness) while data from eleven physiological variables were recorded. Analysis of the data showed correlations between cardiovascular physiological changes and subjectively reported experiences of emotions, as well as differences between the profile of physiological
changes induced by the musical excerpts associated with the different emotions. For example, listening to Samuel Barber’s *Adagio* for strings, which Krumhansl’s subjects associated with sadness, produced a physiological response distinct from that produced when listening to excerpts associated with fear or happiness.

More recently, Charlotte Witvliet and Scott Vrana offered a different view of physiological responses to music based on the facial expressions associated with emotional responses. When we encounter something pleasant, for instance, muscles in the zygomatic region of our face (which are used when we smile) become active. When we encounter something unpleasant, muscles in the corrugator region (which are used when we frown) become active. In their study, Witvliet and Vrana collected electromyography data from these regions while subjects listened to musical excerpts chosen to demonstrate different combinations of level of arousal and emotional valence (Witvliet and Vrana 2007; they also collected heart-rate data and data from muscles associated with eyelid closure). Although Witvliet and Vrana were primarily interested in the effect of repeated listening to musical excerpts—in particular, whether subjects would continue to respond affectively to music with repeated exposure—their results also showed correlations between subjects’ physiological responses to music and different groups of musical excerpts. Thus excerpts selected for positive valence and high arousal evoked markedly different physiological responses than did excerpts selected for positive valence and low arousal, negative valence and high arousal, or negative valence and low arousal.

Although the results reported by Nyklíček and his associates, by Krumhansl, and by Witvliet and Vrana indicate that listening to music can induce physiological changes, the correlation between these changes and emotions remains a matter of speculation (Larsen et al. 2008). As but one example, while the activation of muscles in the corrugator region studied by Witvliet and Vrana can indicate a response to an unpleasant stimulus, it can also indicate effort or concentration, neither of which are necessarily unpleasant. Indeed, there is some evidence that physiological changes may be a consequence of an engagement with music that goes well beyond basic physiological mechanisms. Jaak Panksepp, in a series of experiments prompted in part by the prevalence of “shivers down the spine” reported in Sloboda’s 1991 study, found that the majority of his subjects experienced chills when listening both to music he provided and to music they supplied. One song in particular, “Making Love Out of Nothing at All” by the group Air Supply, yielded a mean of 4.4 chills per minute in a group of fourteen students (Panksepp 1995, 183). When Oliver Grewe and his associates attempted to reproduce Panksepp’s results, however, they found that this same piece produced chill responses in fewer than 10 percent of their subjects; among those who did experience chills, it was at the rate of 1.1 per minute (Grewe et al. 2007, 309). This difference in results, together with other evidence from their study, led them to propose that chills were not a direct physical response to music so much as they were a consequence of the ways listeners use music to influence feelings. (It bears mention that this difference could also be seen to reflect the shaping influence of cultural context on the interpretation of music: the participants in Panksepp’s study were college-age individuals in the United States; the participants in the study by Grewe and his colleagues, by contrast, ranged in age from eleven to seventy-two years of age and lived around Hanover, Germany.)
The lack of precision in all of these assessments of physiological responses to music, together with the markedly different results obtained by Panksepp and by Grewe and his colleagues, suggests that we are far from a state of certainty about the physiological changes associated with listening to different types of music, or about the correlation of such responses with emotions. A similarly tepid assessment of research on physiological responses to music and sound stimuli was offered by Dale Bartlett who, after reviewing studies conducted over a 120-year period, concluded that the most that could be said was that music did in fact have an influence on bodily systems and that this influence was either stimulative or sedative (D. Bartlett 1996).

What is to my mind a more serious methodological problem, and one not limited to studies of physiological responses to music, is that none of the studies I discussed made an attempt to distinguish between the responses induced by listening to music and the responses induced by listening to nonmusical sound stimuli. The distinction is an important one, for without evidence to the contrary we have no way of knowing whether the observed responses are specific to music. Research by Margaret Bradley and Peter Lang seems to argue against just this sort of specificity. As part of a study of affective reactions to both visual and acoustic stimuli, Bradley and Lang had subjects listen to sixty nonmusical sounds while they recorded physiological data very similar to that gathered by Witvliet and Vrana (Bradley and Lang 2000, 2007). The results were almost exactly identical: crying babies, growling dogs, chirping birds, and lowing cows induced the same patterns of response in facial muscles and heart rate as have been reported for music. A later study by Patrick Gomez and Briggita Danuser adopted the basic methodology used by Bradley and Lang to explore more directly differences in physiological responses to environmental noises and musical passages (Gomez and Danuser 2004, 2007). Although some differences emerged, these appeared to be related principally to the resources music offers for processes of entrainment. Such results suggest that, without a robust program of research that makes direct comparisons between responses to environmental sounds and music, we have no way of knowing whether music is substantively different from nonmusical sound stimuli in the physiological or emotional responses it induces (a point also made by Sloboda; see Sloboda 2001–2, 245).

Music and Basic Emotions. The apparent immediacy of emotional responses to music and evidence of general agreement among subjects about the emotions induced by a given musical passage have led a number of researchers to the assumption that music can induce basic emotions (Panksepp and Bernatzky 2002; Juslin and Laukka 2003; Bigand et al. 2005). In truth, if one takes the position that basic emotions reflect evolutionary adaptations, the assumption is a rather odd one, since the time scale for the emergence of such adaptations is several orders of magnitude larger than any proposed for the development of music in the human species. It seems more likely that a broad range of sound stimuli can induce basic emotions (stimuli to which our evolutionary ancestors would have been exposed), and that works of music simply recruit aspects of these stimuli. As noted above, evidence for basic emotions is drawn from differential responses of the autonomic nervous system (Ekman, Levenson, and Friesen 1983). While the studies by Nykliček and his associates and by Krumhansl have shown that music can also induce such differential responses, links between those responses and emotional
states are at best tentative, and the responses have not been shown to be substantially different from those induced by nonmusical sound stimuli.

Most of the evidence for the induction of basic emotions by music has come from categorization tasks. For instance, subjects have been asked to sort musical excerpts into categories identified by words used for basic emotions (Terwogt and Van Grinsven 1988; Dolgin and Adelson 1990; Balkwill, Thompson, and Matsunaga 2004), by movements associated with basic emotions (Boone and Cunningham 2001), or by photographs of facial expressions associated with basic emotions (Kallinen 2005). The premise here, as it is in other work on basic emotions, is that subjects’ ability to perform such tasks accurately reflects an intuitive grasp of discrete emotion states (Ortony and Turner 1990; J. Russell 1994). There is, however, a healthy debate about whether basic emotions represent what Lisa Feldman Barrett calls “natural kinds” (Barrett 2006a, 2006b; Barrett et al. 2007; Izard 2007; Panksepp 2007). In brief, a natural kind is a grouping or category of things that occur in the world, the boundaries of which are given by nature and discovered, not created, by the human mind (Barrett 2006a, 29). Barrett notes that, for the most part, “scientific treatments of emotion are founded on the view that certain emotion categories (such as anger, sadness, fear, disgust, and happiness) carve nature at its joints” (2006a, 30; emotion categories of this sort bear a strong resemblance to the basic-level categories discussed in chapter 2). The cultural perspective on emotion I discussed earlier argues against this view, as does the empirical evidence assembled by Barrett. Indeed, it could be said that the challenges presented by music, as a complex cultural product, highlight the limitations of the basic emotion approach. This is one way to understand Klaus Scherer’s reservations about the applicability of the model of basic emotions to music: as Scherer has noted, emotional responses to music are simply richer and more varied than what can be captured by six or seven basic emotion terms (Scherer 2003, 2004).

A further problem concerns the way basic emotion terms are used to characterize responses to music. In all of the studies of music and basic emotions that I mentioned, subjects were asked to categorize the music according to which emotion the music expressed. There is, however, a significant difference between the emotions that are attributed to some thing or set of events and actually experiencing those emotions: as the philosopher Peter Kivy has observed, characterizing a melody as expressing sadness does not necessarily mean that listening to it makes the listener feel sad (Kivy 1990, chap. 8).

In sum, there is good evidence that a wide range of individuals—some as young as four years of age—can use basic emotion terms to categorize musical excerpts, and that these same individuals demonstrate broad agreement about which excerpts belong in which basic emotion categories. There is, by contrast, little evidence that music can actually induce basic emotions.

New Prospects for Research on Music and Emotion

With the methodological problems associated with physiological responses to music and basic emotions in mind, let me now return to Juslin and Västfjäll’s
proposal that lack of progress on understanding the relationship between music and emotion is due to researchers having neglected the psychological mechanisms that underlie listeners’ emotional responses to music. As steps toward a solution, they identified six mechanisms involved in the musical induction of emotions: brain-stem reflexes, evaluative conditioning, emotional contagion, visual imagery, episodic memory, and musical expectancy (Juslin and Västfjäll 2008, 563). With regard to the first two mechanisms, there is firm evidence that music can give rise to things like a startle response (which is a reflex associated with the neural structures localized in the brain stem), and that music can be paired with an affectively valenced unconditioned stimulus such that hearing the music at some later point evokes the same affective state as the unconditioned stimulus, a kind of evaluative conditioning that produces what John Booth Davies called the “Darling, they’re playing our tune” phenomenon (Davies 1978, 69). This is not, however, to say that there is evidence to suggest that music is unique in this respect—to the contrary, such responses can be induced by a wide variety of sound stimuli, and no existing research has shown that music is better or worse at eliciting such responses, or that the responses are in any qualitative way different from those elicited by nonmusical stimuli.

More promising is the psychological mechanism of emotional contagion. As defined by Elaine Hatfield, John Cacioppo, and Richard Rapson, emotional contagion in its most rudimentary form is “the tendency to automatically mimic and synchronize facial expressions, vocalizations, postures, and movements with those of another person and, consequently, to converge emotionally” (1992, 153–54). Although the basis for this mechanism is not well understood, there is some evidence that mirror neurons may play a role: when I see another person express disgust, for instance, neurons associated with how I would experience and express disgust are activated (Rizzolatti and Sinigalia 2008, chap. 7; I provide a brief review of research on mirror neurons in chapter 4). Hatfield and her colleagues point to considerable evidence that emotional contagion is an important factor in the bonding that occurs between mother and child (Hatfield, Cacioppo, and Rapson 1994, 81–88), and show that it is manifested in a wide range of social interactions. For Juslin and Västfjäll, the application to music is relatively immediate: emotional contagion occurs when a listener “perceives the emotional expression of the music, and then ‘mimics’ this expression internally, which by means of either peripheral feedback from muscles, or a more direct activation of the relevant emotional representations in the brain, leads to an induction of the same emotion” (2008, 565). The problem raised by this application is similar to that raised by research on music and basic emotions: attributing emotions to music (which is how I understand Juslin and Västfjäll’s notion of “perceiving the emotional expression of music”) is quite different from observing a conspecific express emotions, not the least because in the case of music the listener must imagine the physical manifestations of emotion that will then be mimicked.

Juslin and Västfjäll note that the mechanism of emotional contagion in music is related to what Jay Dowling and Dane Harwood called the iconic representation of emotion, through which “the ebb and flow of tensions and relaxations in the music mirror the form of emotional tensions and relaxations” (1986, 205).
Dowling and Harwood borrowed their conception of iconicity from C. S. Peirce, but they also linked it to Susanne Langer’s ideas about morphological connections between music and emotions. Although there has of late been renewed interest in the importance of iconic relationships between musical structures and emotional states (Sloboda and Juslin 2001; Scherer and Zentner 2001), it is heir to the same problem as emotional contagion—that of explaining the relationship between musical structures and the physiological and psychological processes associated with emotions. In the next section I would like to offer a solution to this problem based on Barsalou’s theory of perceptual symbol systems, a solution that may also be applied to the fourth of the psychological mechanisms identified by Juslin and Västfjäll, musical imagery, in which music is correlated with a visual image that has a strong emotional valence.

As for the two remaining mechanisms identified by Juslin and Västfjäll in their 2008 article—that is, episodic memory and musical expectancy—the former is, to my mind, closely linked with evaluative conditioning (in that research has yet to show that the memories associated with musical experience are substantively different from those associated with other experiences), and the latter seems able to account for only listeners’ initial affectual response to music rather than the complete range of emotions reported by listeners (Huron 2006, chap. 2).

In subsequent work Juslin and his colleagues added a seventh component to the model set out in Juslin and Västfjäll 2008: rhythmic entrainment (Juslin et al. 2010, 621). They made a persuasive case that physiological entrainment could quite likely contribute to the induction of emotional states. Although I shall argue in the next chapter that the establishment of temporal frameworks that support entrainment is an important resource for musical grammar, I should also want to note that humans can entrain to a wide range of phenomena. Thus, even if rhythmic entrainment proves to be a strong mechanism for the induction of emotions, it would need to be shown that music is a more effective way to prompt physiological entrainment than are other phenomena.

In drawing attention to the psychological mechanisms that underlie listeners’ emotional responses to music, Juslin and his colleagues have set out an approach through which research on music and emotion can move beyond the fact that music can arouse or modulate emotions—and, despite the lack of agreement about just how music relates to emotional life, this is a fact that has never been seriously questioned—to a focus on the means through which this is accomplished. In order to better appreciate the function of music within human cultures, however, I believe it is important to explore how music uniquely exploits such mechanisms, and with this in mind I would like to explore ways sonic analogs for dynamic processes correlate with affectual responses.

**Music, Emotions, and Sonic Analogs**

As my review of research on emotion, and on relationships between music and emotion, points out, we have much to learn about human emotions and about
the way they may be aroused or modulated by music. It would seem, however, that if music does have the capacity to arouse or modulate emotions this capacity depends in no small measure on the way musical materials are organized. In conformance with the approach to grammar I offered in chapter 1, we might generalize this perspective to say that one of the functions of music in human culture is to shape the emotions of conspecifics and that this function is realized through the forms that musical expression takes. Chapter 2, in turn, provided a theoretical framework for exploring how sequences of musical sounds might correlate with human emotions through analogical processes.

Although humans’ affectual responses to changes in the environment are various, they can be distinguished by both the physiological and psychological components of the response and the way these components develop over time. Were we to apply the perspective provided by Barsalou’s theory of perceptual symbol systems to the way affectual responses are understood (at a cognitive level), we would expect that neuronal activations associated with the various components of an emotional response would be brought together through the distributed neural structure of a convergence zone; repeated instances of a given affectual response would establish an integrated, multimodal sensorimotor representation of that response. The pattern of neuronal activation associated with this representation could then be reactivated even in the absence of the environmental circumstances with which the affectual response was associated. Each such reactivation would create a simulation of the affectual response captured through this distributed neural system. As with the simulation of concepts, the simulation of affectual responses would be partial and would involve only a portion of the sensorimotor states associated with a given response. Ekman’s research on subjects who posed a facial expression and subsequently experienced the emotion associated with the expression points to this sort of simulation, as does the art of the convincing actor who summons the living impression of an emotion as part of a role rather than as a response to a given situation. (I should note that the philosopher Charles Nussbaum has also deployed the notion of simulation to explain emotional responses to music; see Nussbaum 2007, 246–56. His use of the notion, however, does not draw on Barsalou’s theory and adopts a view of musical knowledge rather different from my own. More recently, the philosopher Jenefer Robinson and the music theorist Robert Hatten have made use of aspects of Nussbaum’s approach along with other work in philosophy and in musical topic theory to offer an account of the relationship between music and emotions that bears some similarities to what I propose here; see Robinson and Hatten 2012.)

As research on emotion makes abundantly clear, an affectual response is a coherent sequence of physiological and psychological phenomena that is distributed over time and typified by parametric modulation or change—that is, it is a dynamic process. As I proposed in chapter 2, an analog for a dynamic process must minimally provide a correlate for the parametric modulation or change characteristic of the process and represent its sequence of phenomena within a temporal frame that is proportional to that of the original. As a way of taking a closer look at how the sonic analogs of music could correlate with the dynamic processes that constitute an affectual response, let us return to the fourth movement recitative
from Bach’s cantata “Ich habe genug” and the composer’s setting of Simeon’s “Mein Gott! wenn kömmt das schöne: Nun! (My God! When will the lovely ‘now!’ come),” the opening of which was discussed in chapter 1 (the complete score is given in example 3.3). As I noted there, Bach finds resources through which to summon the aged man’s emotions of despair and agitation. In general, despair—as a response to a hopeless situation—draws together the resources of the organism so that they can be conserved; one of its most apparent physiological correlates is a sense of energy ebbing away. Agitation, by contrast, is a response that sends the resources of the organism out in a flurried burst, equal parts distraction (directed

Example 3.3  The fourth movement recitative from Johann Sebastian Bach’s cantata “Ich habe genug” (BWV 82); numbers indicate text phrases. Text translation: “My God! When will the lovely ‘now!’ come: / When I will journey into peace / and into the cool soil of earth, / And there, near You, rest in Your lap? / My farewells are made, / World, good night!” (Bach 2003)
at the agent that prompted the response) and action (to move the organism out of harm's way). To create an analog for a momentary wave of despair, Bach stretches the two syllables of “Mein Gott!” over three eighth notes, the third of which falls a diminished fifth to D4; the drawn-out delivery of “Gott” that results provides an apt correlate for energy typical of despair. To create an analog for a sudden eruption of agitation, Bach packs the six syllables of “wenn kömmt das schöne: Nun!” into about the same amount of time he set aside for “Mein Gott;” but with the singer’s pitch now rising rapidly and stretching beyond the C5 implied by the ascending A♮4 and B♮4 to reach E♭5; the expanding energy summoned by this gesture provides an apt correlate for the burst of action typical of agitation.

Three points deserve comment. First, despite their aptness, Bach’s sonic analogs are not direct expressions of emotions but musical means to prompt the simulation of emotion processes: we need not assume that the singer is in the throes of despair as she renders “Mein Gott!” any more than we need assume that Bach was despondent as he imagined the setting of the text. Second, to the extent that emotional responses make important contributions to both semantic and episodic memory (LeDoux 1996, chap. 7) these responses will enhance the salience—and thus the syntactic relevance—of the sonic analogs through which they are activated. Third, while Bach’s analogs share certain of the sonic materials through which humans express their thoughts and (intentionally or not) reveal their emotions—we could, for instance, imagine a speaker’s voice falling in pitch as he declaims “Mein Gott!” or rising in pitch and accelerating with “wenn kömmt das schöne: Nun!”—the resources of music extend far beyond the relatively straightforward specification of pitch and of rhythmic relationships that is evident in Bach’s setting of the text. These extended resources include the syntactic processes through which composers organize musical materials to create analogs for sustained and complex affectual states, and it is to a preliminary exploration of such processes that we now turn.

**Syntactic Processes in the Fourth Movement Recitative from Bach’s “Ich habe genug”**

As noted in chapter 1 (and in accordance with the construction grammar approach I adopt) the syntactic processes through which musical materials are organized are themselves sonic analogs for dynamic processes. In the case of Bach’s recitative, these syntactic processes play out over the course of successive textual phrases (which are numbered in example 3.3) and, as a whole, create an analog for the restless, questing journey alluded to in the text.

Phrase 1 (which includes Simeon’s remarkable exclamations) sets out the basic tonal framework for the recitative through both the melody (which includes the change from A♭4 to A♮4 that points the way toward C5) and the accompaniment (organized around the leading tone B♭2—harmonized with a fully diminished seventh chord—proceeding to C3). The process here is one of initiation and is common to a range of temporally ordered communicative media. For instance, the linguist Gilles Fauconnier has proposed that linguistic expressions establish mental spaces
Music and Emotion

(introduced here at the end of chapter 2) through what he called space builders, which include a range of dedicated grammatical constructions and which often take conventional forms (such as “Once upon a time . . .”) (Fauconnier 1994, 17). Speakers use space builders to induce their listener to set up a new mental space, which is filled out by their listener’s contextual knowledge as well as by elements and relations provided in the ensuing discourse (Sweetser and Fauconnier 1996, 10); in a similar fashion, the musical process of initiation sets in motion a sequence of events that typically supports, or even requires, continuation.

Accordingly, the setting of phrase 2 creates a sonic analog for the process of continuation. That process here includes an element of diversion through the introduction of D♭ (first in the harmonization of G2 in the accompaniment, then with the setting of the second syllable of “fahren” in the melody). The conclusion of the phrase—with a dissonant leap in the melody from D5 to E♭4 over another fully diminished seventh chord in the accompaniment—further emphasizes the sense that phrase 2 has moved away from the terrain established in phrase 1. The setting of phrase 3 confirms the direction taken by phrase 2 through the transfer of E♮ to the bass of the accompaniment (where it is again harmonized by a fully diminished seventh chord), followed by an arrival on an F minor chord at the start of measure 4. Two aspects of the singer’s melody ensure that we hear this arrival to be a temporary one: first, the singer concludes on A♭3 rather than on the F4 that would answer the E♮4 at the end of phrase 2; second, this A♭3 takes the singer into the lower reaches of her register (a perfect twelfth below the E♭5 of measure 2). The musical materials of phrases 2 and 3 thus build upon the framework provided by the setting of phrase 1 but also reshape it, enough so that by the end of phrase 3 it has been called into question.

Phrase 4 then enacts a process of return to the tonal framework of phrase 1 through the B♮4 on the first syllable of “Schoße,” harmonized with a fully diminished seventh chord that gives way to a G major chord on the downbeat of measure 5. Both of these harmonies include D♭s that replace the D♭s of measures 2 and 3, and it is on D♭5 that the voice momentarily comes to rest as the question begun with “wenn kommt das schöne: Nun!” is completed with the rhyming “ruhn.” Phrases 5 and 6 then bring the analogical journey to a conclusion, first through a reprise of C minor on the downbeat of measure 6 and second through the cadential formula that guides the voice to C4 on the downbeat of measure 7. The finality of this arrival is rendered momentarily contingent by the harmonization of C4 with an A♭ major chord and then assured through the harmonic embellishment of G2 prior to the C minor chord that brings the recitative to a close.

The syntactic processes employed by Bach—which set out a tonal framework, depart from it to come to a momentary rest in another tonal area and then gradually work back toward the original tonal framework—offer a dynamic representation of a wandering, restless journey not completely consonant with the text of the recitative (with its images of peacefulness and rest, and the almost maternal embrace suggested by “im Schoße ruhn” [“rest in Your lap”]). This lack of consonance between music and text suggests the depth of Simeon’s anguish and the extent of his impatience at being separated from the rewards of his faith. The compositional strategies employed by Bach in the course of the recitative thus build
upon the simulation of emotions prompted by his setting of Simeon’s impassioned outburst to create a highly ramified sonic image of Simeon’s internal conflict, one that has the potential to connect with his listeners’ own experiences of similar conflicts. Even in this brief, relatively straightforward recitative, the way in which sequences of musical sound summon dynamic processes related to the emotions is anything but simple or direct.

It bears mention that the overall syntactic pattern of the recitative—which involves establishing focal material, departing from that material, and then returning to it—is one that is common not only to many forms of musical expression but can also be found in a range of narrative forms. What is important here is not so much this general pattern as the specific means through which it is realized: thanks to the way Bach organizes his materials, the return to C minor accomplished in phrases 5 and 6 of the recitative sounds as one that is fully earned, which is part of the reason why the appearance of the A♭ major chord at the beginning of measure 7 is so surprising.

The resources Bach uses in his setting are primarily those associated with fixed-pitch relationships (which serve to anchor the analogical representation of departure and return). By contrast, the rhythmic resources exploited in the movement, which reflect the conventions of recitative (including a syllabic setting of the text and a close relationship between textual and musical phrases), are for the most part similar to those of prosody. (As noted in chapter 2, there is evidence that prosody can be used to create analogs for dynamic processes in the speech stream; cf. Shintel, Nusbaum, and Okrent 2006; Shintel and Nusbaum 2007; Hedger, Nusbaum, and Hoeckner 2013.) This approach to the temporal organization of musical materials changes in the closing moments of the recitative when the continuo part becomes a more active partner with the voice. The partnership between voice and continuo is achieved in part through the appearance of a moving line in the bass (which is marked “arioso”) and in part through the rhythmic framework initiated by the divisions of the beat described by this line, which contribute to the sense that the first beat of measure 7 is a point of arrival. On the one hand, the temporal expectations set up by this brief passage are of limited extent; on the other hand, they do much to enhance the sense of surprise generated by the A♭ major chord of measure 7. Bach’s change of compositional strategy illustrates the interaction of multiple syntactic layers, which, as noted in chapter 1, is typical of musical grammar. For much of the recitative the only active layer is one that uses relationships among pitches to create sonic analogs for dynamic processes. Beginning in measure 6, however, a new syntactic layer is introduced, one that makes use of patterns of expectation generated by the onset of events and the relative accentuation of these events to create a sonic analog for a process of arrival.

The notion of multiple interacting syntactic layers introduces two significant complications into the conception of syntactic processes. First, the resources exploited by a given syntactic layer may more readily afford the creation of some sonic analogs rather than others. For instance, it is relatively straightforward to create a sonic analog for a dynamic process of departure and return through pitch relationships but somewhat more difficult to summon the dynamic process of deceleration by this means alone. Second, the compositional strategies through
which general syntactic processes are realized within different syntactic layers may themselves be quite different from one another. Creating a sonic analog for the dynamic process of digressive continuation within a syntactic layer focused on pitch relationships, for example, may involve little more than the gradual chromatic modification of previously introduced pitch classes, whereas within a syntactic layer focused on temporal expectation this might be achieved through the introduction of additional accentual patterns (such as those associated with syncopation). Given these complications, I shall set aside the specification of syntactic processes for the moment, a deferral that will allow us to focus on the way musical resources within individual syntactic layers are used to create sonic analogs, as well as how the sonic analogs proper to various syntactic layers relate to one another. With this in mind, I would like to return to the first section of Bach’s “Schlummert ein,” the example I explored in my introduction to this chapter. In crafting this aria, Bach coordinates syntactic layers that involve interwoven rhythmic strands, counterpoint between melody and bass, and the shaping influence of harmony to create a rich analogical representation of a sequence of emotions of profound relevance for the attitude he would have his congregation adopt.

Syntactic Processes and Syntactic Layers in Bach’s “Schlummert ein”

All of the predominant syntactic layers through which Bach shapes the musical discourse of the aria are evident in the first section (which was given in example 3.1). In what follows I shall focus on three of these: the first exploits music’s rhythmic resources; the second, music’s contrapuntal resources; and the third, music’s harmonic resources. Each of these syntactic layers can be conceived of as at least partially independent from the others in that each layer is organized through syntactic processes that create distinctive sonic analogs for dynamic processes. In the complete aria these syntactic layers are, of course, coordinated with one another, but it is through making use of the expressive resources of each layer and carefully crafting their interrelationship that Bach is able to create the unique emotional environment summoned by the aria.

Rhythmic Processes. As I noted in my initial comments on the aria, the steady pulsing bass in the accompaniment provides a convincing analog for a physical activity that is deliberate and measured (such as that required to rock a baby to sleep). The steady eighth notes in the bass are, however, further organized through recurring patterns created by Bach’s handling of pitch and register. Examples include the rhythmic figures created by three or more repeated pitches (in measure 2, the second half of measure 3, the first half of measure 5, and elsewhere) and that begin with a neighbor note (as in the first half of measure 3 or the first half of measure 4). These patterns serve, in the first instance, to group the eighth notes by fours and to thus effectively divide each measure in half; in the second instance, through their recurrence they provide a means to mark out larger patterns within the aria (such as the modified reprise of measures 1–4 in measures 7–10).
In many places the melody of the aria pushes and pulls against the regular framework set out by the bass through agogic accents, such as those introduced on the second eighth note of measure 1 and the second and sixth eighth notes of measure 2. The asynchrony of accents between melody and bass that results (which is one species of syncopation) is then typically resolved in short order, as occurs in measures 3 and 4 when the melody fully synchronizes with the bass for the completion of the first two lines of the text. A slightly different use of agogic accents occurs with the repetition of “schlummert ein” when the D4 that enters on the last quarter note of measure 4 is held over into measure 5, creating a half-note duration that is asynchronous with the patterns of four eighth notes in the bass. Although a measure of synchrony is recovered with the arrival on “ein” on the third quarter note of measure 5, the asynchronous pattern is then immediately reprised, this time beginning on G4. The rhythmic profile of the melody is further distinguished from that of the bass through two instances in which pitches are sustained for what seem to be exceptionally long durations. The first of these, at the beginning of measure 11, revisits the setting of “schlummert ein” from measures 4 and 5; the second, beginning in the second half of measure 15, sets up the evaded cadence of measure 16.

On the one hand, these syncopations and sustained pitches complicate the rhythmic patterns set out by the bass; on the other hand, their coordination with the bass and the regularity with which they occur provide an analog for a dynamic process that is initially complex (because of the diverse relationships among its components) but that then becomes relatively simple (as a result of the relationships becoming more unified). This sonic analog would be fleshed out a bit further by rhythmic activity in the accompaniment provided by the continuo player as he realized Bach’s figures (one solution for which is given in example 3.1). In most cases the accompaniment would conform with the broad patterns set out by the bass and would align with the singer’s melody. Within the performance practice of the eighteenth century, however, the assumption was that the continuo player’s accompaniment would be to some degree independent of the patterns set out by the bass and melody, something underscored by the syncopations suggested by Bach’s figured bass symbols on the second of the eighth notes of measures 5 and 6 (and elsewhere). In consequence, the rhythmic patterns created through the realization of the accompaniment might, in some cases, be out of alignment with those of the bass or the melody, and thus further enrich the means through which Bach’s music could provide an analogy for the transformation from a complex set of relationships among the components of a dynamic process to one that is more unified.

A departure from this overall pattern comes in measure 13, which begins a sequence of two and a half measures in which melody and bass are almost perfectly synchronized (save for the C5 held onto the downbeat of measure 14). This analog for an orderly progression of events is then interrupted by the held G4 of measures 15 and 16, and the fermata and rest at the end of measure 16. Measures 17 and 18 then return to the earlier pattern, with a series of syncopations (beginning on the second beat of measure 17) giving way to the more
or less unified relationships with which measure 18 concludes. The context for the pattern here is, however, somewhat changed: not only does it play out over only two measures, but in its rhythmic distinctiveness it is also very typical of the rhythmic patterns associated with cadence (understood as an analog for a process of arrival).

In sum, then, Bach employs two basic strategies in the sonic analogs set out by the rhythmic layer of syntax in this aria. The first offers analogs for the transformation of a complicated dynamic process into one that is more unified; the second offers an analog for a relatively orderly process that is broken off only to be resumed and brought to completion. These strategies could be seen to correlate with a process through which an infant is lulled to sleep, but they would fit equally with a process of quiet reflection through which the attention is repeatedly brought into focus and, with focus achieved, brings about (with some effort) a resolution of the issues that prompted the reflection.

Contrapuntal Processes. As suggested by my discussion of the melody of Simeon’s recitative (which follows “Schlummert ein” in Bach’s cantata), much of the effect of tonality—and in particular the dynamic processes of departure and return—can be accomplished with a single, sparsely accompanied voice. That said, Bach’s setting of the aria demonstrates how the coordination of a melody with a relatively independent bass voice (a coordination conventionally called counterpoint, and here understood to include the principal signposts of harmonic organization) makes possible the creation of a far richer set of analogical structures. For example, the introduction of F♯5 at the beginning of measure 2 of the aria momentarily pulls the music away from G major and toward C major, a departure that would be reinforced in performance by further F♯s supplied by the continuo player. The F♯5 of measure 3 and additional F♯s provided by the continuo then bring about a return to G major. As illustrated by the analytical sketch of example 3.4, however, the counterpoint between the outer voices holds off any sense that this return is complete: the dominant harmony of measure 3 is prolonged through measure 4 and beyond, for the tonic harmony provided in the interjection of measures 5–6 does little to complete the tonal process initiated in the first four measures. Indeed, it is only after a more marked departure from G major in the first half of measure 13 that Bach offers his listener a convincing return to G major, which is immediately undercut by a reprise of F♯, first in the accompaniment of measure 15 and then as the singer’s sustained G4 falls to F♯4 in measure 16. It is only subsequent to this last deflection that Bach supplies a cadence that brings this section of music (and ultimately the entire movement) to a conclusion.

In sum, then, the coordination of melody and bass in this section of Bach’s aria creates a sonic analog for a gentle and yet focused—perhaps even intense—feeling of anticipation, a consequence of the delay of an assured tonic harmony through interjections (the repetitions of the softly suasive “schlummert ein”) and the brief deflection toward C major that occurs in measure 16. To some extent, the entire arc of the dynamic process of anticipation is completed with the cadential material of measures 17–18, and yet for a moment or two the poignancy and anxiety of the delay remain to color the final arrival on tonic.
Harmonic Processes. Within the tradition of compositional strategies of which Bach was a part, the relationship between contrapuntal syntax and harmonic syntax was a close one, so much so that musicians could hear the outer-voice counterpoint for a passage and have a very good idea how it would be filled in. That said, the process of filling out the harmonies for a passage was far from automatic (see, for example, the discussion in Buelow 1986, chap. 7), and in the choices they made musicians could further shape the course of musical discourse. In Bach's aria this is evident in three tendencies. The first tendency is for a root-position dominant seventh harmony to be followed by a root-position minor chord built on the bass note a whole step above that of the dominant chord, a progression that, when it occurs at the close of a musical section, is often called an interrupted or deceptive cadence. Examples can be found on beats 2 and 3 of measure 1, from beat 4 of measure 3 to the downbeat of measure 4, and (without an explicit harmonic seventh) from beat 4 of measure 2 to the downbeat of measure 3. The second tendency is for harmonic dissonances to resolve quite soon after they are sounded. Examples include the chordal seventh on the fourth eighth note of measure 4, which resolves immediately after on beat 3 of the measure, and the chordal seventh introduced in the first halves of measures 5 and 7, which each resolve with the change of harmony in the second half of the measure. The third tendency is the avoidance of any striking dissonances throughout this portion of the aria—indeed, although fully diminished seventh chords are prominent in other sections of the aria (and are particularly noticeable in Bach’s setting of “Hier muß ich das Elend bauen” [“Here I
must build up misery”]) they are absent here (although there is an allusion to such chords with the melodic leap of a diminished seventh at the beginning of measure 13). Taken as a whole, Bach’s harmonic syntax offers an analog for a gently flowing succession of events that proceeds in an orderly fashion, avoiding abrupt or sudden changes (including any arrivals on tonic that might interrupt the flow with repose).

“Schlummert ein” and the Simulation of Emotions

As I noted in the opening of this chapter, “Schlummert ein” is evocative of a lullaby, a musical genre designed to calm infants—that is, designed to induce in the infant an affectual response that culminates in a relaxed, harmonious state. Although there are significant variations among the lullabies used by different cultures, they typically have a slow tempo, have melodies with a simple contour, and make use of a significant amount of repetition of both textual and musical materials (Unyk et al. 1992; Trehub and Trainor 1998). These design features not only restrict the amount of stimuli presented to the infant but also offer a sonic analog for physical activity that is gentle, regular, and directed toward a single, easily comprehensible goal. With respect to the compositional strategies that would enact such a design, one would expect the syntactic processes through which the sonic analogs of different syntactic layers are organized to be coordinated with one another to yield a musical utterance appropriate to the process of lulling an infant to sleep.

While measures 1–18 of Bach’s aria, with their steady rhythmic pulse and gently flowing harmonic syntax, instantiate many of the basic design features of a lullaby, relationships among the syntactic processes of the aria’s constituent syntactic layers suggest an imagistic landscape more complex than that which would stretch uninterrupted toward sleep. Bach accomplishes this through the careful shaping of the elements of and relations between syntactic layers: in some cases, syntactic processes are aligned with one another, in other cases not. An instance of the latter occurs in the elaborate dance enacted between the rhythmic and contrapuntal layers. As noted earlier, one of the prominent strategies Bach uses in the rhythmic layer creates an analog for the transformation of a complex set of relationships into one that is more unified; in measures 4 and 10, however, the coalescence of rhythmic materials coordinates with an arrival on a contrapuntal goal that demands a consequent (a demand brought to the foreground by the fermata and rest with which measure 10 concludes). Syntactic layers can also, of course, be aligned with one another: beginning in measure 13 the rhythmic, contrapuntal, and harmonic layers all contribute to a sense of forward direction that culminates in the arrival on tonic on the third beat of measure 15; the tightly knit structure then unravels (the sustained note of the melody falling out of rhythmic coordination with the bass, and the harmony introducing Fs that call into question the contrapuntal arrival on tonic), momentarily pausing in silence. Through means such as these, Bach creates a musical utterance that has the potential to correlate with a developing affectual state in which the feeling of being soothed is mixed with a feeling of anticipation, and the balance between the two eventually resolved through a concerted effort.
Perhaps most important for the perspective I wish to develop here, to the extent that measures 1–18 of Bach’s aria give rise to a simulation of emotions the process is anything but straightforward. Not only must the sonic analogs created within various syntactic layers be taken into account, but the relationships among these layers (and thus the overall sonic analog created by Bach) may pull the listener’s attention in a number of different directions. Encounters with this utterance by different listeners, or with the same listener on different occasions, may in consequence give rise to subtly different interpretations of the emotional import—or, more broadly, the meaning—of the utterance.

**Sonic Analogs and Affectual Responses**

As indicated by my analysis of Bach’s setting of Simeon’s remarkable exclamations in the fourth movement recitative from “Ich habe genug,” the sonic analogs through which a musician prompts the simulation of an emotional state may be relatively direct and make use of resources also used for nonmusical evocations of emotions. I would like to suggest, however, that it is more often the case that musicians draw on the resources of musical syntax to create artful representations of emotion processes. In some cases the syntactic resources through which these representations are accomplished are of limited extent—for instance, the musical materials of the recitative that begins with Simeon’s outbursts are for the most part organized around pitch relationships through which his wandering journey can be summoned. In other cases the syntactic resources are rather more involved, as in the three distinct and carefully coordinated layers of musical syntax that can be traced in the opening section of Bach’s “Schlummert ein.”

More generally, I would propose that the emotions induced by Bach’s “Schlummert ein,” by the recitative that follows it, or by any other musical work rely in part on correlations between the dynamic processes proper to affectual responses and the sonic analogs of music. By this means, sonic analogs have the potential to induce a simulation of an affectual response—that is, a partial activation of the multimodal sensorimotor representation of that response. Although such simulations are thoroughly genuine, they are also somewhat different in function from more typical affectual responses. As Klaus Scherer has noted, an organism’s responses to emotional stimuli are, for the most part, adaptive: they help the organism prepare its reaction to the stimuli, and as such they are proactive. Responses to musical stimuli, by contrast, are rather diffusely reactive and not as highly synchronized across different organismic components as those of more utilitarian emotional responses (Scherer 2004, 244).

**Conclusion**

**Music, Emotion, and Cultural Practice**

As should be evident from the analyses I have offered here, although musical structures can reflect the morphology of emotions they can also represent
the morphology of various other kinds of dynamic processes. So why is it that
the trope of music as an emotional art has been so powerful and so enduring?
Two answers come to mind, the first of which I suggested in my introductory
comments: the insubstantial character of sonic analogs makes the connection
between music and emotions easy to draw, as both seem to have their proper
home within our internal experience rather than in the world without. The
second reflects the resources that sonic analogs offer for the manipulation of
emotions. Before the music of Bach’s “Schlummert ein” breaks in to a listener’s
consciousness she could have anything on her mind and be in any emotional
state. Once she begins to attend to the music—and, more specifically, to cor-
relate musical materials with affectual dispositions—she will move at least part
of the way toward the restful but attentive state that seems to be the musical and
liturgical goal of Bach’s aria. The notion that music can manipulate emotions
takes us back to the legend of the Tauromenian youth and to the theories of the
Florentine Camerata, but this particular example also highlights two things left
out of these accounts that are crucial to the broader function of music within
human culture: the social context for musical emotions, and the relationship
between musical emotion and ritual.

It should be kept in mind that Bach’s “Schlummert ein” is taken from a cantata
that was itself but one portion of a worship service that provided the occasion for
a thickly textured set of social interactions. One of the significant moments within
these interactions was the sermon: although early eighteenth-century Lutherans
might view other portions of their four-hour long worship service to be optional,
close attention to the religious guidance provided by the sermon was regarded as
obligatory. In the typical order of service, the cantata was just prior to the sermon
and was thus part of a process of bringing together and settling the congrega-
tion in preparation for the sermon (Leaver 1997; Kevorkian 2002; Gardiner 2013,
chap. 9). Bach designed “Schlummert ein” to be a key part of this process: the aria,
which occupied the central point within the five movements of the cantata and
was set off on either side by recitatives, invited reflection on the part of the listener,
an activity directed and supported by the regular repetitions of the opening music.
(In addition to the three instances of the opening music that occur in the version
from Anna Magdalena Bach’s notebook, within the cantata the music of measures
1–18 of “Schlummert ein” was used as both an instrumental introduction to and
postlude for the aria.) Although this reflection would be, of necessity, individ-
ual, the overall activity of listening to the cantata and being invited by its musical
materials to reflect was one framed by a social process of worship. The emotional
changes wrought by the aria were thus conceived relative to and initially realized
within a communal, rather than individual, context.

The congregational gathering of Lutherans in the early eighteenth century was,
of course, part of a ritual of worship that was conceived of as having real efficacy
for the immortal souls of believers. Within this ritual, the activation of affectual
responses by means of a musical work such as “Schlummert ein” had two basic
purposes: first, to focus listeners’ attention in preparation for the teachings of the
sermon; and, second, to support the religious message of the text. As I suggested
in my analysis of the aria, this message concerned not only the promise of resting
in the Savior’s embrace but also the need to actively turn away from the superficial rewards and very real miseries of life on earth. The emotions experienced while listening to “Schlummert ein” were not ancillary to this message but instead ensured its absorption and thus its ritual efficacy.

The role of music in social and ritual interactions highlighted by Bach’s “Schlummert ein” points to one explanation for why it was important for human cultures to develop tools such as music for manipulating emotions. The basic affectual responses to the aria that I have noted—the establishment of a relaxed harmonious state, and a heightening of attention—are not in themselves extraordinary, and they could be induced by a wide range of phenomena. For instance, before sitting down with a book many individuals will seek to modify their environment—by settling in to their favorite chair, adjusting the lighting to have a clear view of the page—to create a similar affectual state. The responses to “Schlummert ein” are different in that they are meant to be extraordinary: they are prompted by the actions of privileged members of the worship community (extending from the composer through to the performing musicians) and have as their context a ritual intended to prepare members of the congregation for everlasting life. The aria, the cantata of which it is a part, and indeed the whole of the worship service are designed to create a kind of experience that is anything but ordinary, and to do this in large part through the manipulation of the congregation’s emotions.

This perspective on the role of emotions within human culture conforms, for the most part, with the ethological approach to musical emotion developed by Ellen Dissanayake (2001). As she describes it, “In an ethological view, the biological purpose of emotions is to motivate behavior—to make us respond appropriately to the sorts of occurrences in the environment that could affect us, for good or ill” (2006, 32). Dissanayake argues that ritual provides a framework for the manipulation of emotions for social purposes—that is, motivating specific forms of behavior—and that music provided human cultures with unique resources for accomplishing this goal. Chief among these resources is the ready correlation of the sonic analogs of music with various movement patterns (such as represented in prototypical form by the rocking of a baby in time to the rhythm of a lullaby; cf. Dissanayake 2000) and the various strategies of musical grammar predicated on an active engagement with the progress of successive dynamic processes, of which Meyer’s expectation-realization model is but one example (Dissanayake 2006).

As was evident in my survey of research on the emotions over the past few decades, affectual responses have both psychological and physiological components. This suggests that the process of recalling an affectual response would involve both—from the perspective provided by Barsalou’s theory of perceptual symbol systems, the recollection (or, in his terms, simulation) of an emotion gives rise to a partial neural reenactment of the sensorimotor states associated with that emotion. With respect to emotions, then, the body serves as a way of knowing: on the one hand, the dispositions of the body are a crucial component of what an emotion feels like; on the other hand, the body serves as a resource through which emotions are expressed (Colombetti 2014). To a certain extent,
Music and Emotion

the same is true of music: as I shall try to show in the next two chapters, the sonic analogs of music are consistently grounded in and correlated with bodily movement. And so, while both emotions and music may seem insubstantial, they are both thoroughly incorporated—the “internal experience” that seems to provide them their proper home is in fact anchored in and expressed through bodily experience.

As I hope to have made clear, the way music is used to manipulate emotional states—if and when it does—is anything but simple and direct. What music can offer, and frequently does in the service of social cohesion and ritual practice, is a means of simulating affectual responses. Through the generality of analogical reference, such simulations also offer a means for exploring the features of such responses and how they connect with other phenomena: the tonal peregrinations that summon the anguish felt by the soul in its separation from God might equally summon a wandering journey through a trackless waste; the regular rhythms and harmonies that suggest the tranquility of the nursery at evening might equally suggest an aimless afternoon spent floating on a summer’s lake. Indeed, these possibilities point to the broader function of music within human cultures, for music’s sonic analogs provide a way to grasp—or, better, to enact—a wide range of dynamic processes, to embody their most essential features, and to do so in a way that makes them immediate to others.

**Elements of Musical Grammar: Syntactic Prerequisites**

In my discussion of the movements from Bach’s “Ich habe genug,” I employed—in an admittedly informal fashion—two interrelated elements central to musical grammar: syntactic processes, and syntactic layers. As previously noted, the syntactic processes used to organize the basic elements of musical grammar are themselves analogs for dynamic processes. The possibility of multiple syntactic layers, each organized by syntactic processes appropriate for that layer, is one of the distinctive features of musical grammar. Although syntactic layers are typically coordinated in a musical utterance, as we saw in my analysis of Bach’s “Schlummert ein,” this coordination may be relatively loose depending on the function that utterance is to realize. In what follows I would like to make an initial move toward formalizing my account of syntactic processes and syntactic layers by considering the cognitive constraints on each.

**Syntactic Processes.** As a species of construction grammar, each of the grammatical elements in musical grammar, including syntactic processes, are combinations of form and function. A ready example is provided by the syntactic process of repetition (which might vary from the exact to the approximate). Although repetition is a common and characteristic feature of musical utterances (Margulis 2014), its capacity for building meaning can be exploited through language. In rhetoric, for instance, repetition is exemplified by the figure known as anaphora. Pseudo-Longinus, in his first-century account of rhetorical practice, demonstrated anaphora in an excerpt from a speech by Demosthenes: “By his manner, his looks, his voice, when he strikes you with insult, when he strikes you like an enemy,
when he strikes you with his knuckles, when he strikes you like a slave” (pseudo-
Longinus 1995, 237). The steady repetitions of “strike” provide a sonic analog for
the action of striking, which, as Mark Turner observed, automatically involved
the members of Demosthenes’s audience in the basic structure of the meaning
he aimed to impart with his words (Turner 1998, 50–51). Indeed, these repeti-
tions draw attention to the simulation of motion they prompt, guiding listeners
to reflect on the motivations for the repeated metaphorical blows. Within the
musical domain, when the syntactic process of repetition is applied to sequences
of sounds without regard to their pitch content—repeated tapping on a tabletop
being one example—it provides the raw material for musical meter; implement-
ated without variation, it provides a sonic analog for the dynamic process of active
stasis. When the syntactic process of repetition is applied to pitches—the open-
ing motto of Beethoven’s *Fifth Symphony* is a reliable example—the result can be
something closer to anaphora: our attention is drawn to the basic configuration
that is repeated; this momentarily brings that configuration to the foreground,
prompting reflection on the motivation for the configuration of pitches and for
making it prominent through repetition.

In her exploration of musical meaning (which adopted a cognitive perspective
similar to that which informs my approach), Candace Brower identified a host of
dynamic processes (which she framed in terms of image schemas) that were rel-
relevant to the organization of musical materials that contributed to the construction
of meaning. These included processes related to pursuing a path toward some goal,
processes involved with achieving a sense of balance or stability, processes focused
on departure and return, and processes concerned with movement through the
topography of pitch space (Brower 2000, 333–52). One of the things that distin-
guishes all of these processes is a basic appeal to causality: paths typically lead to
a goal; instability prompts a search for ways to achieve stability; departures often
(although not necessarily) motivate return; differentiated topographies suggest
possibilities for motion. Although causality of this sort might seem relatively sim-
plistic, as noted in chapter 2 (in connection with a discussion of Jean Mandler’s
work) there is evidence that attention to the causal entailments of such processes
is key to infants’ ability to form concepts. The syntactic processes exploited by
music could thus be seen to be a deeply ingrained aspect of the knowledge through
which we guide our understanding of the world.

On the one hand, the notion of a syntactic process I have proposed is intention-
ally quite broad and limited only by the kind of coherent sequences of phenomena,
distributed over time and typified by parametric modulation, that could be repre-
sented analogically. On the other hand, as realized in musical grammar, syntactic
processes are necessarily applied to sequences of sound events. As I noted above,
a given sonic analog may be more easily realized with certain configurations of
musical sound rather than others; as a correlate, the same syntactic process—for
instance, repetition—may suggest one sonic analog when applied to sequences of
unpitched sounds and another when applied to a sequence of pitches.

As a provisional summary, then, the syntactic processes of musical grammar
will tend to be constrained by a basic logic of causality and by the limitations of the
sonic resources to which they are applied.
Syntactic Layers. One of the significant challenges presented by the multiple syntactic layers of musical grammar is keeping track of the processes in each layer. In most cases, the coordination of these layers is such that it is indeed not necessary to keep track of the processes unfolding in each layer. For instance, a standard tonal cadence, which involves processes that exploit rhythmic, melodic, contrapuntal, and harmonic resources, constitutes a single holistic structure. That said, should circumstances demand we can discriminate among cadences based on the specific character of individual processes, allowing us to distinguish between those that are equivocal and those that are emphatic. (For a description of some of these processes in music from the Classical repertoire see Sears 2015.) The possibility of teasing out the processes specific to individual syntactic layers is something that may be demanded of individuals who are involved with constructing musical utterances. This was at least the perspective of Johannes Brahms, who, in his lessons with Gustav Jenner, emphasized that cadences in songs had to conform to the structure of the text that was being set.

Discriminating between different syntactic layers is also common among performers faced with a challenging passage: learning the passage is simplified if, for instance, rhythmic patterns are first isolated from pitch materials. And an attention to syntactic layers may also occur when musical utterances are constructed whole or in part through the act of performance, as happens within certain improvised traditions.

The notion that musical utterances involve multiple syntactic layers is perhaps most useful for developing accounts of how utterances that are in some respects quite similar to each other nonetheless produce markedly different effects. For instance, there is good evidence that a significant number of the composers working in Europe during the late eighteenth century drew on a common stock of contrapuntal patterns to create their music (Gjerdingen 2007). We nonetheless find greater reward in our encounters with some of the musical works that resulted from this tradition of shared compositional strategies than we do with others, suggesting that it is the way such contrapuntal patterns are coordinated with syntactic
processes unfolding in other syntactic layers that makes it possible to create truly remarkable musical utterances.

The nature of syntactic layers, and the syntactic processes that are deployed in these layers, will become somewhat more evident in the discussion of temporal and tonal frameworks toward the end of chapter 4.

**Open- and Closed-Class Forms**

As noted in chapter 1, in his discussion of the way grammatical structure contributes to the meaning of an utterance Len Talmy proposed drawing a distinction between open-class and closed-class forms. Open classes have a large number of members to which more are easily added. Closed classes, by contrast, have relatively few members and are more difficult to augment. Open-class forms contribute most of the content of an utterance, whereas the closed-class forms determine most of its structure. With respect to musical utterances, open-class forms can be conceived of as exploiting two features of musical sound. The first concerns the actual sonic materials used to realize a particular sonic analog: a harmonic cadence requires different materials than a melodic cadence, and a rhythmic cadence different materials still. The second feature concerns the way syntactic layers are coordinated: a strong harmonic and melodic cadence, coupled with a weak rhythmic cadence, will not have the same effect as if all of the implementations of the sonic analog for closure are equally strong. Thus while the array of closed-class forms may be relatively limited—the catalog of processes assembled by Brower in her discussion of musical meaning might be a good place to start—the way these processes would be realized through musical sound is practically unlimited.
To think of Fred Astaire is to think of the lithe, effortless dancer who gave presence to music through his movement. Yet Astaire was also an accomplished singer, one known not so much for vocal facility or beauty of tone as for his ability to deliver a song with commitment. Nowhere is this clearer than in a scene from the 1936 movie musical *Swing Time*, which introduced Jerome Kern and Dorothy Field’s “The Way You Look Tonight.” In the film, Astaire plays a dancing gambler (or gambling dancer) who has recently come to New York to earn some money. He encounters Ginger Rogers, who plays a dance instructor, and he soon has designs on her, both as a dance partner and as a possible romantic interest.

In the scene in which “The Way You Look Tonight” is introduced, Astaire is trying to win Rogers over—he has managed to get in to her hotel suite, and, even though she has repeatedly spurned him, he makes one more effort to win her over by singing Kern and Field’s song, accompanying himself at the piano (as is shown in example 4.1). Rogers, for her part, has locked herself in her room and started shampooing her hair, an activity that contributes to a dramatic situation that is both intimate and comedic.

As is typical of films of the period, the scene is shot from a number of camera angles and involves multiple takes. There is, however, one sustained take that uses only one camera angle and that provides an opportunity to study in detail the way Astaire’s movements correlate with his delivery of the first verse of the song. Before turning to those movements, which, on the view I want to develop here, are an integral part of his rendering of the song, let me briefly summarize a few of the salient features of the first verse, the score for which is given in example 4.2. (In conformance with the conventions of popular song, I have notated the melody on a treble staff; Astaire’s light baritone actually sounds one octave lower.) One way to both describe and hear Kern’s melody is in terms of a number of smaller units. There is, for instance, the falling fifth of measures 1–2 (“Some-day”), which is answered by the arch-like rise and fall through a third that occupies measures 3–4 (“when I’m aw-f’ly low”). The latter motion is replicated (a step higher) in measures 5–6 (“when the world is cold”), and again in measures 7–8 (“I will feel a glow just think-ing”). This last arch turns out to be a bit more complicated, for as soon as it has returned to its starting G4 it ascends once more, this time continuing up to the D5 of measure 9 and then falling to the D4 of measure 10 (“of you”). If our focus is on Fields’s words it makes sense to consider measures 7 through 10 as a whole, for the positive feelings of the protagonist are a consequence not simply of “thinking,” but of “thinking of you.” If, however, our focus is on the music our perspective on the passage shifts a bit: the change from the continuous quarter
notes of measures 7–8 to the whole notes of measures 9–10 not only brings about a marked change in the rate at which notes appear, but it also hearkens back to the opening measures; there is further a change from a stepwise succession of pitches to a descent through an octave that can be heard to echo and expand the descent through a fifth of measures 1–2. (The connection between these two moments in

Example 4.1 Still image of Fred Astaire at the piano from *Swing Time*


Andantino
the melody is all the stronger in that they both stand at the beginning of an eight-measure subphrase, and both descents conclude on D4.) The octave descent of measures 9–10 is followed by a figure that ascends and descends (“And the way you look to-night”), much as did the figures of measures 3–4 and 5–6, but with important differences: the arch described by the succession of pitches is not as smooth; the durations of the notes become longer (in successive stages); and the entire figure ends on D4, a whole step below where it started.

Not surprisingly, as Astaire sings he moves—not a lot, for he is anchored by the conceit of accompanying himself at the piano (an illusion all the more charming for the string orchestra that joins him on the soundtrack)—and he does so in a way that seems thoroughly natural, not least because of the correlation of his movements with the structure of Kern’s verse. For my discussion of these I will rely on line drawings created from still images taken from the movie; although these drawings will necessarily leave out some very interesting details, they will allow us to focus on the principal features of Astaire’s movements.

The first of these drawings is given in example 4.3 and shows Astaire’s position at the beginning and end of the gesture that accompanies his singing of “someday.” In the course of this gesture Astaire moves toward the right side of the frame, lifting his shoulders and chin slightly, and then concludes the movement by dipping his head toward the keyboard (as shown in the second image of example 4.3). This set of movements is typical of those he makes in his performance, restricted as he is to those he can make with his head and his torso. The movements he makes with his arms are much less noticeable and, in that the camera angle takes a view of Astaire from over the top of the piano, are for the most part obscured.

As Astaire sings “when I’m awf’ly low” he makes a slight circular motion with his shoulders, moving first to the left of the frame (a displacement not shown in example 4.4) and then back past the center, holding this position for his arrival on the long note that sets “low.” This same basic movement is repeated as he sings “when the world is cold.” Astaire then uses a varied version of this movement for the fourth line of lyrics: as shown in the first image of example 4.5, he begins “I will feel” in the center of the frame; as he sings “a glow just thinking” he moves toward the right and top side of the frame, reaching and holding the furthest
extent of this motion on the note that sustains “of” (shown in the second image of example 4.5). For the completion of the line, Astaire then moves forward and dips his head, such that as he sings the word “you” his chin reaches its lowest point in this sequence of movements. It is worth noting that this is an expansion of the movement he made at the opening of the verse when he sang “someday,” something that can be seen by comparing the images of example 4.3 with the first and last image of example 4.5.

For the last line of lyrics, Astaire makes use of a new sort of movement as well as returning to one he used previously. The beginning of the line (“and the”) finds him dipping his head just a bit further down than it had been when he sang “you” and moving toward the left of the frame (compare the last image of example 4.5 and the first of example 4.6). As he continues with “way you look to-” he moves straight back away from the keyboard, reaching and holding the furthest point of the entire sequence. The gesture and the verse conclude as he moves back toward the keyboard, lowering both head and torso, for “night.”

In summary, then, there are three basic sorts of movements that Astaire makes as he sings the first verse of “The Way You Look Tonight.” The first, which accompanies his singing of “someday” (example 4.3) results in a dip of his head; the second, which is used for “when I’m awf’ly low” and “and the world is cold” (example 4.4) involves a circular movement from the left of the frame to the right; the third (skipping over “I will feel a glow just thinking of you” for the moment),
which is employed for the first part of “and the way you look tonight” (example 4.6), finds him moving straight back away from the keyboard. This last movement concludes with a reprise of the dip of the head used for “someday”—it can thus be analyzed as a combination of two gestures, the second recalled from earlier in the sequence. In a similar fashion, Astaire’s movement as he sings “I will feel a glow just thinking of you” (example 4.5) can be analyzed as a combination of the circular motion used for “when I’m awf’lly low” and “and the world is cold,” and the dip of the head used for “someday” (and, ultimately, for “tonight”). It is important to note that these movements are in close correspondence with basic features of Kern’s melody: Astaire’s dip of the head corresponds with the arrival on D4 (whether approached through a falling interval, as in measures 2 and 10, or as part of a cadential formula, as in measure 13); his circular motion corresponds with the arch-like melody of measures 3 through 8; and his movement away and back toward the keyboard corresponds with the cadential formula initiated in measure 11. Further, combinations of musical materials give rise to combinations of gestures. As I observed earlier, while the words for measures 7 through 10 express what is apparently a single thought (“I will feel a glow just thinking of you”) the music makes recourse to two different kinds of material (the arch-like rise and fall—and further rise—of measures 7–8, and the falling interval of measures 9–10); the gesture that corresponds with this phrase, responding to the music rather than the words, involves two distinct movements (circular movement plus dip of the head). Indeed, the complications of music and gesture here point to an important shift within the scene constructed by the words: with “of you” we are moved from vague musings on the future to a recognition of the agency of another individual—that is, the person addressed by the singer—which, indirect as it is, will make it possible to face uncertain times.

Given the limited repertoire of movements Astaire makes and their correlation with basic features of Kern’s tune an argument can be made that these movements are an integral and almost certainly unscripted part of his rendering of “The Way You Look Tonight.” In this they are similar to the spontaneous gestures that accompany speech that have been studied over the past three decades by Adam Kendon, David McNeill, Susan Goldin-Meadow, and their colleagues (Kendon 1980, 2004; D. McNeill 1992, 2005; Goldin-Meadow 2003a). Such gestures have been shown to make use of recurring patterns of movement that correlate with, but are in some measure independent from, basic units of speech. Most of the evidence suggests
that these gestures are an essential aspect of human thought, and that they fulfill a role somewhat different from that of language. (It should be mentioned that I use “language” as a term more comprehensive than “speech.” Thus American Sign Language is a language that does not involve speech. It also, however, makes use of gestures independent of the physical movements through which signs are produced; with regard to these see Taub 2001, Liddell 2003, Wilcox 2004, and Goldin-Meadow and Brentari forthcoming.)

As noted in chapter 1, the term “gesture” is often used—both formally and informally—to refer to holistic sonic units that are analogous to various physical movements. The idea that there would be an affiliation between music and gesture is hardly a new one—Etienne Bonnot de Condillac’s *langage d’action* (also mentioned in chapter 1) had its basis in gesture and, in Condillac’s originary myth, led directly to music (as well as to language and poetry; for a compact summary, see Zbikowski 2008a, 283–84). In general, there are two basic perspectives on how music and gesture relate to one another. Richard P. Blackmur, in an essay that first appeared in 1942, saw gesture as something that exceeded conventional linguistic meaning and that was presented in its purest form in music (Blackmur 1952, 11), a view that was later taken up by Edward T. Cone (1974, chap. 8); we might call this approach music as gesture. A slightly different perspective regards music and gesture as autonomous modes of expression that are frequently coordinated and interdependent (a view that is, in the main, common to the essays gathered in Godøy and Leman 2010); we might call this approach music and gesture.

In this chapter I would like to explore both of these perspectives, and to do so with the idea that music and gesture, along with language, are ways humans communicate with one another. While the model for human communication has long been based on language, studies of gesture suggest that this model should be broadened to include the spontaneous movements that often accompany speech. In the first section that follows I shall review some of what recent research has taught us about the way humans communicate through gesture, with particular attention to the way gesture provides physical analogs for dynamic processes (analogies that can thus be seen to share structural features with music’s sonic analogs). As it happens, humans are not the only species to make use of physical gestures in the course of communication, and in the second section of this chapter I will explore the infrastructure that sets human communication apart from that of other primates. This exploration reveals important features of human communication, features that are exemplified in different ways by language, gesture, and music. Within these two sections, my observations about music and gesture will for the most part be concerned with how they are similar to one another—a music as gesture approach, if you will. In the third section I shall shift to a music and gesture approach, and I will focus on how these modes of communication are different from one another (other than the simple fact that one uses sound and the other physical movement). The example I shall use to illustrate these differences is drawn from a well-known scene from Charlie Chaplin’s *The Great Dictator*, one in which music directly shapes our understanding of what is communicated through physical movements. This shaping is facilitated in part
through the resources provided by musical grammar, resources that include frameworks for temporal and tonal organization that I discuss in the fourth main section of the chapter.

In my concluding comments I return to the movements Astaire makes in his performance of “The Way You Look Tonight” to consider what they can tell us about musical grammar. For the moment, however, I would like to note two remarkable features of those movements. First, to the extent that Astaire’s movements reflect the structure of Kern’s tune, we can say that these movements are a consequence of musical organization: Astaire’s gestures are a visual instantiation of the sonic analogs for dynamic processes that make up Kern’s tune. Second, the relative independence of Astaire’s movements from language and their close coordination with music attest to the importance of analogical structures to human thought: music and gesture are, in their different ways, means of giving phenomenal presence to dynamic analogical structures. Both of these features point to what a careful consideration of music as gesture, along with music and gesture, can tell us about the nature and structure of musical grammar.

**Gesture and Language**

One could say that the study of relationships between gesture and language that blossomed in the 1970s and 1980s was a result of the ready availability of audiovisual recording technologies. As Kendon has noted, interest in gesture goes back to antiquity, but a crucial factor in making it possible to study gesture was the development of relatively affordable means to capture speakers in the act of gesturing (Kendon 2007, 17). In particular, videotape allowed researchers to develop experimental protocols through which they could compare the ways different subjects gestured as they spoke, and to explore in greater detail the constituent elements of gestures and their relationship to language.

One thing that the technologies of both film and videotape made evident was that humans move in a wide variety of ways when they speak. That said, it appears that there is broad agreement on which of these movements are relevant to discourse: Kendon reports an experiment done in the 1970s in which he showed a group of individuals a film of a New Guinea highlander addressing a large gathering.

The observers, who did not hear the sound track of the film, were asked to describe what movements they saw the man make. All of them recognized that he was speaking to a large gathering, all of them recognized the same segments of movement as being related to his speech and all of them distinguished these quite sharply from other movements that, they were all agreed, had nothing to do with his speech. Thus arm extensions, elaborate movements of the hands in the space in front of the body, were all recognized as belonging to his speech performance. Self-touching, postural movements, movements involved in the manipulation of an axe—these were all regarded as being quite separate. (Kendon 1980, 208)
Researchers interested in gesture have developed a number of ways to categorize the movements that belong to speech performance. Building on Kendon’s work and focusing on the relationship between various kinds of movement and language, McNeill arranged gestures into what he called Kendon’s continuum, a version of which is shown in example 4.7. McNeill observes, “As one moves from gesticulation to sign language, two reciprocal changes take place. First, the degree to which speech is an obligatory accompaniment of gesture decreases. Second, the degree to which gesture shows the properties of a language increases” (2011, 344). McNeill’s preferred term for what Kendon called “gesticulation” was simply “gesture,” and it is on gesture so conceived that most of the research on the use of movement in human communication has been focused. In his work over the past two decades McNeill has distinguished four basic types of gesture: iconic, metaphoric, deictic, and beats (D. McNeill 2005, chap. 2). Iconic gestures are strongly imagistic, whereas metaphoric gestures tend to be somewhat more abstract; deictic gestures involve pointing, at times to objects that are actually present, at other times to “virtual” objects summoned through discourse; beats (which, in appearance, look much like movements that beat musical time) typically mark items of discourse for attention, providing (through physical movement) a material anchor for different elements of the speech stream.

An alternative way to classify gestures proposed by Cornelia Müller focuses on the function of gesture within discourse. Müller’s classification distinguishes between three basic types of gesture: discourse gestures, which structure an utterance (for instance, counting off the points one is making on one’s fingers); performative gestures, which enact speech acts (for instance, using a sweeping motion of the hand to dismiss an offer); and referential gestures, which can refer to something concrete or to an abstract concept (Müller 1998, 110–13; Cienki 2008, 8; Müller’s approach to the classification of gestures is similar in important respects to that offered in Rimé and Schiardatura 1991, 242–49). Whereas the first two types of gesture can be understood to be strongly associated with the basic function of language in human cultures as I characterized it in chapter 1—discourse gestures serve to help the speaker and her audience negotiate a path through the ideas current within the shared referential space of discourse, and performative gestures provide a corporeal representation of the illocutionary force of a speech act—the third type draws on the more broadly deployed resources of analogical reference discussed in chapter 2. From the perspective of Müller’s classification, then, McNeill’s iconic and metaphoric gestures differ only in the things to which they refer rather than in the mechanism of referral. To illustrate, consider an attested case in which a speaker accompanied the

Example 4.7 Kendon’s continuum: toward the left of the continuum, the connection between movement and language is obligatory; toward the right, the connection between movement and language is not obligatory; adapted from D. McNeill 2011

Gesticulation → Speech-Linked → Pantomime → Emblems → Sign-Language
utterance “And he went up a winding staircase” with a spiral motion he made with his index finger pointing upward (Iverson and Goldin-Meadow 1997, 458). The speaker’s gesture thus provided an analog for the path an agent would make in ascent: the finger correlates with the body of the agent, and the movement of the hand as a whole with the path of the agent in ascent. Having observed this, the speaker’s audience could then infer that the staircase is arranged in a spiral, information that is not provided by the utterance alone. But this same gesture could accompany an utterance such as “And things started to get out of control,” in which case the gesture would represent an abstract concept (with the apparently aimless spiral motion serving as a metaphor for loss of control). For McNeill, the first gesture would be classed as iconic, the second metaphorical; for Müller, they serve the same basic function within discourse, namely referring to concepts that might be either concrete or abstract in nature. I would also note that Müller’s broad notion of referential gestures conforms with the approach to iconic signs that I developed in chapter 2, and in particular the idea (which I developed within a musical context) that such signs could be situated along a continuum ranging from those with great fidelity to the process they are intended to represent to those which bear relatively few structural correlates with their target.

The example of the winding staircase and the spiral gesture gives a clear illustration of the singular contribution made by referential gestures: they provide information not found in the accompanying utterance (in the first case, the path an agent takes in ascending a staircase; in the second case, a visualization of what loss of control involves). There is also evidence that gesture may actively shape thought. As part of his research McNeill would videotape subjects as they described a short scene from an animated cartoon. In his analysis of these videotapes, he noted subjects’ making recourse to recurring patterns of gestures, which he came to call catchments. (The recurring patterns of movements that Astaire makes as he sings “The Way You Look Tonight” could be interpreted as analogous to catchments. For a discussion, see Zbikowski 2011.) According to his interpretation, catchments momentarily capture salient features from the flow of thought that motivates both speech and gesture so that these features can be related to one another. This then creates a dialectic between the holistic analog elements of gesture and the combinatorial symbolic elements of language (D. McNeill 1992, chap. 9; 2005, chap. 4). The dialectic is manifested in what McNeill calls a growth point, in which an opposition between the images associated with gesture and the formal units of language creates a momentary instability in thought. This instability, although relatively brief, determines the ultimate shape taken by an utterance, and thus the ideas the speaker conveys through gesture and speech. Kendon offers a handy summary of this aspect of McNeill’s work: “The growth point is the ‘idea unit’ from which an utterance arises and which the utterance, when fully fledged, still contains” (2004, 78). As McNeill’s analyses of his subjects’ descriptions show, over the course of a description a subject can be seen to be not only shaping her thought through gestures (as suggested by the various gestural catchments she produces) but also reshaping thought as gesture and language interact with one another (D. McNeill 2005, 112–21).
A different sort of evidence for the way gesture shapes thought is provided by Goldin-Meadow’s research on how young children use gestures. Goldin-Meadow noted that in some cases a child’s gestures and speech are in close conformance with one another, but in other cases they are not. Particularly interesting are cases in which gesture introduces new information that is not mentioned in speech (which Goldin-Meadow characterizes as a gesture-speech mismatch), for these indicate that gesture and speech are, if only temporarily, drawing on different sources of knowledge. “A speaker who has produced a gesture-speech mismatch knows (at some level) the information conveyed in both modalities. However, the speaker has not yet developed a framework . . . within which those pieces can be fitted together” (Goldin-Meadow 2003a, 29). Once the child has developed the necessary framework, she will be able to bring her speech and gesture into conformance with one another—that is, the information associated with the two different modalities will become conceptually unified.

One issue that has bedeviled research on gesture and language is the ultimate relationship between the two modes of expression. Both McNeill and Goldin-Meadow take the position that, at root, gesture and language are manifestations of a single unified system of communication (D. McNeill 1992, 2; 2005, 24–29; Goldin-Meadow 2003a, 16). It must be noted, however, that the goals for a system conceived in this way come almost exclusively from language: there is no independent motivation for communicating with gestures. Although McNeill recognizes the possibility of gestural communication in the absence of language (in the form of pantomime), he construes it in terms of linguistic communication with speech suppressed: “‘Pantomime’ is dumb show, a gesture or sequence of gestures conveying a narrative line, with a story to tell, produced without speech” (2005, 5). In brief, while gesture often provides information not given in the speech stream and may shape the ideas expressed through gesture and speech, the theoretical framework offered by McNeill and Goldin-Meadow does not see gesture as providing conceptual information distinct from that of speech.

Remedying this situation requires at least two expansions of our perspective on human communication. The first builds on the three defining characteristics of concepts I set out toward the end of chapter 2: (a) concepts are cognitive constructs stable enough to be stored in memory; (b) they are an important resource upon which organisms draw to guide present and future actions; and (c) concepts of one sort (say, musical or gestural) can be related to concepts of another sort. As such, conceptual knowledge can obtain in a wide variety of expressive media, including music and gesture, and need not be seen as directly dependent on language.

The second expansion of our perspective on human communication follows from the idea, also set out in chapter 2, that humans have available to them two different forms of reference: symbolic reference and analogical reference. In the case of combinations of speech and gesture, language makes use of symbolic reference whereas some (but by no means all) gestures make use of analogical reference. More specifically, gestures often serve as kinesthetic analogs for dynamic processes. Although the gesture involving a spiral motion with the index finger pointing upward produced with the utterance “And he went up a winding
Music and Gesture

"staircase" has typically been regarded as an example of a straightforward iconic gesture (with the spiral motion a visual picture of a spiral staircase; this is the interpretation offered in Iverson and Goldin-Meadow 1997), I would argue that the dynamic interpretation that I offered—the finger correlating with the body of the agent, the movement of the hand as a whole correlating with the path of the agent in ascent—provides a richer account of the contribution made by gesture. From this perspective, then, the referential gestures that accompany speech have a deep affinity with the basic elements of music: both are concerned with providing analogs—either through movement or through sound—for dynamic processes.

One thing left out of the notion of reference developed in chapter 2 was an account of the social and cognitive resources that make it possible to use symbolic or analogical reference to communicate. Indeed, the system of communication that results is apparently unique to humans: we know other species have, at the very least, a limited capacity for analogical reference and that some species appear to understand (albeit with the intervention of training by humans) symbolic reference, but none have created anything like the complex system of communication that is negotiated with apparent ease by children as young as three or four years of age. What other species appear to lack is an infrastructure of shared intentionality, and it is to an examination of this infrastructure that I would now like to turn.

The Infrastructure of Human Communication

Communication by Apes, Communication by Humans

Over the past two decades a number of investigators have explored the gestures made by apes (a category that includes chimpanzees, bonobos, gorillas, and orangutans). As Michael Tomasello has pointed out, we now know that apes can make a range of gestures (including those directed toward a specific individual), can follow the gaze of whoever they are interacting with, and have an understanding of the goals or intentions of others (Tomasello 2006, 507–10; see also the essays collected in Call and Tomasello 2007). These skills allow apes to use gestures in one of two principal ways: as a way to signal an intention (such as when an infant touches the back of its mother so that it might get a ride); or as a way to get attention (such as slapping the ground and looking at another ape with which it wants to play) (Tomasello 2008, chap. 2). But, despite these skills, there is one relatively simple gesture that apes do not perform: pointing. Tomasello has proposed that this reflects apes’ inability to enter into a joint attentional scene (which, as noted in chapter 1, is one of the abilities demonstrated by human children as they approach their first birthday) and their lack of an interest in helping others (an interest shown even by preverbal human children). Without these two things, apes have very limited capacities for cooperating with one another and, as Tomasello argues, it is this aspect of cooperation, rather than the ability to use symbols, that distinguishes human communication from other forms of communication in the animal world (Tomasello 2006, 2008, chap. 2).
As an example of the fundamentally cooperative nature of human communication, consider a situation in which I meet you in the concert hall prior to a performance and point to an empty seat. My gesture could be interpreted in any number of ways: I could be indicating where I intend to sit, where you should sit, where a mutual friend will be sitting, or the fact that the concert is almost completely sold out and only one empty seat remains. At least two things are required to interpret my gesture: a shared frame of reference (relative to which the information I provide can be understood to apply to myself, to you, to our friend, or to the popularity of the concert) and the knowledge that I offer this information in order to be helpful. Within the context of interactions between humans, the communication engendered by my pointing to the empty seat involves a collaborative effort between the two of us that takes advantage of a referential frame that is shared by both of us, as well as shared ideas about how the gesture of pointing can be helpful. As Tomasello sees it, this collaborative effort takes advantage of what the psychologist Herbert Clark has called common ground—that is, shared knowledge among those involved in a communicative act (Tomasello 2008, chap. 3; H. Clark 1996, chap. 4). One of the ways such knowledge is built up is through what Clark calls perceptual copresence, in which those involved in communication attend to the same things with some confidence that they have done so in the same way; indeed, in Clark’s view gestures are one way to establish common ground through mutual attention to the movements of communicators (1996, 112–13).

Although there is more to Tomasello’s account of human cooperative communication, I would like to pause for a moment to consider how Clark’s idea of perceptual copresence—something that is, incidentally, basic to the performance practice of a wide range of musical ensembles—might be developed. For his part, Clark simply assumes that the sharing of attention contributes to a sense of shared experience (and thus the construction of common ground), but recent research on the neuronal structure of the motor cortex of primates has changed significantly our conception of what it means to share perceptual information.

The Influence of the Motor Cortex on Cognitive Processes

For much of the twentieth century the neurological structures through which motor movements are controlled, which are collectively called the motor system, were assumed to be relatively passive: commands that originated elsewhere in the brain told the motor system what to do, and it did it. Beginning in the late 1980s, however, it became evident that the design of the motor system was not quite as simple as this. As Giacomo Rizzolatti and Corrado Sinigaglia have noted, research done during that time showed that the motor system is made up of a complex web of cortical areas that are anatomically and functionally different, and in which clear distinctions between perceptual information and motor action are blurred (Rizzolatti and Sinigaglia 2008, 20). Three striking discoveries contributed to this view, all of which came out of research on macaque monkeys and which focused on an area within the monkey brain called F5, long known to be associated with motor actions. First, neurons within the motor system are associated with specific
motor tasks: the neurons involved in grasping, for instance, are different from those involved in tearing (Fadiga et al. 2000, 171).

Second, a portion of the neurons in F5 respond to the visual presentation of an object, and these are again differentiated: those that fire when the monkey views a ring are, in most cases, different from those that fire when the monkey views a sphere (Fadiga et al. 2000, 172). Researchers came to call these neurons “canonical neurons,” to distinguish them from the neurons involved in specific motor actions. Francesca Garbarini and Mauro Adenzato, in a review of this research, observed,

The most interesting aspect of canonical neurons is that the same neuron fires not only in response to the same object, but also in response to a group of objects that have the same characteristics, in terms of the type of interaction they allow. At this level of description, an object can be codified “on relational terms,” i.e., it can be identified and represented in relation to the type of action that it affords an interacting subject. This means that the type of interaction that is established with an object is a constitutive part of the representation of the object itself. (Garbarini and Adenzato 2004, 102)

The emphasis that Garbarini and Adenzato give to “affords” reflects their belief—shared with other researchers in this area—that the presence of canonical neurons supports James J. Gibson's notion of affordances (Rizzolatti and Sinigaglia 2008, 34; Gibson 1979, 206). For Gibson, perception involved not simply the relatively passive gathering of information by an organism but the consequences for actions afforded by this information. Drawing examples from Eric Clarke, we could say that, for a human being, a chair affords sitting, a stick affords throwing, raspberries afford eating, and a sharp pencil affords writing (E. Clarke 2005, 38). For a termite, by contrast, the chair (if wooden), stick, and pencil would all afford eating, while raspberries would afford sitting. Affordances thus reflect how—and, to a certain extent, why—the perceptions of an organism shape its actions and its actions shape its perceptions (Chemero 2009, chap. 7).

The third, and perhaps most striking, discovery produced by research on the motor system was of a further set of neurons that became active both when the monkey executed a motor act and when it observed an experimenter doing the same motor act; these have come to be called mirror neurons. Rizzolatti and Sinigaglia characterize the unique properties of mirror neurons in this way:

The motor properties of the mirror neurons are identical to those of other F5 neurons in that they discharge selectively during specific motor tasks, but their visual properties differ significantly. Unlike the canonical neurons, mirror neurons do not discharge at the sight of food or other three-dimensional objects, nor does their behaviour appear to be influenced by the size of the visual stimuli. In fact, their activation depends on the observation of specific motor acts in which a body part (such as a hand or the mouth) interacts with an object. (Rizzolatti and Sinigaglia 2008, 80; translation adapted)

Further research has demonstrated that a portion of F5 mirror neurons respond to hand actions even when the final part of the action—for instance, the grasping
of an object—is hidden from the monkey (Umiltà et al. 2001, 160). As Vittorio Gallese has observed, “These results seem to suggest that inferences about the goals of the behavior of others appear to be mediated by the activity of motor neurons coding the goal of the same action in the observer’s brain. Out of sight is not ‘out of mind’ just because, by simulating the action, the gap can be filled” (2005, 33). Although Gallese’s use of the notion of simulation is somewhat different from that proposed by Lawrence Barsalou (which was described in chapter 2), there is converging evidence that simulation—whether of particular motor actions, or of objects and events as they are recalled—is one of the key cognitive processes through which we interact with the world (Barsalou 2008; 2009).

It should be noted that most of the recent research on canonical and mirror neurons has been conducted on monkeys, using invasive techniques of a sort not suitable for human subjects. That said, there is evidence both from brain-imaging studies (Calvo-Merino et al. 2005; Bangert et al. 2006) and from single-neuron studies (Mukamel et al. 2010) that similar structures are a feature of the human brain. It also bears mention that while there has been a fair amount of emphasis on the activation of motor neurons by visual stimuli, there are studies showing similar responses to aural stimuli for both monkeys (Kohler et al. 2002) and humans (Gazzola, Aziz-Zadeh, and Keysers 2006; for a review and discussion, see Overy and Molnar-Szakacs 2009). All of this work suggests that there is substance to Clark’s notion of perceptual copresence, with the caveat that what is being shared is not simply attention but patterns of motor activity appropriate to the circumstances at hand. The common ground basic to human communication, through its manifestation in the motor system, is thus a part of our basic cognitive architecture.

### Intentional Understanding

**Intentional Action.** The changed view of cognitive architecture that follows from research on the motor system has had as its consequence a much better understanding of how cognitive processes are grounded in embodied experience. There is, however, at least one significant challenge presented by current interpretations of the significance of this research, and that is how knowledge about the goals of actions—as distinct from the actions themselves—comes to be encoded at the neuronal level. As noted above, there is experimental evidence that mirror neurons can be activated even when the salient object—the goal of the action—cannot be seen. This evidence has led Gallese and others to argue that individual neurons encode not only actions upon objects but the goal of the action itself. Were this so it would be truly remarkable, inasmuch as goals are connected not only with the motor actions through which they might be achieved but with the needs of the organism as a whole. An alternative interpretation, offered by Gergely Csibra, divides the neuronal activation of motor programs into low-level and high-level processes. Mirror neurons, which are tuned to visual and aural information about actions, become automatically activated whenever an action matches that which is coded by the neurons and are thus relatively “low level.”
Knowledge about the goal of an action, by contrast, requires at the very least an analysis of the components of an action hierarchy—for instance, the coordinated movement of arm, hand, and fingers used to interact with some object, a sequence that will take one form when picking up a heavy book, another when picking up a china cup—which must then be reconstructed when the action is cognitively simulated. As Csibra explains,

> Low-level mirroring (i.e. “direct matching”) and bottom-up propagation supports a simulation system that facilitates the understanding of the goals behind observed actions, while high-level mirroring and top-down propagation (i.e. “action reconstruction”) allows predictive emulation of observed actions on the basis of the high-level interpretation of the action achieved without low-level motor activation. (Csibra 2008, 442)

Put another way, the observer must have some knowledge of the intentions and goals of the individual doing the actions to be able to simulate those actions.

According to the perspective developed by Csibra, then, knowledge about the goals of actions is not encoded at the neuronal level, but instead reflects an understanding of the intentions of the individual being observed. In their recent work on the origins of cultural cognition, Tomasello and his colleagues have proposed that the capacity to understand intentional action is one that is shared across primate species (Tomasello et al. 2005). They also note, however, that while this capacity supports cultural learning by apes and by human children up to around the age of one year it is not sufficient for the more complex forms of social interaction—including those associated with sharing information in a helpful way—that distinguish human cultures. For these sorts of interactions, individuals need not only to understand that others have goals and intentions, but also to be able to share these goals and intentions.

**Shared Intentionality.** Midway through his 1957 essay on meaning, the philosopher H. P. Grice offers as an example two contrasting cases. In the first, he imagines showing Mr. X a photograph of a particular situation; in the second, he imagines making a drawing of the same situation and showing it to Mr. X (1957, 382). Grice argues that the difference between the two cases is that Mr. X will take what he will from the photograph even if it were left laying around, but in the case of the drawing he will realize that, in making the sketch, Grice intends that he should develop a particular belief about the situation that is depicted. Grice then gathers together various points he has developed in the course of the essay, setting out what he believes is necessary for the sort of meaning in which he is interested:

> Perhaps we may sum up what is necessary for A to mean something by $x$ as follows. A must intend to induce by $x$ a belief in an audience, and he must also intend his utterance to be recognized as so intended. But these intentions are not independent; the recognition is intended by A to play its part in inducing the belief, and if it does not do so something will have gone wrong with the fulfillment of A’s intentions. Moreover, A’s intending that the recognition should play this part implies, I think, that he assumes that there is some chance that it will in fact play this part,
that he does not regard it as a foregone conclusion that the belief will be induced in the audience whether or not the intention behind the utterance is recognized. (Grice 1957, 383–84)

Two important things follow from the argument Grice develops. First, without the audience's recognition that A intends to induce by x a belief—that is, that he intends, in the simplest sense, to change their minds—the meaning of x will be lost. The audience must thus share A's intention. And, as Grice observes, A’s successful induction of the belief associated with x is predicated on the audience's recognition of his intention: A recognizes that the audience has to recognize his intention if they are to develop the belief he intends. Second, the shared intentionality basic to the sort of exchange with which Grice is concerned is independent of the system of symbolic reference basic to language: Grice's drawing, as an analogical representation of a particular situation (and, as such, something close to that form of the hypoicon Peirce called a diagram), has meaning only within a context of shared intentionality.

Within the context developed by Grice, this shared intentionality has three components. First, the individuals involved in the interaction must be mutually responsive to one another. Second, the individuals must share a goal—here, the development of a belief. Third, the individuals must coordinate with one another to achieve the goal: if A realizes that the audience has not recognized his intention he must find a way to get them to recognize it; if the presentation of A’s utterance of x is somehow marred, the audience must ask him to restate it or reformulate it so that the belief it is intended to induce becomes manifest. Within the more general framework developed by Tomasello and his colleagues, these three components of shared intentionality are basic to the cooperative interactions that typify human cultures—that is, for cooperation to occur, the goals and intentions of each individual must include as content something of the goals and intentions of the others involved in the activity (Tomasello et al. 2005, 680). And it is cooperation of this sort that was necessary for our forebears to hunt successfully certain forms of prey and which is also essential to the various forms through which humans communicate with one another.

A Cooperative Model for Human Communication

In the model of human communication developed by Tomasello in his subsequent work, the capacity to understand that others have intentions—and, more specifically, that some of these intentions reflect a willingness to share information in a helpful way—is crucial. As he sees it, the symbolic structure of language rests on a nonlinguistic infrastructure of intentional understanding and common conceptual ground (Tomasello 2008, 58). This infrastructure provides the basis for the sort of cooperation that marks human communication, which he proposes is guided by three social motivations. The first of these is requesting. As Tomasello notes, getting another to do something that you want is characteristic of the intentional communication signals of all apes but is elaborated by humans into imperatives
Music and Gesture

that range from orders to polite requests to suggestions to hints (Tomasello 2008, 84). The second motivation, which is apparently unique to humans, is altruistic helping—that is, offering assistance to others even without being requested to do so. Just why such a motivation might have developed in our species is a matter of speculation, but there is some experimental evidence that chimpanzees tend to choose to help or cooperate with others who have a good overall reputation for helping and cooperating (Tomasello 2008, 200), suggesting that other primates might have an inclination—but no more—toward cooperative behavior. The third social motivation is a desire to share feelings and attitudes about things with others. Tomasello illustrates this motivation thus: “For example, on a beautiful day it is quite common to say to your officemate upon arrival at the office, ‘What a beautiful day today!’—which derives not from any imperative or informative motive involving help, but rather from a purely social one. This kind of communicative act is simply a sharing of attitudes and feelings so as to expand our common ground with others” (2008, 86). And it is through expanding our common ground with others that we develop affiliations that strengthen our membership in a particular group. As Tomasello notes, “To an unprecedented degree, human groups mark themselves to make sure of group membership, even having group-specific greetings—a unique kind of speech act—that serves, in part, to solidify group membership” (2008, 209–10).

Consistent with his overall approach, Tomasello’s discussion of the sharing of attitudes and feelings is framed around the social uses of language. I would be remiss, however, if I did not note that a musical instantiation of his example of sharing attitudes is provided by Rodgers and Hammerstein’s “Oh, What a Beautiful Morning” (briefly mentioned in chapter 1). As shown in example 4.8, the verse of the song (which sets the opening soliloquy) has a relatively restricted range (staying within a perfect fifth until just before the chorus) and makes use of predominantly stepwise motion. In contrast to this, the melody for the chorus (which sets “Oh, what a beautiful mornin’; Oh, what a beautiful day”), through the sustained vocal athleticism of running arpeggios and a marked expansion of the vocal range, proceeds with an exuberance that matches—if not exceeds—the sentiment of the words. And as if this were not enough, in the arrangement used for the appearance of the song in the 1943 musical Oklahoma! the accompanying instruments remain silent until measure 24. The lilting waltz that typifies the chorus is but hinted at prior to measure 29; once the chorus begins, however, both melody and accompaniment adopt the rhythm of the dance, which, through its manifest physicality, gives an embodied dimension to the sharing of attitudes and feelings implicit in the lyrics.

In brief, then, Tomasello’s model for the sort of communication that is apparently unique to humans begins with an infrastructure of intentional understanding and common conceptual ground. While there is some evidence that other primates share this infrastructure (especially if we understand common ground as shaped by the apprehension of patterns of motor activity) they do not appear to share the social motivations that lead to complex kinds of requesting, to altruistic helping, or to a desire to share feelings and attitudes as a means to expand common ground. Although Tomasello’s principal goal in elaborating this model was to develop a
Foundations of Musical Grammar

fuller understanding of the role of language in human culture, much of the model applies to the gestures that often accompany speech but that may also—as in the case of my example of pointing to the empty seat in the concert hall—occur without the immediate support of speech. And, to the extent that such gestures involve analogs for dynamic processes (as distinct from a simple act of deixis), I believe the model also has clear relevance for musical communication.

Music and Human Communication

If we think of music as not simply patterned nonlinguistic sounds but as a set of practices shared between everyone within the compass of those sounds—that is, as an essentially social activity—its contribution to common conceptual ground is
immediately evident: those sounds are one way to achieve the perceptual copresence that underlies common ground. Further, research showing that motor neurons can be activated by aural stimuli suggests that this copresence extends to a fully embodied experience: the common ground established by music would thus include not only shared attention to a sequence of sounds but shared knowledge about the motor activities associated with the production of these sounds. This sonic common ground can be put to a number of uses within human societies, including (as suggested in chapter 3) the establishment of a shared context for the performance of a ritual. Within such a context music is especially important as a medium for sharing feelings and attitudes, thus serving as a means to cultivate the social affiliations that are central to the sort of cooperation that distinguishes human communicative interactions (Dissanayake 2006, 43–49; Tomasello 2008, 208–10; Kirschner and Tomasello 2009, 2010).

I should note, however, that I regard success in the endeavor of sharing of feelings and attitudes through music as a consequence of using patterned nonlinguistic sound in an intentional way. That is, the listener needs to hear these sounds as music. If, for instance, the opening of the chorus of Rodgers and Hammerstein’s “Oh, What a Beautiful Morning” were used as a ringtone for a mobile phone or as a cue that a door operated by an electronic lock could be opened, its function as music would be deprecated to the point of becoming negligible. When, by contrast, that sequence of sounds is used as a means to shape thought (rather than simply as a prompt to answer a phone or open a door)—when it is recognized as being intended to do just this (an intent made manifest by the musical utterance that extends beyond and expands upon the opening moment of the chorus)—it becomes a means to a particular type of communication. According to the theoretical perspective I develop here, this type of communication involves sonic analogs for dynamic processes. As such, it is not very well adapted to making requests, and may be of limited use in altruistic helping (although the use of music in therapeutic situations suggests how this might be possible), but may be more effective than language for sharing feelings and attitudes.

The Infrastructure of Human Communication

A strong argument can be made that what makes human communication unique is its essentially cooperative nature. Built on an infrastructure of common conceptual ground and shared intentionality, human communication is guided by three social motivations: making requests, altruistic helping, and sharing feelings and attitudes. As I see it, the infrastructure of human communication is necessary for both symbolic and analogical reference, which, in turn, make it possible to request, help, and share through various combinations of sonic, physical, and imagistic tokens.

Both language and gesture manifest all of the social motivations for cooperative communication proposed by Tomasello—indeed, these motivations are, in part, reflections of the view he has developed of the ways language and gesture function within human cultures (Tomasello 1999, 2003, 2008). Music, by contrast,
puts greater emphasis on sharing feelings and attitudes, not least because it offers ways to represent through sonic analogs the dynamic processes with which feelings and attitudes are associated.

As I noted in my introductory comments for this chapter, most of my attention in this and the preceding section has been on features shared by music and gesture—the music as gesture approach—and in particular on their mutual reliance on analogical reference for the representation of dynamic processes. In the following section I shall take up the matter of differences between music and gesture that become manifest when the two combine to create what Nicholas Cook calls an instance of multimedia (1998, chap. 3), a combination that reflects the conjunction of music and gesture. My aim will be to develop a clearer understanding of how the resources recruited by musical grammar shape the part music plays in such instances, and indeed how those resources are different from what can be offered by gesture alone.

Music and Movement

The 1941 film The Great Dictator was, for Charlie Chaplin, both a personal reaction to the policies and images of Adolph Hitler’s Third Reich and a deliberate effort to expand artistic horizons. In the film, Chaplin plays the dual roles of Adenoid Hynkel (a direct parody of Hitler) and a character known simply as the Jewish Barber (based on Chaplin’s the Tramp). These doubles are separated by character (Hynkel is a craven, psychotic despot endowed with relentless energy; the Barber, a mild-mannered Everyman with few aspirations extending beyond his barbershop and yet possessed of an underlying nobility) and by their regard of fate (Hynkel obsessed with fashioning destiny; the Barber concerned only with adapting to the circumstances of the moment). About midway through the film the storyline shifts to a scene in the barbershop, introduced by the voice of a radio announcer proclaiming, “This is the Happy Hour Program. Make your work a pleasure; move with the rhythm of music. Our next selection: Brahms’ Hungarian Dance Number 5.” We are then treated to a marvelously improbable and exquisitely choreographed scene in which the Barber shaves a customer in time to Brahms’s intensely energetic and self-consciously dramatic music as it is performed in an arrangement for full orchestra. (In point of fact, the dance is not, as Brahms apparently believed, from Hungarian folk traditions but is a csárdás entitled “Bártfai emlék” (“Memory of Bartfeld”) written by Béla Kéler; see Kalbeck 1976, 1: 65. For the sake of simplicity I will continue to refer to the work as “Brahms’s Hungarian Dance No. 5” since this indexes the version used by Chaplin.)

One of the keys to the success of the scene is Chaplin’s close attention to the relationship between movement and music, according to which he matched, wherever possible, discrete patterns of physical action to structurally significant spans of music. For instance, Chaplin correlates the second statement of the principal theme of the dance (shown in example 4.9) with the application of shaving cream to his customer’s face (example 4.10). Over the course of the
first eight measures Chaplin's movements are quite uniform, with each successive application of shaving cream matched to one measure of music. Against the varied and more animated version of these measures that follows, his strokes become shorter and more rapid until, at the end of the passage, he dispenses with the brush and uses his hand to smooth the shaving cream over his customer's face with a speed and frequency that matches the sixteenth notes of the music. (For a slightly different interpretation of Chaplin's movements in this scene, see Godøy 2010, 104–7.)

Part of the humor of the scene comes from the Barber's willingness to follow the letter of the announcer's exhortation, bringing the various actions associated with shaving into conformance with the "rhythm" of Brahms's dance. Another part
comes from the correlation of the music with actions that are either soundless or that produce sounds very different from those provided by the music. Yet neither of these can account for why this particular scene is as effective as it is—there are a number of ways that the Barber could have responded to the Happy Hour music and a range of sounds, musical or otherwise, that could have been correlated with the actions of shaving. Indeed, most of Chaplin’s movements, as a sequence of actions taking place in a barbershop, are not especially comedic: if one watches the scene without the soundtrack one is struck by the occasionally frenetic pace of Chaplin’s actions and the way certain movements are exaggerated, but these things are not in themselves funny, just a bit odd. As a set of movements closely coordinated with Brahms’s music, however, they do become funny, for they take on a purposiveness that exceeds that associated with the simple act of giving a customer a shave. This purposiveness is a consequence of the way Brahms has arranged his musical materials to create an analog for a dynamic process that moves relentlessly forward to its conclusion.

To explore the way Brahms accomplished this, I should like to return to the analytical approach introduced in chapter 3, considering first the syntactic processes used in various syntactic layers before moving on to discuss interrelationships between syntactic layers. This consideration will also provide an opportunity to specify further the features of syntactic processes as they relate to the resources exploited by different syntactic layers, which will in turn shed light on correlations between music and physical movement.
Syntactic Processes and Syntactic Layers in Brahms's Hungarian Dance No. 5

Rhythmic Processes. In that it aims to evoke the energy and drive of the csárdás, Brahms's dance hews closely to the traditions of compositional strategies for social dance in general and for the csárdás in particular (Szabolcsi 1964, 63). These include a very regular rhythmic framework: as can be seen in example 4.9, the left hand of the piano sets up a chugging rhythm in measures 17–28 with single pitches on beats 1 and 2 (the pitch on beat 2 lower in register than that on beat 1) and higher-register chords dividing these beats. What might be regarded as one of the less remarkable features of this framework—its unwavering regularity—points to a dynamic process essential to the syntax of temporal expectation: repetition. Although a single instance of a pattern may generate a degree of expectation, that expectation will largely be a function of stylistic norms extrinsic to the pattern (Huron 2006, chap. 11). A second instance of that same pattern, however, leads to the expectation that a third will follow. When expectations generated by the process of repetition become particularly robust, entrainment—which has both psychological and physiological components—may result, and entrainment has been shown to be important for the comprehension of musical spans (Thaut 2005) as well as for musical meter (London 2012, chap. 1). In the first instance, then, the syntactic process of repetition used in measures 17–28 gives rise to a framework for temporal expectation that provides a basis for entrainment and for understanding relationships between temporal events that co-occur with this framework. Set against this framework are the various rhythmic figures of the right hand which, framed by the temporal expectations set up by the left hand, create an analog for a process that grows in animation before coming to a rather sudden halt.

One of the principal ways the rhythmic figures of the right hand are organized is through duple groupings. Measures 17–18, which share a long-short rhythmic pattern, form one pair, and measures 19–20—which embellish the preceding pattern with two sixteenth notes and then replace it with a held note—another. The rhythmic pattern of measures 19–20 is repeated in measures 21–22 (confirming the pairing of measures) and then disrupted by the sudden burst of activity in measure 23 followed by the held note of measure 24. (As noted in chapter 1, changes of rhythmic strategy of the sort exemplified by measures 23–24 are often associated with the syntactic process of cadence, which serves as a sonic analog for arrival.) Measures 25–28 return, for the most part, to the rhythmic pattern of measures 17–20, the exception being the sudden flurry of sixteenths introduced in measure 25 that lead to the downbeat of measure 26. The replication of the pattern of rhythmic figuration set out in measures 17–24 is, however, disrupted in measures 29–30 by another and longer succession of sixteenth notes that includes off-beat accents (created by octave reinforcements of the third sixteenth note in each group of four), which momentarily pull the listener's attention away from the first beat of the measure. While the succession of sixteenth notes continues into measure 31, that measure brings the pattern of rhythmic accents back into line
with that created over the course of measures 17–28. This alignment achieved, the succession of sixteenth notes breaks off on the second beat of the measure; this change coincides with the first interruption of the steady sequence of eighth notes in the left hand since measure 17, after which the hands come together for the two emphatic chords of measure 32.

In summary, the syntactic processes through which Brahms organizes the rhythmic materials of this passage begin with the regularly repeated patterns of the left hand, which encourage the listener to entrain to the succession of musical events and thus provide a framework for the comprehension of the rhythmic figures set out by the right hand. The right hand’s figures are in turn organized by syntactic processes that deploy, on the one hand, repetition to construct and reinforce duple groupings and, on the other hand, variation to distinguish groups from one another at the level of both two- and four-measure groupings. As implemented by Brahms, these variations animate the rhythmic surface (through the introduction of sixteenth notes) and complicate expectations (through the off-beat accents of measures 29–30). The result is a sonic analog for a process that gradually increases in complexity over a relatively brief span of time, culminating in a pell-mell rush (measures 29–30) that leads to a relatively sudden conclusion (measures 31–32).

Harmonic Processes. One of the attractions of Brahms’s Hungarian Dance especially evident in measures 17–32 is the brusque simplicity of his harmonic language. Making do with little more than tonic, dominant, and subdominant he charts a confident and unwavering harmonic course through the tonal landscape of the dance. To be sure, the broad outlines of this landscape are part and parcel of the traditions of Western European dance music; that said, the specific course Brahms traces is a consequence of the syntactic processes he deploys, which organize the constituent sonic analogs of the passage to create an analog for a dynamic process that moves surely and somewhat precipitously toward its goal.

The sonic analogs set out by the passage are of three different types, each of which spans four measures. The first type, exemplified by measures 17–20 and 25–28, provides an analog for a stable, albeit variegated, dynamic process. Musically, this is accomplished by regular appearances of F♯ at the beginning of each measure, the melodic arpeggiation of tonic harmony in measures 17–18 (or 25–26), and the neighbor-note motion provided by the leading-tone harmony of measure 19 (or 27) that leads back to tonic in measure 20 (or 28). The second type, set out in measures 21–24, provides an analog for a dynamic process that oscillates between instability and stability. This is accomplished by an alternation of harmonies in successive measures (which, of course, speeds up the rate of harmonic change): subdominant–tonic–dominant–tonic, with the last alternation marked by the appearance of E♯ at the beginning of measure 23. The third type, set out in measures 29–32, provides an analog for a dynamic process that begins in a state of fluid instability, which suddenly gives way to anchored stability. This is accomplished by a succession of parallel 6/3 chords (shown, in example 4.9, with the number 6 between the staves of measures 29 and 30) that leads to the cadential formula of measures 31–32.

These basic materials are organized into a larger whole to create a sonic analog for a goal-oriented dynamic process that unfolds in measured stages and that
becomes more complex: the first stage (measures 17–20) sets the basic ground; the second (measures 21–24) introduces an element of instability (which nonetheless returns to stability); the third stage (measures 25–28) returns to the focus on tonic of the first, albeit with a considerable expansion of the melodic arpeggiation of tonic harmony in the first two measures; and the final stage (measures 29–32) momentarily veers toward total instability (with the stepwise succession of first-inversion harmonies) before returning abruptly to stable ground. Although each stage concludes with an appearance of tonic, the effect is less one of a return after a departure and more one of a succession of variegated gestures, each more extravagant than the last but each concluding in a similar fashion.

Contrapuntal and Registral Processes. Although the processes proper to the outer-voice counterpoint of Brahms’s dance can, to some extent, be conceived of as separate from its harmonic processes, in truth the two are so closely intertwined that they yield very similar sonic analogs. This is less the case with the way Brahms disposes his pitch materials across the register of the piano, for this disposition offers a sonic analog for a process that begins in a controlled, steady state but that then suddenly and vigorously pushes at the boundaries it had previously established. Over the course of measures 17–24 Brahms’s disposition of pitch materials is quite normative: for instance, even with octave doublings the pitches of the right hand remain within a thirteenth until the arrival on tonic in measure 24. This normative disposition is suddenly abandoned in measure 25 when the right hand explodes into the upper reaches of the piano’s tessitura. The disposition of pitches in register changes again in measure 29: although the main melodic line of the right hand returns to its previous disposition, sporadic octave doublings keep the registral boundaries open; the left hand, for its part, moves into the tenor register in measures 29–30, returning to its previous register for measures 31–32. (In the original four-hand version, the primo piano part continues in the upper register throughout measures 29–32.)

It bears mention that the effect Brahms achieves through the way he disposes his pitch materials is somewhat attenuated in the orchestral arrangement of the dance by Martin Schmeling that was used for the soundtrack of The Great Dictator. There the registral expansion occurs only the first time these sixteen measures are heard; the second time—that is, in measures 17–32—the principal melodic voices are kept in the upper register throughout. What does occur in these measures is an enrichment of orchestral timbre (as the full complement of winds and brass is used throughout), together with a dropping out of the bass voices in measures 29 and 30. In consequence, on the soundtrack the sense of intensification wrought by the registral contrast between measures 17–24 and 25–32 in the piano version is replaced by the timbral contrast between measures 1–16 and measures 17–32 in the orchestral version.

Relationships among Syntactic Processes in Brahms’s Hungarian Dance No. 5. As might be expected in a work intended to evoke a rustic dance, there are numerous correspondences among the syntactic processes that play out in the different syntactic layers of Brahms’s composition. It is worth noting, however, that there are at least a few moments when processes are not synchronized: for instance, although processes within both the rhythmic and contrapuntal/registral layers offer analogs
for increasing intensity beginning in measure 25 this sort of increase is more weakly represented in the harmonic layer (the brisk arpeggiation of measures 25–26 notwithstanding), which returns in the main to an analog for a stable dynamic process of the sort with which the excerpt began. And when, beginning in measure 29, analogs created within both the rhythmic and harmonic layers are at their most animated the intensity of the sonic analogs set out by the registral layer has decreased markedly. (It is worth observing that the intensity of the sonic analogs created within the registral layer in the four-hand piano and orchestral versions remains constant over these measures, neither decreasing nor increasing.) As a whole, however, the interplay between syntactic layers yields a passage that summons a dynamic process moving relentlessly forward, becoming more intense as it does so. That this is a consequence of the way Brahms organized the sonic analogs encompassed within these measures is made evident if the two eight-measure phrases given in example 4.9 are switched, as they are in example 4.11. Instead of a passage that drives to a compelling conclusion we have one that begins with apparently unmotivated musical histrionics and that concludes in a decidedly anticlimactic way. Put another way, if the function of measures 17–32 of Brahms's dance is to provide a consistent framework for patterned movement with a forward-thrusting dynamic shape, the music of example 4.9 is grammatical and the music of example 4.11 is ungrammatical.

Meaning Construction in the Barbershop Scene from Chaplin's The Great Dictator

The compositional strategies that Brahms used to create a sonic analog for a dynamic process that moves with direction and purpose toward its goal are not, of course, unique to Brahms. Beyond the fact that he retained many of them from Kéler's version of the dance, they reflect a tradition of compositional strategies developed by composers within the Western tradition over the course of the seventeenth and eighteenth centuries. These strategies made it possible for composers to give their musical utterances an extraordinary sense of being directed toward both short- and long-term goals (McClary 2000, chap. 3). As a consequence, when the musical materials of Brahms's Hungarian Dance No. 5 are correlated with the movements through which a customer is given a shave—a correlation made possible by Chaplin's careful choreography—those movements acquire a sense of purposiveness they would not otherwise have. One way to account for this effect is through a conceptual integration network such as that shown in example 4.12. Following the basic methodology introduced in chapter 2, one of the mental spaces that serves as an input to this network is structured by the four four-measure units of Brahms's dance, each of which is distinct from the next. The other mental space is structured by Chaplin's actions as he prepares to shave his customer: a smooth application of shaving cream with his brush, followed by more rapid movements of the brush, followed by his hand vigorously finishing the application of the shaving cream. The generic space that captures the basic topology of the network involves dynamic processes that comprise clearly defined and differentiated units. In the blended space, elements of the input spaces are brought
together, giving rise to a mental space in which Chaplin's movements proceed with purpose toward a clear goal.

In themselves then—viewed without the soundtrack—Chaplin's movements in the scene have no larger significance beyond what would accrue to any film of one man shaving another. This is not to diminish the meaning such movements could have within a film—presented to us as an object for our attention we would almost certainly assume that they had some significance—but only to emphasize their relatively unexceptional character (and thus potential transparency). When correlated with Brahms's Hungarian Dance No. 5, however, these movements contribute to a conceptual blend in which the Jewish Barber's actions proceed with a
clearly focused direction and purpose. Within the overall storyline of *The Great Dictator* this imaginative transmutation is telling: up to this point the barber has been more or less a somnambulist, happy enough to have realized his dream of returning to his barbershop but having few ambitions other than to keep it tidy and to welcome the occasional customer. Subsequent to this scene he becomes much more engaged with his surroundings and, although ever a reluctant participant, begins to take actions that will transform him from a passive subject into an agent of change.

The interpretation of Chaplin’s movements that is made available by their correlation with music calls to mind a distinction made by Susan Goldin-Meadow and Sian Beilock in their work on action and gesture. Goldin-Meadow and Beilock propose that gesture is a special sort of action: where most actions have a direct effect on the world (in the sense of changing the features or dispositions of physical objects in the world), gestures do not and instead serve a representational function (Goldin-Meadow and Beilock 2010). Thus the movements through which we arrange various small blocks of wood into this or that pattern are actions, and
the movements we would use to characterize aspects of the resulting arrangement (or the means through which we achieved it) are gestures. (It bears mention that Alexander Refsum Jensenius and his colleagues, in their work on musical gestures, adopt definitions of action and gesture that are almost exactly the opposite of this. See Jensenius, Wanderley, Godøy et al. 2010, 13.) From the perspective I have developed in this chapter, however, I think it would be more productive to think of actions as movements that are not regarded as a constituent element of cooperative communication, whereas gestures are. And so, while the movements of a dancer at practice have no direct effect on physical objects external to herself—the dancer’s only goal is to make a particular movement—they correspond well with what we generally think of as actions. Were these same movements deployed within the context of cooperative communication (for instance, as part of an unspoken narrative; see, for instance, the discussion of dance practice within Australia’s Tiwi community in Grau 1995, 49–57), they would begin to fulfill a role closer to that associated with gesture. In the scene from The Great Dictator, the communicative function of Chaplin’s movements becomes manifest through their correlation with Brahms’s music. Within the conceptual blend consequent to this correlation, the movements participate in an instance of meaning construction through which the Jewish Barber’s character is developed, and thus they become something closer to communicative gestures.

According to this interpretation, the communicative aspect of Chaplin’s movements comes principally from the music: in the conceptual integration network shown in example 4.12 direction and purpose come from the mental space set up by the music rather than from the mental space set up by the movements. As Mary Ann Smart has shown, however, in the performance of opera—and specifically French and Italian opera of the first half of the nineteenth century—there is a rich tradition in which performative gestures are correlated with musical materials with the aim of creating a larger expressive whole (Smart 2004). In such cases, both music and movement—here understood as involving communicative gestures—would have the potential to make equal contributions to the cooperative communication proper to the blended space. As will be developed more fully in the discussion of relationships between music and words in chapter 6, the correlation of musical utterances with other communicative media provides an opportunity to specify in more detail what each medium contributes to the construction of meaning, and thus to deepen our understanding of each medium’s unique characteristics.

This brings us back, at length, to the movements Astaire makes as he sings “The Way You Look Tonight.” As I suggested earlier, Astaire’s gestures can be viewed as a visual instantiation of the sonic analogs for dynamic processes that make up Kern’s tune. In that music and movement are in close correlation, they can give rise to a conceptual blend in which communication is accomplished through both. Were Astaire’s movements not in close correlation with the music, the emergence of such a blend would be unlikely. Indeed, this is what happens with the movements Astaire makes with his hands as he mimes playing the piano: their lack of correspondence with the sounds that are central to the scene renders them incidental rather than essential. This is not to deny the potential effectiveness
of a performance of “air piano” (which Astaire, being an accomplished pianist, could have rendered), only that for a convincing performance—and effective communication—the correlations between music and movement should be many rather than few (something evident in the research by Rolf Inge Godøy and his colleagues on “air instruments”; see Godøy, Haga, and Jensenius 2006).

Music and Gesture

On the analysis I have offered here, Astaire’s movements provide a set of visuospatial images coordinated with both the music and the words of the song. They are, however, different from the spontaneous gestures that accompany speech studied by Kendon, McNeill, Goldin-Meadow, and their colleagues in at least one important way: Astaire’s movements are not part of a process through which thought is shaped, but instead reflect the shaping of thought accomplished through musical materials. In broad terms, this shaping of thought reflects the role of music in human cultures, and in particular the part it plays in the sort of cooperative exchanges that typify human communication. More specifically, however, there are two important resources that make possible intentional communication through sonic analogs: the establishment of a temporal framework for utterances involving sonic analogs; and making recourse to pitch specificity through which successive utterances, in part or in whole, can be compared to one another.

Temporal Frameworks for Musical Utterances

As we saw in the excerpt from Brahms’s Hungarian Dance No. 5, the left hand’s pattern of bass notes alternating with chords provided a contextual framework for the varied rhythmic figures of the right hand. Regular patterns of this sort support entrainment and offer a means to measure durational relationships: the chugging rhythm of the left hand creates a temporal yardstick that makes it possible for the listener to appreciate that the first melody note of measure 17 lasts three times as long as the last. Further, the interaction between rhythmic strands has the potential to set up larger patterns of expectation, giving rise to the two- and four-measure groupings evident in Brahms’s dance. In a similar fashion, the accompaniment for Astaire’s performance of “The Way You Look Tonight” (which, although less frenetic, is not unlike that for Brahms’s dance) sets out a framework for Kern’s melody, which falls into two-, four-, and eight-measure groups. These larger patterns in turn offer a backdrop for the dynamic shapes Kern creates through repetition and variation: the accumulation of quarter notes over measures 3–8 culminates in a return to the same rhythmic figure with which the melody began (the two successive whole notes of measures 9–10), marking the beginning of the second half of the tune and offering a harbinger of the disruption that will be created by the cadential pattern of measures 11–14.
The frameworks for temporal expectation set up by Brahms’s dance and Kern’s song fall under the rubric of musical meter, and they have long been associated with the regulated patterns of human respiration and movement (Kirnberger 1776–79, 2: 114; W. McNeill 1995, chap. 3; Caplin 2002). Such frameworks also provide, at the level of both basic materials and syntactic organization, the basis for sonic analogs for a range of temporal processes, including those that involve acceleration or deceleration, or the interaction of multiple rhythmic strands. Not every musical utterance, of course, will establish a robust framework for temporal expectation, the recitative from Bach’s cantata “Ich habe genug” discussed in chapter 3 being one example. That said, musical utterances that exploit temporal frameworks contribute to the shaping of listeners’ musical understanding in two important ways. First, such frameworks make it possible for musicians and listeners to anticipate coming events and to imagine possibilities for motor movement (something evident in Steven Friedson’s discussion of the rhythmic practices of the Tumbuka [1996, chap. 5]). Second, the complex relational structures associated with temporal frameworks bear a resemblance to the notion of a schema as it was employed by Frederick C. Bartlett in his work on human memory (F. Bartlett 1932, chap. 10). As such, the most significant elements of such structures (for instance, the essential features of various musical meters) would be part of long-term memory, from which they could be called up to evaluate novel utterances—that is, they would be syntactic structures.

**Tonal Frameworks for Musical Utterances**

The use of fixed pitches, which allow a musician to return to the “same” pitch after some interval of time, is such a commonplace in musical expression that its importance for organizing musical utterances is easy to underestimate. Equally important, however, are the ways fixed pitch can be used to anchor musical concepts. In Kern’s “The Way You Look Tonight,” for instance, both the descending fifth with which the melody begins (which sets “someday”) and the descending octave that marks the culmination of the melodic process that unfolds over the first eight measures (which sets “of you”) conclude on D4, a strategy that forges a direct connection between two moments in the melody some twelve seconds apart (assuming the tempo Astaire adopts for his performance). (It bears mention that a temporal gap of this extent challenges the capacity of working memory, as discussed in chapter 2.) To be sure, these two descending intervals are also linked by the rhythmic figure that sets their constituent pitches (two successive whole notes) and their placement within the metric scheme (both begin an eight-measure section), but I would like to propose that the reprise of D4 is crucial to hearing the musical concepts of measures 1–2 and 9–10 as being of the same sort. The D5 of measure 9 can thus be heard as a goal of the accumulating processes of the preceding measures as well as the beginning of another syntactic process, one that concludes with the cadence of measures 13–14.

A given pitch can also serve as a point of reference within a musical landscape sketched by successions of pitches. For instance, the D4 of measure 2 acts as an
anchor for the melody as a whole: as noted in my opening discussion, measures 3–9 trace successive arcs that gradually rise in register before the octave descent of measures 9–10 brings the melody back to D4; measures 11–13 describe a more compact transit, which rises to A4 before returning once more to D4.

It is, of course, hardly surprising that the pitch-class D would be prominent in a D major tune; the significance of that pitch can also be understood to arise from the structural properties of the diatonic collection to which it belongs (Clough 1979). My point, however, is less about D as a pitch and more about fixed pitch as a resource for structuring sonic analogs. Syntactic processes that exploit this resource can give rise to analogs for processes of departure and return (as we saw in the movements from Bach’s “Ich habe genug”), goal-oriented processes (of the sort evident in Brahms’s dance), or a range of paths through the virtual topographies created by successions of musical pitches (all of which were noted at the end of chapter 3). And, keeping in mind the way longer temporal spans challenge the capacity of working memory, the anchor provided by fixed pitch reference may help to connect unfolding sequences of musical events with large-scale syntactic structures that are stored in long-term memory.

*Grammatical Resources for Music and Gesture*

The resources offered by temporal frameworks of the sort basic to musical meter and tonal frameworks established by the careful management of relationships among fixed pitches allow musicians to create sonic analogs with a degree of precision not in evidence in the sonic analogs associated with sound effects or linguistic prosody. Indeed, frameworks like these are an essential substrate for musical grammar, making it possible to specify how sonic analogs unfold through time and what sort of shape they take (with “shape” framed relative to the topography created through pitch relationships). Frameworks of a similar sort are not generally available for the spontaneous gestures that accompany speech, which may be why they are generally not thought to adhere to any sort of grammar (Goldin-Meadow 2003a, chap. 13; D. McNeill 2005, chap. 2; 2012, 191–94). That said, the sort of contextual anchors so evident in musical organization do, on occasion, crop up within the domain of gesture. A striking example is provided by Astaire’s movements as he sings “The Way You Look Tonight.” Note the position of Astaire’s chin in the second image of example 4.3 (“day”), the last image of example 4.5 (“you”), and the last image of example 4.6 (“night”). In example 4.3 and example 4.6 his chin ends up in almost exactly the same place; in example 4.5 his chin is somewhat lower but is in approximately the same place with respect to the midline of the frame. Of course, each of these images captures an arrival point—that is, the end of a gesture. In general, there is no requirement for gestures to terminate at close to the same point in space, but this is just what happens with all three of Astaire’s movements. It is worth emphasizing that the musical structures that correlate with these movements conclude in exactly the same place: the melodic pitch D4. What gesture does only rarely and approximately, then, music that exploits the perceptual anchors provided by fixed pitch does frequently and with a high degree of
precision, a capacity that provides a reliable contextual framework for the sonic analogs of musical grammar.

Conclusion

Music as/and Gesture

The notion that Astaire's movements as he sings “The Way You Look Tonight” reflect the musical organization of Kern's tune might well seem counterintuitive: our usual habit of thought is to conceive of music as gesture, drawing on the visible, incorporated features of gesture to render concrete the ephemeral sonic analogs of music. Be that as it may, I take the position that both music and gesture (as well as other nonlinguistic expressive media) have their basis in conceptual knowledge independent of that of language, and that this knowledge can be shared through the resources offered by analogical reference. This particular species of communication—which, on the best evidence we have, is uniquely human—is built on an infrastructure of common conceptual ground and shared intentionality and, in the case of music, guided by the social motivation to share attitudes and feelings. Drawing on the perspective on the embodied nature of affectual responses developed in chapter 3 and the sketch offered here of the ways the activation of the motor cortex shapes cognitive processes, this sharing of attitudes and feelings should be understood to be thoroughly grounded in bodily experience: a given sonic analog might correlate directly with a sequence of bodily movements or it might simply recollect the general features of such movements, but both point to yet another way the body serves as a means of knowing.

The careful coordination of movement and music in the barbershop scene from Chaplin's The Great Dictator provided a somewhat different view of the relationship between music and gesture, one that emphasized the correlation of the two rather than the features they shared as nonlinguistic modes of communication. Here the syntactic resources of music—including the coordination of multiple syntactic layers and the frameworks provided by meter and tonality—gave rise to a sense of direction and purpose less evident in Chaplin's movements, which (on the analysis I offered) ultimately reshaped how these movements could be understood.

Although there has long been speculation that music, gesture, and language share a common origin, it is perhaps a safer bet to proceed on the assumption that these unique modes of human communication are built on a shared infrastructure and are shaped by diverse and sometimes conflicting social motivations. It seems plain that the study of gesture, as an overtly corporeal expressive medium, has much to tell us about music; it seems equally plain that a study of the sonic analogs for dynamic processes offered by music can tell us much about the role human movement plays in social and cultural practices, something taken up in greater detail in the next chapter.
Foundations of Musical Grammar

Elements of Musical Grammar: Referential Frameworks

The temporal and tonal frameworks I introduced above are both examples of referential frameworks. Such frameworks are syntactic structures (stored in long-term memory) that provide perceptual anchors for sonic analogs. Although I shall continue to focus on temporal and tonal frameworks in the account of musical grammar I develop in the next two chapters, this is simply an expedient adopted in the interests of clarity. As I conceive of them, referential frameworks could be based on various sonic phenomena: although the temporal and tonal aspects of musical organization have been proposed as the primary parameters for musical utterances (Meyer 1973, 44; 1994, 247–48), there is no a priori reason why some other aspect of musical sound could not serve equally well as a perceptual anchor for sonic analogs.

An assumption shared by many music scholars is that referential frameworks of this sort—especially those associated with pitch—have an ontological status that renders them independent from musical practice. Such an assumption has, for instance, guided what might be broadly called scale theory since the late Middle Ages (including accounts of various modal systems) as well as research in music perception and cognition at least since Hermann von Helmholtz ([1885] 1954, chap. 13). As suggested by my account of the parts such frameworks play in Brahms’s Hungarian Dance No. 5 and Jerome Kern’s “The Way You Look Tonight,” I view referential frameworks as enacted through the syntactic processes that organize the basic elements of musical grammar. They are thus an emergent property of, rather than a precondition for, musical practice.
In the autumn of 1819 Martha Wilmot Bradford (1775–1873) set off from England with her husband, three children, and two servants to take up residence in Vienna, where her husband was to assume the post of chaplain to the British embassy. Although William Bradford had previously held a comfortable position at a parish in the small village of Storrington, West Sussex, Martha was more than happy to take on this European adventure, having acquired a wider view of the world when, in the years before her marriage, she served as personal companion to the Russian Princess Ekaterina Romanova Dashkova (herself an intimate of Catherine the Great; Wilmot 1935, xv). The Bradfords were soon swept into the political and social whirl of Viennese diplomatic life. In a letter to her sister Alicia, Martha described a ball held by the Spanish ambassador on December 27 of that same year, attended by the entire imperial family save the emperor himself. Martha noted in particular the sequel to the Empress Caroline's circulation through the crowd:

When she [the empress] had talked to almost everyone she proceeded to the ball room. Then begun the waltzes. There were not many dancers, except the Court, but if they were not glittering Waltzes never did I see any. *Cinderella* in her glass slippers and Fairy Gifted finery was *dull* to the brilliancy of every creature. (Wilmot 1935, 46)

Indeed, social life in Vienna during much of the nineteenth century centered around the ballroom and the waltz. This was true not only for the ruling classes—the grand balls that punctuated the gathering of heads of state at the Congress of Vienna in 1814–15 invariably concluded with waltzing (Alsop 1984, chap. 6)—but also for the emerging bourgeoisie, who packed the twenty or thirty dance halls sprinkled through Vienna’s suburbs (Hanson 1985, 163). The centrality of dancing to Viennese social life was evident in a letter Martha wrote to Alicia in October of 1827, in which she expressed various concerns about a visit to Vienna planned by their nephew Edward Wilmot. Although Martha worried that Edward’s health would suffer the effects of Vienna’s capricious winter weather, her most significant reservations focused on the likelihood that he was ill-equipped to confront the challenges of Vienna’s social life. Martha’s prose, often elegant and witty, begins to break down in the face of the obstacles her nephew would confront:

A young englishman arriving at Vienna, if quite unknown, is driven to theatres etc. in very despair; if known and introduced into different houses, he has the honour to sit in their drawingrooms during quarter of an hour where *nothing* but french
or German is spoken, and where the young ladies seldom appear, and go from one visit to another, all to be over by 8 O’Clock when it is supposed each has his private engagement, and this sort of thing lasts till near the Carnival, when gaiety begins. Then comes on balls thick and threefold, and woe be to him who does not Waltz, for tho’ there is a blaze of beauty and the balls are quite lovely to look at once or twice, yet no girl will speak to a man who does not dance, and (the men say) very little to those who do. They dance uninterruptedly, and O what an exhibition it is. (Wilmot 1935, 304)

In nineteenth-century Vienna, success in society required being able to dance and, in particular, to dance the waltz.

Social dancing was, of course, important throughout Europe, and during the seventeenth and eighteenth centuries it ran the gamut from relatively simple dances of the countryside to the highly refined dances of the court. As the eighteenth century gave way to the nineteenth, however, changes in the economic and social order brought to prominence dances that could be learned quickly and practiced in a variety of settings. Among the dances that became popular, the waltz was distinguished by the simplicity of its step—which, in its most basic form, was a single pattern that was repeated indefinitely—and by its energetic mode of performance, which required that partners hold each other in close embrace as they spun around their common axis. These selfsame features also gave rise to a certain amount of anxiety: the waltz was often characterized by critics as a dance both primitive and (in its wanton intimacy) emblematic of moral decline (Ullrich 1948; Reeser 1949, 16–25; Yaranan 2002, 5–10). At the same time, the disruptions to the social order that had transformed dance culture led to a greater emphasis on the correct performance of socially structured interactions and resulted in a proliferation of books and pamphlets on etiquette during the late eighteenth and early nineteenth centuries. Within this context the waltz—teetering on the edge of the socially acceptable but also making possible a rapid circulation of partners—provided a unique means of social exchange based on close, if temporary, personal contact. As such, the dance could serve both the bourgeoisie (as a means to explore and expand social alliances) and heads of state (as they negotiated international relations transformed by the Napoleonic wars).

The importance of social dance was reflected in the industry with which entrepreneurs built lavish dance halls and staffed them with conductors and performers who could provide music for the dancers who filled them. As an example, the Apollosaal, which opened in the Viennese suburb of Mariahilf in 1808, was an opulent facility that featured a long, narrow dance floor polished to facilitate fast dances such as the waltz, and that employed (among others) the conductor/composers Josef Hummel, Sigmund Sperber, Thomas Storch, and Joseph Lanner (Hanson 1985, 166; Stöger 1897). The success of both halls and musicians depended on the availability of music that could capture the movements of the dance, something that, as Thomas Wilson noted in his 1816 treatise on the waltz, could not be taken for granted. After remarking on the availability of collections of waltzes that provided proper music for various different forms of the waltz, Wilson observed:
There are, however, a great number of tunes, composed in three-eight and three-four, not possessing, in the least degree, any of the style that, in tunes properly adapted in their composition to Waltzing, is capable by good players, of being expressed with a tenderness and feeling that greatly promotes good Waltzing.

The reason is that, Waltzing having become so fashionable, has afforded a stimulus to composers, to employ their talent in the composition of waltzes; and, though very few musical composers have a sufficient knowledge of dancing, to enable them to understand the nature or composition of Waltzing; yet, their productions, if composed in three-eight or three-four, whether adapted to Waltzing or not, are almost invariably without hesitation denominated waltzes. (Wilson 1816, liv–lv)

It was, perhaps, with similar thoughts in mind that Adolf Bernhard Marx, writing some two decades later, began his instructions on how to compose a waltz with a description of the basic movements of the dance. “The waltz has two movements: first each pair of dancers turns itself in a circle around its own center; second the pair progresses with these continuous turns in a greater circumference until it reaches its starting place and the circle is closed. Each little circle is performed in two-times-three steps and is, as it were, the motive of the dance” (Der Walzer hat zweierlei Bewegungen: erstens dreht sich jedes Paar der Tanzenden im Kreise um seinen eigenen Mittelpunkt; zweitens bewegt es sich mit solchen fortgesetzten Wendungen in einer grössern Kreislinie fort, bis es wieder an seinen Ort gelangt und der Kreis geschlossen ist. Jene kleiner Kreiswendung wird in zweimal drei Schritten ausgeführt und ist gleichsam das Motiv des Tänzes”) (Marx 1837–38, 2: 55). The result of these movements is a distinctive spiraling circular path that the dancers trace through successive repetitions of their smaller circles, which Sevin Yaraman illustrated with the diagram reproduced in example 5.1. Marx went on to argue that the basic materials of the music had to conform with the basic materials of the dance: “At the very least the [music for the] waltz must bring into prominence this basic motive of movement. Each measure, or, better, each phrase of two measures, must answer to the dance motive marking the first step firmly, and also the swinging turn of the dance. Where the measures do not point it out they must still favor it, by a melody which spiritedly turns away from the first note” (“Das Geringste, was der Walzer zu leisten hat, ist die Hervorhebung dieses Grundmotivs der Bewegung. Jeder Takt, oder besser: jede Klausel von zwei Takten muss also dem Tanzmotive entsprechen, den Antritt fest bezeichnen und die schwingende Wendung des Tänzes—wo nicht andeuten, doch begünstigen durch eine vom ersten Tone sich schwungvoll abwendende Kantilene”) (Marx 1837–38, 2: 55). The music for the waltz could thus be seen as a sonic analog for the dynamic process of performing the waltz’s revolving embrace.

The importance of the waltz in the social life of early nineteenth-century Vienna conforms well with Pierre Bourdieu’s understanding of the way bodily experience informs the habitus, the “structuring structures” that shape the behavior of individuals in social contexts (1990, 53). Bourdieu conceived of bodily experience as involving affective responses produced when the body confronts its environment; this experience gave rise to bodily dispositions that shaped the
habitus. As Bourdieu noted, “We learn bodily. The social order inscribes itself in bodies through this permanent confrontation, which may be more or less dramatic but is always largely marked by affectivity and, more precisely, by affective transactions with the environment” (2000, 141; on the bodily aspect of the habitus, see also Noland 2009, 7). Viewed from this perspective, the practice of waltzing—of spinning one’s body across the ballroom floor—was both a manifestation of the social life of Vienna and part of the way the social life of Vienna was constructed. Given the close relationship between the steps and the music of the waltz, waltz music could be understood as an equal participant in the construction of the social: listeners heard a waltz as an invitation to the dance and as a guide to an elaborate form of social interaction.

The practice of waltzing in early nineteenth-century Vienna thus provides us with a glimpse of musical construction grammar in action: the form of the musical utterance basic to the dance is what Marx aimed to capture with his notion of the waltz motive; the function of this utterance was to provide the dancers with a template onto which they could map their movements and thereby enter into—and indeed realize—the social exchanges of the dance floor. In this chapter I use the specific case of social dance to exploit and expand the view of musical grammar developed over the course of the preceding chapters—indeed, the relationships that obtain between the music for social dances and the steps the dancers make provide an ideal laboratory within which to probe the foundations of musical grammar. One of the things that will quickly become evident is the role music’s temporal and tonal frameworks (described in chapter 4) play in social dance. Temporal frameworks, with their interrelated rhythmic strands and

*Example 5.1* Sevin Yaraman’s diagram showing the path that dancers make while waltzing (2002)
schematic potential, provide a rich array of musical events that dancers can use to coordinate and plan their movements. Tonal frameworks—which facilitate the creation of analogs for dynamic processes like departure, return, and arrival—make it possible for a composer to construct a sonic topography that is in dialog with the spatial topography activated by the dancers’ movements (a topography fully realized through the four-dimensional patterns that the dancers create on the dance floor).

In what follows I shall focus on two rather different dance practices. The first is associated with France’s ancien régime. Thanks in part to Louis XIV’s interest in dance, French choreographers developed a number of different dance notations during the late seventeenth century that make it possible to study closely the way the steps and music of a dance were correlated with one another. In the first section that follows, I use an example of dance notation published in 1700 to treat in some detail the way the movements of the dancers correlated with the music and to consider what the relationship between the two can tell us about how dance practice shaped the habitus of the social milieu of the ancien régime. In the second section, I return to the waltz as it was practiced in early nineteenth-century Austria and Germany, and to waltzes by Joseph Lanner and Franz Schubert. As might be expected of a composer associated with the Apollosaal, Lanner’s waltzes are clearly intended for the ballroom, but Schubert’s are somewhat more varied and suggest a number of different contexts for performance. Attending closely to the syntax of these waltzes as revealed through the sonic analogs Schubert employs, I explore how each might have helped construct a different social context, including one in which—following a suggestion by Robert Schumann—there was no dancing at all.

**Music and Dance in the Ancien Régime**

*An Introduction to Dance Notation and French Noble Dance*

In 1700, having been granted exclusive rights of publication by royal privilege the preceding year, Raoul-Augur Feuillet published his *Chorégraphie, ou l’art de décrire la dance* (Feuillet [1700] 1968a). Feuillet’s “art of describing dance” was one that involved an exhaustive system of figures and diagrams that made it possible to represent both the steps of a dance and the paths the dancers took as they moved across the floor. Although Feuillet’s notation was flexible enough to accommodate as many as nine dancers (Hilton 1997, 94–97, 430–31; Feuillet [1700] 1968b, 67–84), most of the choreographies he published in the years that followed were for one or two dancers. The importance of *danses à deux* was especially apparent in the collection of nine ballroom dances by Louis Guillaume Pécor that Feuillet published in 1700, each of which was for a man and a woman (Feuillet [1700] 1968b).

The first page of Feuillet’s notation for Pécor’s choreography *La bourrée d’Achille* is shown in example 5.2. At the top of the page are the first eight
measures of a bourrée to which the opening of the choreography is set. (*La bourrée d’Achille* takes its music from the “Entrée des Genies de Talie” in the prologue to Jean-Baptiste Lully’s *Achille et Polixène* [Lully (1687) 2007]. This part of the entrée consists of a bourrée followed by a minuet, which Pécout arranged into a short suite by following the minuet with a reprise of the bourrée; all told, the choreography published by Feuillet runs to eleven pages. It bears mention that Lully completed only the overture and Act I of *Achille et Polixène*; the Prologue and Acts 2–5 were written by Pascal Collasse.) The space below the musical notation represents a bird’s-eye view of the performance space for the dance. The portion of the performance space corresponding to the top of the page was reserved for what was known as the Presence, which would be occupied by the person, or persons, of highest rank (Hilton 1997, 85). At the bottom of the page are two symbols that represent the starting positions for the two dancers: the man on the left (a semicircle joined to a straight line), the woman on the right (two nested semicircles joined to a single straight line); showing due deference, they begin

Example 5.2 The first page of Raoul-Auger Feuillet’s notation for Louis Guillaume Pécout’s choreography *La bourrée d’Achille* ([1700] Feuillet 1968b, 2)
by facing toward the Presence. Immediately above the symbols for the dancers are two small circles to which short lines are connected; these indicate that the dancers start with their feet in fourth position. Two solid lines extend toward the top of the page; these indicate the track that the dancers will take across the floor. Midway up the page the track shows each dancer making a right-angle turn; the man to his left, the woman to her right, such that they briefly proceed in opposite directions. Each dancer then makes another turn (the man to his right, the woman to her left) and then follows an arc toward the top of the room and each other. The dancers then make one more sharp turn (the man to his right, the woman to her left) and, once again side by side, move toward the center of the room. The track itself is crossed by a number of very short perpendicular lines, each of which corresponds to a bar line of the music shown at the top of the page; the first portion of the track (when the dancers move directly toward the Presence) corresponds to two measures of music, their movement in opposite directions from each other after their first turn corresponds to one measure, and so on for the eight measures of the bourrée given on this page. Symbols for the steps each dancer performs are drawn to either side of (and occasionally across) the track; those on the left side of the track are for the left foot, and those on the right side of the track are for the right foot.

In her work on French noble dance (a large portion of which was given over to the interpretation of Feuillet’s notation) Wendy Hilton offered helpful distinctions between the step (a passage of the foot forward, backward, or to the side), a step-unit (made up of a number of steps), and a step-sequence (a sequence of step-units that reaches a point of conclusion; Hilton 1997, 73). In the opening of the choreography shown in example 5.2, the male dancer begins with three steps (left-right-left), which together make up the step-unit known as the pas de bourrée. (The female dancer also does a pas de bourrée, but starting on the right foot, thus creating a mirror image of the movements of the male dancer.) As do many of the step-units within the French noble style, the pas de bourrée begins with a mouvement—bending the knees down and then up again (a plié followed by an élevé)—that leads directly to a step, a combination known as a demi-coupé (Hilton 1997, 75).

One of the challenges faced by those notating choreographies was the precise coordination of music and movement: in some cases a number of steps were correlated with a single musical event (such as a held note); in other cases a single step was correlated with a number of musical events (such as the sequence of notes that make up a short scalar passage). A further complication was that most step-units could be used interchangeably in duple and triple time: the pas de bourrée, for instance, could be employed equally well to dance a bourrée (in duple time) or a sarabande (in triple time; Hilton 1997, 186). As a means of clarifying the relationship between music and movement, Feuillet’s Chorégraphie included a brief section that described in slightly more detail how various aspects of musical rhythm corresponded with different dance movements (Feuillet [1700] 1968a, 87–92); most subsequent dance tutors included similar sections, often given the title “Treatise on Cadence.” In the “Traité de la cadence” appended to his 1725 Abbregé de la nouvelle méthode dans l’art d’écrire ou de tracer toutes sortes de danses de ville,
Pierre Rameau made one aspect of the relationship between movement and music particularly clear: in all cases the plié of the initial demi-coupé in a step-unit—resulting in a downward movement—occurred on the upbeat to the measure, and the élevé—resulting in an upward movement—on the downbeat (Rameau [1725] 1972, 103–11). The élevé of the demi-coupé was thus associated with an accent, a feature that could, in some instances, be used to create cross rhythms between movement and music (as occurred with the second demi-coupé of the typical pas de menuet step-unit; Hilton 1997, 191).

In general, the step-unit was a point of focus for much of the theory and practice of French noble dance. In its role as a means of organizing how dancers thought about the varied movements they needed to make, the step-unit is analogous to a basic-level category, representing a compromise between the details of individual steps and the larger successions of movements proper to step-sequences. (The role of basic-level categories in human cognition and the relationship of such constructs to the elements of musical grammar were discussed in chapter 2.) And, as do the categories humans use to structure their understanding of the world, the conception of step-units demonstrates typicality effects: in his Chorégraphie, for instance, Feuillet gathers together all of the variants of the pas de bourrée into a table spanning eight pages and including ninety-four variants, which are grouped into various subtypes and which begins with the most basic version of the step-unit and concludes with rather more complex versions (Feuillet [1700] 1968a, 63–70). The importance of the step-unit was similarly reflected in the music—indeed, one of the principal functions of the music for French noble dance was to provide a sonic template into which the step-units of a choreography could be fit.

**Pécour’s Choreography and Collasse’s Music for La bourrée d’Achille**

As with many of Pécour’s choreographies for danses à deux, that for the initial bourrée from *La bourrée d’Achille* involves a relatively small number of step-units (Witherell 1983, 7–15). In addition to the pas de bourrée, these include the contretemps de chaconne (which, after the initial plié, involves a spring from one foot onto the same foot, followed by two plain steps; Hilton 1997, 214), the temps de courante (a mouvement that leads to a single sliding step; Hilton 1997, 201), the coupé soutenue (a demi-coupé followed by a sliding step; Hilton 1997, 177), the coupé simple (a demi-coupé followed by a single plain step; Hilton 1997, 177), and the pas de sissonne (a hop, or jeté, that finishes on two feet, followed by another hop that finishes on one foot; Hilton 1997, 223). Table 5.1 provides a list of the various step-units used by the male dancer in Pécour’s choreography for the initial statement of the bourrée from *La bourrée d’Achille*, together with the measure that corresponds with each step-unit.

A more complete account of Pécour’s choreography for *La bourrée d’Achille* would have to include a much more thorough description of the various steps and step-units as well as a fuller account of the movement of the dancers’ arms, the small differences between the step-units used by the male dancer and those used by the female dancer, and the paths the two trace across the floor (including
movements forward and backward) as they perform the dance. For the purpose of understanding the relationship between movement and music in the French noble style, however, it will be sufficient to limit consideration to the most distinctive aspects of the step-units given in table 5.1 and the musical events with which they were correlated. To this end, I would like to introduce an analytical notation developed by Meredith Little for the purposes of studying relationships between music and movement in French noble dance (Little 1975). Little used five symbols, shown in example 5.3, for the main features of different step-units: these include symbols for the plié and élevé of the mouvement, as well as for the jeté, the plain step, and the glissé (the sliding step used in the temps de courante and coupé soutenue). Example 5.4 provides the score for the whole of the bourrée in a keyboard reduction of the harmonization used in the prologue to *Achille et Polixène*, together with the appropriate symbols for the main features of the various step-units. Brackets beneath these symbols show each of the step-units, and the letters above the symbols show whether the right foot (R), left foot (L), or both together (T) are being used.

![Table 5.1 Step-units used by the male dancer and corresponding measures for Pécour’s choreography for the initial statement of the bourrée from *La Bourrée d’Achille*](https://example.com/table5.1)

<table>
<thead>
<tr>
<th>Step-unit</th>
<th>Measure</th>
<th>Step-unit</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>pas de bourrée</td>
<td>1</td>
<td>pas de bourrée</td>
<td>5</td>
</tr>
<tr>
<td>pas de bourrée</td>
<td>2</td>
<td>pas de bourrée</td>
<td>6</td>
</tr>
<tr>
<td>contretemps de chaconne</td>
<td>3</td>
<td>contretemps de chaconne</td>
<td>7</td>
</tr>
<tr>
<td>temps de courante</td>
<td>4</td>
<td>coupé soutenue</td>
<td>8</td>
</tr>
<tr>
<td>pas de bourrée</td>
<td>9</td>
<td>pas de bourrée</td>
<td>17</td>
</tr>
<tr>
<td>pas de bourrée</td>
<td>10</td>
<td>pas de bourrée</td>
<td>18</td>
</tr>
<tr>
<td>pas de bourrée</td>
<td>11</td>
<td>pas de bourrée</td>
<td>19</td>
</tr>
<tr>
<td>pas de bourrée</td>
<td>12</td>
<td>pas de bourrée</td>
<td>20</td>
</tr>
<tr>
<td>pas de sissonne</td>
<td>13</td>
<td>pas de sissonne</td>
<td>21</td>
</tr>
<tr>
<td>pas de sissonne</td>
<td>14</td>
<td>pas de sissonne</td>
<td>22</td>
</tr>
<tr>
<td>pas de bourrée</td>
<td>15</td>
<td>pas de bourrée</td>
<td>23</td>
</tr>
<tr>
<td>coupé soutenue</td>
<td>16</td>
<td>coupé simple</td>
<td>24</td>
</tr>
</tbody>
</table>

*Example 5.3*  Meredith Little’s analytical symbols for the main features of step-units

\[
\begin{align*}
\checkmark & \text{ plié} \\
\wedge & \text{ élevé} \\
\curve & \text{ jeté} \\
\mid & \text{ plain step} \\
\ && \text{ glissé}
\end{align*}
\]
Example 5.4  Keyboard reduction of the bourrée from the “Entrée des Genies de Talie,” from the prologue to Jean-Baptiste Lully’s *Achille et Polixène* by Pascal Collasse, with analytical symbols for Pécour’s choreography
Although Pécour’s choreography was designed to fit with Collasse’s music, I should note that one of the distinctive features of the French noble dance was the close coordination of music and dance in both conception and practice: those known chiefly as composers were intimately familiar with choreographic conventions (and were often, as was Lully, dancers themselves), and those known chiefly as choreographers were also performers and composers of dance music (Beaussant 1992, chap. 2; La Gorce 2002, chap. 3; Hilton 1997, chap. 2). As a result, it is necessary to think of the grammar of French noble music and dance—that is, how they were organized for expressive purposes—as coconstituted. Although choreographies were designed to fit the music, the music was designed to fit with the step-units that shaped the dancers’ understanding of movement sequences.

Both Collasse’s music and Pécour’s choreography for the bourrée fall into two main parts: Section 1 comprises measures 1–8, and Section 2 comprises measures 9–24. Each of these is divided by a reprise: in Section 1, measures 5–8 restate the material of measures 1–4 (with the material for the last measure changed as a way to mark the end of the section); and in Section 2 measures 17–24 restate the material of measures 9–16 (with the last measure changed to signal the conclusion of the dance). In my analysis I shall consider the music and choreography for each section in turn, but I begin with a discussion of the temporal and tonal frameworks set up by the music in the opening measures of the dance. (For a slightly different account of relationships between music and dance in the first bourrée of La bourrée d’Achille, see Zbikowski 2014.)

**Section 1 of the Bourrée.** Although the step-units used in Pécour’s choreography span durations of two half notes in Collasse’s notation, the basic temporal framework established by the music over the course of measures 1–8 spans four half notes (or two measures). The principal anchors in this framework occur on the first and second beats of the notated measures (keeping in mind that there are two main beats in each measure in this time signature): the first beat is marked by a change of harmony, and the second beat is invariably preceded by an eighth note in the bass. These beats are more or less regularly divided by quarter notes and grouped by the consistency of the harmony—measure 1 is focused on the dominant, measure 2 on the tonic, measure 3 on the dominant, and so on throughout the section. The result is a temporal framework that allows musicians, listeners, and dancers to anticipate coming events and, especially so for the dancers, to plan their motor movements accordingly. Two aspects of this framework deserve further comment. First, as is typical of the bourrée, the framework begins with an upbeat, imparting to the dance a sense of liveliness for which it was known (Little and Jenne 2001, 37). Second, the propulsive effect of moving from the upbeat to the downbeat (correlated with the dancers rising from their plié, and a change of harmony) is replicated over the two measures of the basic framework by the relationship between dominant and tonic: the dominant is generally rendered less stable than the tonic, either through a bass line that leads away from the root of the chord (measures 1 and 5) or with an auxiliary harmony on the second quarter of the measure (as in measures 3 and 7). As a whole, then, the musical materials that establish the basic temporal framework for the dance create a sonic analog for
a dynamic process that moves steadily forward in regular increments. Aspects of these same materials contribute to establishing the tonal framework for the dance: the sense of stability associated with the tonic (and confirmed through agogic accents in measures 4 and 8) is reinforced by the consistent and careful diatonicism of measures 1–8 (only the pitches of a D major scale are used, and the sixth step of the scale [B] is added only in measure 3) and through a melody focused on the principal notes of the scale.

Over the course of the first section of the bourrée, the temporal framework provides a means to anchor sonic analogs that animate the rhythmic syntactic layer. Although the temporal framework is articulated through regular successions of quarter notes (dividing the main beat), the regularity of these successions is disrupted in a number of places: the first halves of measures 1, 2, 4, 5, and 6 all include a dotted-quarter note leading to an eighth note or two sixteenths; measures 3 and 7 include eighth notes on the second quarter of the measure, and a dotted-quarter/eighth-figure in the last half of the measure; and measures 4 and 8 begin with a dotted half note in the upper parts (which, in the case of measure 8, coincides with a dotted half note in the bass). These disruptions to the flow of quarter notes inform the rhythmic design of Section 1 in two ways: first, most measures begin with an agogic accent; second, the two measures that do not begin with an agogic accent (namely, measures 3 and 7) have an agogic accent on the second main beat of the measure, which is preceded by an auxiliary harmony embellished with eighth-note motion. This rhythmic design together with the harmonic strategies deployed by Collasse produce an effective analog for a series of regular movements (measures 1–2 and 5–6) that are disrupted (measures 3 and 7) and then pause (equivocally in measure 4, more definitively in measure 8).

As I have described it, the dynamic process represented by the music of Section 1 conforms, in most respects, with the dynamic process enacted by Pécour’s choreography. Pécour begins with two pas de bourrée (which set up something close to a series of regular movements), followed by a contretemps de chaconne (which introduces a jeté and disrupts the alternation of L–R–L and R–L–R), leading to a temps de courante that momentarily halts the progression of steps. This pattern is then repeated almost exactly, save for the introduction of a coupé soutenue—a step-unit that typically concludes a step-sequence—in measure 8. One aspect of the dance not as clearly represented by the music is the regular succession of mouvements, which provide a physical correlate for the arrival on each strong beat. Although these arrivals are also marked by a change in harmony, the music does not provide a strong analog for the physical tension created on each upbeat as the weight of the dancer’s body is taken on flexed knees.

Section 2 of the Bourrée. In comparison with Section 1 the harmonies of Section 2 are rather more varied and participate in a compositional design that unfolds over a longer span of musical time. Collasse opens the section with the submediant in 5 3 position. This gives way to a succession of 6 3 harmonies (beginning on the fourth quarter of measure 9 and extending through the first half of measure 11, mediated only by the 5 3 chord on the last quarter of measure 10) that leads, in the latter half of measure 11, to the dominant of A major, embellished by a 4–3 suspension. (All of the features of measures 9–16 are, of
course, reprised in measures 17–24.) The subsequent arrival on A major leads, almost immediately, to a return to the tonic. Collasse then repeats the melody of measures 9 through 12 but provides it with a new harmonization: measure 13 (which corresponds to measure 9) begins with the tonic in $\frac{5}{3}$ position, and measure 15 (which corresponds to measure 11) is focused on the dominant of D major. Perhaps most important, however, are two details of this reharmonization. First, although the opening of measure 14 is identical to that of measure 10, Collasse harmonizes the last quarter of measure 13 with the leading-tone chord of D major, thus connecting the tonic in $\frac{5}{3}$ position on the third quarter of measure 13 with the tonic in $\frac{6}{3}$ position on the first quarter of measure 14. Second, beginning with the fourth quarter of measure 14 the harmonies set out an orderly circle of fifths progression (F# minor, B minor, E minor, A major), which culminates in an arrival on tonic in measure 16.

In contrast to the comparatively subtle play of rhythm in Section 1, the rhythmic strategies used by Collasse in Section 2 are both simpler and of more marked effect. The half notes on the second quarter of measures 10, 12, and 14 create an agogic accent at a moment in the rhythmic cycle that has not, up to this point, received any emphasis at the same time that they preclude the possibility of marking a moment—the third quarter of the measure—that has received emphasis. The contrast provided by these measures is mitigated in two ways: first, measures 9, 11, and 13 each make use of rhythmic patterns established in measures 1 through 8; and, second, measures 15 and 16 revisit the basic materials of the rhythmic cadences of measures 3–4 and 7–8. As a consequence, the displaced accents introduced in Section 2 function as interruptions of the normative rhythmic patterns established in Section 1, rather than as a permanent shift of the pattern of metrical accents.

Within Section 2 Collasse uses the resources offered by harmony and rhythm to craft a succession of musical events that suggest a dynamic process that begins with the upbeat to measure 9 and continues through the downbeat of measure 16 (and that is revisited in measures 17 through 24). The midpoint of this temporal span is articulated by the arrival on A major in measure 12—an arrival prepared by giving over the last half of measure 11 to the dominant of A, complete with a 4–3 suspension in the alto voice and an octave skip in the bass—which is rendered contingent by the displaced accent on the second beat of the measure. (Compare, for instance, the effect of measures 11–12 with that of measures 3–4.) The fluidity—even uncertainty—associated with successions of $\frac{6}{5}$ harmonies and displaced accents only starts to resolve with the circle-of-fifths sequence and more regular rhythmic structure that commences on the last quarter of measure 14 and that reaches its conclusion with the arrival on the tonic in measure 16. As a whole, the passage provides an analog for a dynamic process that moves away from previously established safe ground, ventures over novel topography, and then slips almost effortlessly back to the place where it began.

In contrast to the clear sense of departure occasioned by the introduction of the submediant at the opening of Section 2 (and affirmed by the displaced accents of measures 10 and 12), the first portion of Pécour’s choreography begins with four successive pas de bourrée—that is, an unbroken succession of the same step-unit
that served to anchor the choreography of Section 1. With regularity of movement thus having been established, Pécour then introduces the most animated step-units thus far: two successive pas de sissonne, which comprises four successive hops. These are then followed by another pas de bourrée, and the step-sequence concludes with either a coupé soutenu (in measure 16) or a coupé simple (in measure 24). The overall pattern is one of regular movement interrupted by a brief flurry of activity and concluding with a smooth approach to closure. The dynamic process enacted by the dance steps of Section 2 is, in consequence, markedly different from that suggested by the music. Emblematic of this difference is the mismatch between movement and music on the third quarters of measures 10, 12, and 14 (as well as measures 18, 20, and 22), when the dancer’s step (or hop) onto his left foot is met with no corresponding musical event.

It bears mention that while Collasse’s music and Pécour’s dance steps both reach a conclusion in measure 16 (and again in measure 24) they do so through different means. The music initiates in a rather precipitate manner the sequence of events that culminate in measure 16; once on its way, however, the music proceeds as though its goal is never in doubt, each two-measure unit leading securely to the next. The dance steps, by contrast, continue with their pas de bourrée as though oblivious to the adventurous course of the music but then hurry to catch up, becoming slightly unpredictable at the very moment that the music catches sight of the course it will take to cadence. Here, then, we see a divergence between music and dance not countenanced by Marx’s conception of the relationship between the steps of a dance and its music but that nonetheless yields an effective, and expressive, whole.

Music and Dance in La bourrée d’Achille. In summary, then, Pécour’s choreography for the first statement of the bourrée in La bourrée d’Achille demonstrates two slightly different strategies for the correlation of music and dance. In Section 1, music and dance fit closely with one another: the dynamic process summoned by Collasse’s music finds a visible and physical representation in the movements of Pécour’s dancers. In Section 2, music and dance are more independent of one another, although they ultimately arrive at similar goals. In practice, then, there are a range of possibilities for correlating music and dance: at times the music and steps of a dance will be in very close correspondence with one another, but the two may also diverge for expressive or strategic reasons. This situation is further complicated by two different configurations of the analogical relationships between music and dance.

In the first configuration, music and dance relate directly to one another: the dynamic process constituted by performing the steps of the dance is given a sounding representation by the music for the dance (a perspective on music and dance embodied in Marx’s conception of the waltz), or the dynamic process specific to a succession of musical materials is given a visible and physical representation through dance (which is one way to view Pécour’s choreography of Collasse’s bourrée). In the second configuration, music and dance each provide an analogical representation of a dynamic process drawn from some other domain. For instance, the physical movements that take a dancer across the floor could serve as an analog for a journey. As it happens, this possibility is exploited by Pécour’s
choreography, which leaves the dancers quite distant from one another at both of the main cadence points in Section 2, thereby offering an analogical representation of journeys that have diverged. Indeed, it is only at the concluding cadence of the reprise of the bourrée—that is, in the very last pass through the material of Section 2 at the end of *La bourrée d’Achille*—that the choreography once again brings the dancers together to dance side by side: the intricate traceries of the dancers’ journeys, which over the course of the bourrée, minuet, and bourrée of *La bourrée d’Achille* have repeatedly drawn them together and then sent them apart, finally allow them to conclude as one. Collasse’s music could similarly be understood to provide a representation of a journey, one that begins with the upbeat to measure 1, proceeds in distinctive ways from one cadence point to the next (more orderly in Section 1, more adventuresome in Section 2), and ends with the arrival on tonic in measure 24. It should be noted, however, that the journeys depicted by music and dance are far from identical: Collasse’s music provides the same sort of cadence whether the dancers conclude far apart (as at the end of the first statement of the bourrée) or in close synchrony (as at the end of the suite of dances); Pécour’s dancers mark cadences not by a return to a specific point in the performance space (analogous to the harmonies that are in the same “place” in the virtual topography of Collasse’s music) but with a step-unit that interrupts the steady flow of movement that typifies the preceding steps.

These two configurations reflect different kinds of analogical mappings. The first configuration involves structural mappings between two correlated domains (here, “music” and “dance”), which are supported by correspondences between the elements and relations specific to each domain. For instance, structural features of a cadence within the domain of music (which, in measures 3–4 of *La bourrée d’Achille*, includes a disruption of a previously established rhythmic pattern, the introduction of a contrapuntal configuration not used elsewhere, and a pause) can be correlated with a cadence in the domain of dance (including a disruption of a previously established pattern of step-units, the introduction of a step-unit not used elsewhere, and a momentary cessation of movement; for a discussion of cadential formulae used in both music and dance in the French noble style, see Pharo 1997). The second configuration of analogical mappings, by contrast, involves an additional domain (in the case of the example I offered, that of “journeys”) and, as such, the integration of concepts typical of a conceptual blend. As noted in chapter 2, conceptual integration offers further resources for the construction of meaning but of necessity renders more abstract the structural relationships that typify an analogical mapping.

With respect to the musical grammar of Collasse’s bourrée, it is, of course, conditioned by its function: music and dance are a coordinated whole, building from consistent relationships between the succession of musical events that make up a measure and the succession of movements that make up a step-unit. While the patterns marked out by the music may diverge from those marked out by the dance, they are anchored by cadence points at the ends of musical phrases and step-sequences. The phrase or the step-sequence thus constitutes a larger syntactic whole, distinguished here by the general strategy of establishing a pattern for a sequence of events, diverging from that pattern, and then bringing the sequence to a close.
I should emphasize here that these observations about the grammar of the music for French noble dance are preliminary—indeed, the relationships between music and dance could be considerably more complicated and more subtle than those demonstrated by a ballroom dance such as La bourrée d’Achille. There are, for instance, numerous examples of dance phrasing that runs contrary to musical phrasing and strong evidence that composers and choreographers played with this variability in their compositions (Schwartz 1998; Harris-Warrick 2000). One of the advantages of Feuillet’s notation is that it makes it possible to discover such complications, for it was expressly designed to show how dance steps and music related to one another. As do all notations (whether for music or dance), it has its limitations. Since the music is only shown in a single stave, details of harmony or counterpoint have to be inferred—it is quite possible that the version of Collasse’s music envisioned by Pécour was somewhat different from that used in the published prologue to Achille et Polixène (and, indeed, there are minor differences between Collasse’s melody and that which accompanies Pécour’s choreography). Understanding how the steps of the dance were to be performed requires an equal measure of interpretation, although thanks to the work of Hilton and others there is now far less guesswork than there once was. With that in mind, I would again emphasize that my analytical account of Pécour’s choreography is radically simplified (having for the most part omitted mention of movements through the performance space, having nothing to say about steps performed backward or with turns, and having ignored the movements of hands and arms); as my brief comments on the conclusion of La bourrée d’Achille might suggest, there is also much to be learned from Pécour’s choreography for the minuet and for the reprise of the bourrée, both of which give further insight into how Pécour responded to Collasse’s music. But even with these limitations in mind, the notation and practice of French noble dance provide valuable insights into ways music correlates with other communicative media, into the way dynamic processes can be represented analogically, and the means through which patterned sequences of events can be organized into larger syntactic wholes.

The Culture of Dance in the Ancien Régime

The centrality of dance to the court culture of late seventeenth- and early eighteenth-century France was a consequence of a number of factors: the young Louis XIV’s gift for dancing; the deployment of this gift in entertainments for the court organized by Jules Mazarin, prime minister during the regency of Louis’s mother, Anne of Austria (which began in 1643); and Louis’s determination, once he assumed absolute power, to exclude other members of the nobility from active roles in the governance of the country and to keep them occupied with courtly entertainments, participation in which required learning complex choreographic routines (Hilton 1997, 7–9). The study of dance thus had political as well as practical dimensions, both of which are evident in the decree through which Louis XIV established the Académie Royale de Danse in 1661, the same year he became the absolute ruler of France:
The Art of the Dance has always been recognized as one of the most honorable and necessary for forming the body, and giving it the first and most natural dispositions for all sorts of exercises, and among others for the exercise of arms, and consequently has been considered one of the most advantageous and useful for our Nobility, and for others who have the honor of approaching us, not only in wartime in our armies, but even in peacetime in our Ballets. (Franko 1993, 176)

Under Louis’s rule, then, dance took on unparalleled importance within French culture (something discussed further in Franko 1993, chap. 5; and McClary 2012, chap. 8). On the one hand, the noble classes came to view dance as a means of refining the physical disposition through which they distinguished themselves from those whom they ruled (Rameau [1725] 1970, xii). On the other hand, dance provided an opportunity for the nobility to give corporeal presence to the aesthetic of galanterie (Cowart 2008, 14). Although galanterie was hardly a stable construct during the time of the ancien régime, to the extent that it was construed as a desirable attribute it was an intensely social one, the essence of which was to gain distinction by bringing pleasure to others (Viala 1997). Accordingly, the ballroom settings that Feuillet and other dance masters had in mind were those in which the dancers performed before an audience of their peers, beginning with the couple of highest rank and proceeding through the rest of those assembled (Rameau [1725] 1970, 37–39). Within the ancien régime, dance was thus central to the development of a comportment that pleased, the display of which was crucial to securing and maintaining a place within the social order of Louis’s court.

Given this, it is perhaps not surprising that one of the results of the establishment of the Académie Royale de Danse was the development of dance notations through which choreographies could be preserved and taught to others. By the 1680s there were as many as four kinds of dance notation in circulation (Harris-Warrick and Marsh 1994, 83–87; Pierce 1998), although that of Pierre Beauchamps—at that time composer of the royal ballets and private dance tutor for the king—was by far the most successful. One reason for the success may have been the science of bodily movement intimated by Beauchamps’s notation; it was, indeed, Beauchamps who received credit for codifying the five canonical positions used in French noble dance, which were later adopted for ballet (Rameau [1725] 1970, 5; Hilton 1997, 98–99). And it was as a kind of science that Feuillet’s Chorégraphie proceeded, with attention given to each aspect of position and movement captured by Beauchamps’s notation (without, however, acknowledging Beauchamps as its inventor). By this means the influence of French noble dance spread further: the guide provided by Chorégraphie, together with the annual collections of ballroom dances Feuillet published (such as the one that included Pécour’s La bourrée d’Achille), made possible the dissemination of the French dance culture far beyond Louis XIV’s court. (Two English translations of Chorégraphie appeared in 1706 subsequent to the expiration of the royal privilege granted Feuillet. It was in fact through these translations that Beauchamps’s role in the development of dance notation became more widely known, for each acknowledged him as the inventor of the notation that had been so successfully exploited by Feuillet [Hilton 1997, 46–49].)
It was with some justification, then, that Pierre Rameau could, in his *Maitre à Danser* of 1725, claim a preeminent place for the culture of French noble dance: “We can say, to the glory of our race, that we have a real gift for beautiful dancing. Foreigners, so far from denying this, have, for nearly a century, come to admire our dances, and to educate themselves at our performances and in our schools. There is hardly a Court in Europe where the dancing master is not French” (Rameau [1725] 1970, xii). There is little doubt that this preeminence was in part a reflection of France’s political power during the seventeenth and eighteenth centuries, but it also bore testament to the influence of an expressive culture in which dance and music were carefully adapted to one another such that an elegance and appropriateness in movement had its correlate in an elegance and appropriateness of sonic materials. A sense of the pleasure afforded by such an adaptation is readily apparent in Kellom Tomlinson’s account of the effect of a close coordination of music and dance:

> Every Turn, Step, Spring or Bound seen in one will be at the same Instant observed in the other, in such an exact Symmetry and Harmony of the Parts agreeing with the Notes of the Music, as to cause the most agreeable Surprize in the Beholders of the two Dancers; or admitting a Dozen or more in Number, by observing them all to move as only one Person. This is the natural Effect of good Dancing adorned with all its Beauties, in that the Music seems to inspire the Dancing, and the latter the former; and the Concurrence of both is so requisite to charm those who behold them, that each of them in some Measure suffers by a Separation. (K. Tomlinson [1735] 1970, 150)

The magic wrought by this concurrence of dance and music speaks to the power of the enactment of galanterie: one who can dance well can, through pleasure given, hold his or her audience in thrall.

As noted in chapter 4, the patterned sound of music makes it possible, through sonic analogs, to share attitudes and feelings. As made clear by the practices associated with dance in the ancien régime, the patterned movements of dance also make possible the sharing of attitudes and feelings. This is most obvious in the case of dancers whose conjoined movements enact a dance, but, as suggested by Tomlinson’s remarks on the effect of music and dance properly coordinated, is equally the case for those observing the dance. Indeed, one could regard spectatorship as one of the design principles of French noble dance; this was dance that was meant to be watched by an audience with knowledge about and appreciation for every detail of the steps involved. As Wendy Hilton has remarked, “Choreographically, even the most virtuosic steps were not stressed, but merely slipped into the rhythmic flow of the dance. The audience, most of whom danced themselves, would have been kinesthetically aware of the difficulties they were watching and able to appreciate apparent ease in performance” (1997, 37–38). Empirical support for the kind of kinesthetic awareness proposed by Hilton has, of course, been provided by recent research on the motor areas of the brain (as discussed in chapter 4) and, more specifically, through brain-imaging studies of expert dancers (Calvo-Merino et al. 2005). It seems likely,
then, that the social practice of French noble dance, and in particular the corporeal expression of galanterie, was predicated on shared bodily knowledge that was facilitated by the way the dance was performed. It must also be emphasized, however, that full appreciation of the spectacle of dance and music would have been granted only to those whose habitus had been shaped by the incorporation of the choreographies of French noble dance—that is, the select group of courtiers, dancers, and musicians with which Louis XIV had surrounded himself. The dance practices of the ancien régime created cohesion within this social group and excluded all others.

Another of the design principles of French noble dance—and one quite evident in Pécour’s *La bourrée d’Achille*—was an intimate relationship between music and dance. As my analysis of the first bourrée showed, the interplay between music and dance was one as lively as it was subtle and made possible the enthralling effect described by Tomlinson. This is due in no small part to the sonic analog for movement created by Collasse’s music, one so convincing that listeners may need to be reminded that the impression of movement created by the music is an illusion rather than a fact.

The effect achieved by dances such as Pécour’s *La bourrée d’Achille* was, as I have emphasized here, no accident but instead a result of a careful arrangement of musical materials and choreography, and a thorough training of dancers and musicians alike. As I noted in the opening to this chapter, however, dances that could be learned quickly and practiced in a variety of settings became much more important over the course of the eighteenth century. The result was a change not only in dance practice, but also in the organization of the music to which dancers coordinated their movements.

**Of Waltzes and Waltz Music**

**Waltzes and the Ballroom**

Martha Bradford’s letters to her sister offer a glimpse of visitors’ fascination with the waltz and of the importance of dancing within Viennese social life, but for a full sense of the intoxicating effect of the waltz, we turn to the French diplomat Auguste Louis Charles La Garde-Chambonas, who described in some detail the splendor of the balls that were an integral part of the Congress of Vienna. After a brief sketch of an opening pantomime performed by children for the benefit of the gathered heads of state, La Garde-Chambonas turned to the dancing proper:

Once the sovereigns had left, the orchestras began to perform waltzes. An electric shock seemed immediately to be transmitted to the crowd. Germany is the homeland of the waltz; it is in this country and above all in Vienna that, on account of the musical ear of its citizens, it has acquired all of its particular charm; this is where you have to [be to] see, in this whirling race that is always controlled by the beat, the gentleman support and carry away his partner, and to see her yield
to this gentle being-swept-away, while a kind of dizziness gives her gaze a vague expression that enhances her beauty. Thus, is it difficult to imagine the irresistible influence exercised by the waltz? As soon as the first measures are heard, faces light up, eyes come alive, and a quiver [of pleasure] is communicated from each person to his or her neighbor. The graceful whirls become organized, begin to move, cross one another, overtake each other, while the spectators whom age has rendered immobile mark the time and the rhythm, uniting in thought and memory with a pleasure that they are denied. [Après de départ des souverains, les orchestres se mirent à exécuter des valses. Aussitôt une commotion électrique parut se communiquer à cette immense assemblée. L‘Allemagne est la patrie de la valse; c‘est dans ce pays et surtout à Vienne que, grâce à l‘oreille musicale des habitants, elle a acquis tout le charme qui lui est propre; c‘est là qu‘il faut voir, dans cette course tourbillonnante et toujours réglée par la mesure, l‘homme soutenir et enlever sa compagne, celle-ci céder à ce doux entraînement, pendant qu‘une sorte de vertige donne à son regard une vague expression qui augmente sa beauté. Aussi a-t-on peine à concevoir l‘empire qu‘exerce la valse? Dès que les premières mesures se sont fait entendre, les physionomies s‘épanouissent, les yeux s‘animent, un frémissement court de proche en proche. Les gracieux tourbillons s‘organisent, se mettent en mouvement, se croisent, se devancent, tandis que les spectateurs que l‘âge réduit à l‘immobilité marquent la mesure et le rythme (sic), s‘unissant par la pensée et le souvenir au plaisir, qui leur est refusé]. (La Garde-Chambonas 1901, 36–37)

The music played at such balls was decidedly utilitarian—the focus was, after all, on social interaction rather than on musical reflection—and yet, as suggested by Wilson’s and Marx’s observations on the waltz, all dance compositions were not created equal. The composers whose work found most favor with early nineteenth-century audiences (judging by the number of waltzes they published and their commercial success)—including Johann Strauss, Sr., and Joseph Lanner—married a thorough understanding of the practice of waltzing to musical designs that reframed and recon textualized the unvarying revolutions of the dance (McKee 2012, chap. 3).

An example of this marriage is provided by Joseph Lanner’s Terpsichore-Walzer op. 12, published in 1827—the same year that Martha Bradford wrote to her sister Alicia with concern about her nephew Edward Wilmot’s readiness to join into Vienna’s social life. Although the Terpsichore-Walzer was presented as a work for piano (the medium for which the vast number of Lanner’s waltzes were published) it should be kept in mind that versions like this often served as a short score for a performance by a small orchestra, the members of which would extract the melody, accompaniment, and bass parts from the piano arrangement (a practice that was used by dance bands well into the twentieth century). Accordingly, the work comprises an introduction, six waltzes, and a coda, which together would provide sufficient music to temporarily exhaust the dancers (at least until the next set of waltzes began). Although triple time prevails throughout and no tempo changes are indicated in the score, the different sections clearly have different musical functions: the introduction serves to get the attention of the dancers, not least by activating the two-measure rhythmic structure typical of the waltz; the sequence of waltzes provides a secure foundation for uninterrupted dancing; and the coda provides a breathless culmination for the whole. These functions are also supported by key relationships among the different sections of the work: the introduction is written in a stormy and dramatic C minor; the waltzes alternate between C major (No. 1), F major (No. 2), C major (No. 3), G major (No. 4), C
major (No. 5), and finally (adding significant harmonic coloring to the whole) A♭ major (No. 6); and the coda concludes in a resolutely affirmative C major.

A fuller sense of the support that Lanner’s music provides for the practice of dancing emerges with a close consideration of some of the compositional strategies he uses in the first three waltzes of the Terpsichore-Walzer, the scores for which are given in example 5.5.

One thing that even a quick survey of example 5.5 makes evident is the prevalence of two-measure groups that correlate with the steps of the dance. These groups are a consequence of a number of compositional strategies that are combined in various ways, including the following:

- agogic accents in the melody on the first beat of the measure (prominent in Waltz No. 1 and the B section of Waltz No. 2, but rarer in the melody of Waltz No. 3)

Example 5.5  Waltzes No. 1–3 from Joseph Lanner’s Terpsichore-Walzer, op. 12
• a “guitar-bass” pattern (Maintz 1995, 132) in the accompaniment provided by the left hand, in which a single low-register note alternates with a pair of two- or three-note simultaneities in a higher register; in these three waltzes this accompaniment pattern predominates, changing only at cadence points (which typically end with a single bass note followed by a lone chord) and in measures 9–10 and 13–14 of Waltz No. 2 (when an oscillating eighth-note pattern momentarily takes over)

• groupings of melodic materials into two-measure units, either through direct repetition (cf. measures 1–2 and 5–6 of Waltz No. 1) or through shared rhythmic patterns (cf. measures 9–10 of Waltz No. 2)

• a harmonic rhythm that marks two-measure units (cf. measures 1–4 of Waltz No. 3, where two successive measures of tonic harmony are followed by a change to the dominant seventh followed by another measure of tonic)
As I suggested earlier, the two-measure groups that prevail in these waltzes serve as a sonic analog for the dynamic process of performing the waltz’s revolving embrace. And, as La Garde-Chambonas’s description of the waltzing that took place during the Congress of Vienna indicates, the music of the waltz not only provided a guide for the dancers’ steps (“this whirling race always controlled by the beat”) but also activated the motor responses of the spectators “whom age has rendered immobile.”

As was evident in the opening dance from *La bourrée d’Achille*, the temporal framework that serves as a guide for listeners and dancers alike is a product of a number of musical materials working together in concert. Aspects of these same materials contribute to the tonal framework for the dance—indeed, to a certain extent the tonal framework emerges from the periodicities at the core of the temporal framework. In Waltz No. 1, for instance, the repeated C major harmonies in the opening measures together with the waltz motive’s emphasis on 5 and 1 provide a solid anchor for the pitch materials that follow. Similar strategies shift the tonal focus to F major for Waltz No. 2 and then back to C major for Waltz No. 3.

Although the primary purpose of waltz music was to provide a sonic analog for the steps of the dance, in practice the music of most waltzes was in a wordless dialog with those steps. In some cases, this dialog emphasized what Nicholas Cook, in his work on multimedia, would call conformance (Cook 1998, chap. 3): in measures 1–2 and 3–4 of Waltz No. 2, for instance, Lanner’s music provides a relatively straightforward sonic analog for the sequence of six steps that sets the dancers spinning around one another. In other cases, the music for the waltz and the steps of the dance are in a relationship closer to what Cook would call contest: in measures 1–2 and 5–6 of Waltz No. 1, for example, the immediate repetition of melodic material creates a cross-rhythm that pushes against the six steps of the turning figure, dividing it into two groups of three. A similar situation obtains with the four-measure groups that also feature in the design of Lanner’s waltzes. In some cases—for instance, in measures 1–4 and 5–8 of Waltz No. 1—these larger groups are a natural consequence of the concatenation of two-measure groups and thus would conform directly to the steps of the dance. In other cases, however, the four-measure grouping takes precedence, as in the A section of Waltz No. 3 (the eight measures of which have an aa’ structure); these four-measure units, while not in direct contest with the steps of the dance, would give the sense that the sequence of steps that correlates with the first two measures begins a movement completed by the steps of the second two measures. It also bears mention that the B section of each waltz typically has more variety than the A section, something evident in the contrasting melodic and accompanimental material of measures 9–16 of Waltz No. 2 as well as the circle-of-fifths harmonies in measures 9–12 of Waltz No. 3. This variety would create contrasting sonic environments for the sequences of turning figures performed by the dancers: the musical context for the eight turning figures they described during two statements of the A section is invariably different from the musical context for the eight turning figures they described during two statements of the B section.
It should be noted that this variable relationship between the music and steps of the waltz is somewhat at odds with the view espoused by Marx in his guidelines on how to compose a waltz. As pointed out earlier, Marx emphasized that “each measure, or, better, each phrase of two measures, must answer to the dance motive marking the first step firmly, and also the swinging turn of the dance” (“Jeder Takt, oder besser: jede Klausel von zwei Takten muss also dem Tanzmotive entsprechen, den Antritt fest bezeichnen und die schwingende Wendung des Tanzes …”) (Marx 1837–38, 2: 55). To illustrate his point, in the first edition of his composition treatise Marx called upon a rather rustic waltz that featured prominently in the first act of Carl Maria von Weber’s *Der Freischütz* and that was indeed pervaded by the waltz motive. As Marx noted, in writing this dance Weber “wanted to portray an uncultivated group giving itself to pure pleasure in waltzing, recklessly; thus he had nothing sound out except the pure waltz figure” (“… eine ungebildete Menge sich der blossen Walzlust rücksichtslos hingeben, und daher nichts als die rohe Walzfigure erklingen lässt”) (Marx 1837–38, 2: 56). While Weber’s waltz provided a clear illustration of the relationship between the music and movement that Marx wanted to emphasize, Marx was aware that the example obscured the more complex relationships that might obtain between music and dance. In the third edition of his treatise he removed Weber’s waltz, replaced it with a few fabricated examples of waltz motives, and (reflecting on the possibilities offered by chains of waltzes) observed that “Strauss, Lanner, Labitzky, Gungl, among others, have, in this more convivial than artistic form, occasionally accomplished most attractive things and, in the later passages of such dance sequences, have known how to compose very freely without, however, losing sight of the dance motive” (“Strauss, Lanner, Labitzky, Gungl u. A. haben in dieser mehr geselligen als künstlerischen Form bisweilen höchst Anziehendes geleistet und sich in den spätern Partien solcher Tanzreihen sehr frei zu bewegen gewusst, ohne doch das Tanzmotiv aus den Augen zu verlieren”) (Marx 1846–47, 2: 94).

Indeed, if we approach the published version of Lanner’s *Terpsichore-Walzer* as a short-score version of a sequence of dances that would be performed in their entirety in a ballroom, certain aspects of the work that would have informed the experience of the dancers come to prominence. For instance, the more adventurous B section of each waltz would contrast with the more staid A section of the waltz that followed (with the B section of Waltz No. 6 answered by the vigorous affirmation of C major and the waltz motive that is offered by the Coda). The dynamics of this basic design are then enhanced by the character of each successive waltz. For instance, Waltz No. 1 sets off as a proper beginning for a chain of waltzes—predictable in its grouping of two-measure units and large-scale alternation of tonic and dominant—but over the course of its sixteen measures it introduces features that suggest that the beginning it offers is not as assured as first appears. These include the somewhat abrupt return to tonic in measure 8 and ambiguities in the way in which measures 9–16 should be grouped. The latter are a consequence of the melody and harmony of these measures, which make it possible to affiliate measure 12 with measure 13 or measure 14 with measure 15; each of these affiliations cuts across the two-measure units and six-step patterns that are typical of the waltz. Waltz No. 2 provides a contrast not so much through
the change to F major (which remains a trifle ambiguous until the B♭s of measure 5) as through its straightforward projection of the two-measure waltz motive, through the steady eighth notes and cross-rhythms provided by the melody of the A section (the registral shifts and dynamic contrasts of which divide each measure into two equal parts), and through the contrast offered by its B section. The A section of Waltz No. 3 then returns to C major and a less complex rhythmic surface, the stability of which is subsequently called into question by the notable if mild chromatic peregrinations of the B section. Lanner’s artistry here is thus less a matter of the design of individual waltzes and more a matter of the way he constructs a dynamically unfolding sonic topography through which the dancers move.

Over the course of these and the following waltzes, the dancers would continue with their swinging turns and circuits through the ballroom. In the eight or so minutes it would take to perform Lanner’s six waltzes and coda, a pair of dancers who began waltzing on the first measure of Waltz No. 1 and continued through the end of the coda would make a total of 186 complete revolutions. Although dancers who could summon this degree of dedication and athleticism were probably the exception rather than the rule—couples might join or leave the dance at a number of points, or pause briefly to switch partners—what would not change were the steps of the dance, which remained fundamentally the same throughout. This represents a marked contrast with the dance practice embodied in Pécour’s La bourrée d’Achille (in which the step-units were constantly changing) and points to a significant shift in the structure of social dance. In the dance practice of seventeenth- and eighteenth-century France, the emphasis was on relatively subtle relationships between the steps of a dance and its music, relationships that were borne out by dances properly executed. Should a couple performing Pécour’s La bourrée d’Achille stop in the middle of the dance it would be regarded as a failure: a breakdown in the sophisticated knowledge necessary to perform a complex choreography, in the coordination of this choreography with music, and in the projection of galanterie. Should a couple dancing to Lanner’s Terpsichore-Walzer stop, it could equally be regarded as a failure, especially if it was clear that one or both of the dancers was not in command of the steps of the waltz or their correlation with the music. It could also, however, simply be a sign that one or both were ready to change partners, or had become fatigued, or had noticed that someone entering the ballroom needed to be properly greeted. Collasse’s music was thus designed to support and complement a carefully worked-out choreography, one that would bring pleasure to those who understood its challenge and nuance. Lanner’s music, by contrast, was designed to support and complement continuous movement, for it was by such movement that the social function of the ballroom—understood as a site for the exploration and expansion of social alliances—could be realized. With respect to musical grammar, this shift meant that the social function of the waltz was realized not through music’s providing one component of a delicately coordinated multimedia experience—an experience shared by musicians, dancers, and audience members alike—but through providing an invariant sonic template onto which the dancers could map their movements. Put another way, to realize the social function of French noble dance, the steps and music of each dance had
to be performed; to realize the social function of the Viennese waltz, the steps and music of the dance had to be inhabited.

Waltzes and the Salon

Although works like Lanner’s *Terpsichore-Walzer* were intended as self-contained wholes, in the form they reached the music-buying public—that is, in versions that could be played by a single pianist—they might be performed in a variety of contexts. One could, for instance, imagine a student building her piano skills by playing only one or two of the waltzes rather than all six, or a few of the waltzes being excerpted for the round of dancing that typically concluded entertainments at a bourgeois salon (Hanson 1985, 119). Indeed, waltz collections ranged from a sequence of waltzes that could be performed as a complete set, to collections of waltzes by a single composer from which individual dances could be selected for performance or dancing, to collections of popular waltzes written by a number of different composers. As one example of a particularly popular waltz, consider Franz Schubert’s “Trauerwalzer,” most likely written in 1816 but not published until 1821 when it appeared as the second of thirty-six waltzes that made up Schubert’s op. 9 (D. 365), commonly known as the “Erste Walzer.” As Maurice Brown observed,

The widespread popularity of this dance in [nineteenth-century] Vienna is something almost unbelievable; it was passed on from individual to individual, from group to group, sometimes by transcriptions, sometimes, evidently, purely by ear, until the whole city knew it, and began to call it the “Trauerwalzer”—literally, the “mourning” or “mournful” waltz. Schubert himself, first learning of the nickname, said, “What ass would compose a ‘mournful’ waltz!” Only Schubert’s immediate circle knew of his authorship of the dance; to the rest of Vienna its anonymity gave it the status of folk-music. (Brown 1966b, 292–93)

While we may never have a fully satisfactory account of the popularity of Schubert’s waltz (the score for which is given in example 5.6) certain of its features suggest why it might have caught the attention of waltz-obsessed Vienna. These include the following:

- Schubert’s use of the two-measure waltz motive to recontextualize melodic and harmonic relationships. Thus the material of measures 1–2 (tonic leading to subdominant) is transposed up a step for measures 3–4 (supertonic leading to dominant), and the material of measures 9–10 (dominant seventh leading to tonic minor) is transposed down a third (enharmonically respelled) for measures 11–12 (the dominant seventh of E♮—standing in for a chromatically lowered submediant—leading to an E♮ major chord, respelled in measure 13 as an F♭ major chord). In a similar fashion, measures 5–6 are restated almost exactly in measures 7–8, but with the important difference that the melody of measure 5 skips up a fourth to A5 on the
Music and Dance

Example 5.6 Franz Schubert, “Trauerwalzer,” op. 9 no. 2 (D. 365)

downbeat of measure 6, whereas the melody of measure 7 skips down a fifth to A4 on the downbeat of measure 8, a cadential gesture echoed and supported by the changed accompaniment of measure 8.

- Schubert’s use of a contrapuntally more active bass line in the first four measures. The rising third—A♭2–B♭2–C3—of measure 1 not only participates in a voice-exchange with the melody’s C5–B♭4–A♭4 but also connects with and brings to prominence the D♭3 on the second beat of measure 2. There is then further attention directed toward the second beat of the second measure of the waltz motive—which would, of course, be replicated in measure 4—provides an elegant way to articulate the last of the two-step components of the waltz’s turning figure.

- The subtle relationship between melodic whole-steps and half-steps that plays out over the course of the waltz. As noted above, accented passing tones that resolve downward by step are an important part of Schubert’s melody in measures 2 and 4, and similar stepwise sequences also feature
in measure 5 (as an appoggiatura) and measure 7 (as an accented neighbor tone). Schubert employs similar sequences in the B section but makes use of chromatic modifications to stretch the fabric of A♭ major. In consequence, the F5–E♭5 of measures 5 and 7 becomes F♭5–E♭5 in measure 9; the suspension figure in measure 10 resolves not to C5 but to C♭5; and when the replication of that suspension figure in measure 12 is restated in measure 13 A♯4 (alias B♭4) becomes B♭4. Capping off this series of chromatic inflections and restoring the fabric of A♭ major, the C♭5 of measure 10 is transformed back into C♭5 in measure 14, and the F♭ first heard in measure 9 is replaced by F♮ in measure 15.

In addition to providing a sonic analog for the dynamic process of performing the steps of the waltz, then, the “Trauerwalzer” also gave texture to that performance through melodic and harmonic relationships between statements of the waltz motive, through the articulation of the steps of the waltz, and through a sonic journey away from and then back to the terra firma of A♭ major. It also bears mention that the deflection toward A♭ minor and the chromatic half-steps through which this is effected make it possible to correlate the B section of the waltz with a somewhat more somber emotional realm. If not quite the stuff of mourning, these features suggest that the waltz inhabits an emotional register somewhat more melancholic than that typically associated with waltzes.

On the one hand, it seems likely that some or all of the nuances of Schubert’s waltz that I have described would have been lost on average Viennese citizens; on the other hand, the popularity of the dance is undeniable, and in the absence of other evidence it seems appropriate to attribute a portion of this popularity to Schubert’s compositional craft. The “Trauerwalzer” also points to another function of waltz music in early nineteenth-century Vienna, and that was as a prompt and focus for reminiscence and reflection. The social experience of the ballroom and the role of dancing in that experience were clearly important to the Viennese, but balls were not everyday events: although they might, as Martha Bradford remarked, come on “thick and threefold” during Carnival, there were also times during the year when dancing was prohibited (Hanson 1985, 151). At such times the performance of a memorable composition like the “Trauerwalzer” could recall, if only for a moment, the heady press and rush of the ballroom and the personalities, sounds, and images through which it was enlivened. Again, the waltz—understood as sequences of patterned movement correlated with sequences of patterned sound—played a significant role in shaping the habitus of the citizens of early nineteenth-century Vienna by providing a social environment they could inhabit both in the moment and through simulations prompted by hearing waltz music.

The salons in which the “Trauerwalzer” might be played—either as a vehicle for reminiscence or to accompany dancing—were also a site for Schubert to explore compositional practice through playing and improvising waltzes. One of Schubert’s friends, Leopold von Sonnleithner, recalled that Schubert “sometimes went to private balls at the houses of families he was intimate with; he never danced, but was always ready to sit down at the piano, where for hours he improvised the most beautiful waltzes; those he liked he repeated, in order to remember
them and write them out afterwards” (Deutsch 1958, 121). We can only speculate about the dances that Schubert produced during these performances, but the benefits of using the waltz as a template for improvisation include the rhythmic structure dictated by the dance (with successions of two-measure waltz motives providing forward impetus as well as a basis for four- and eight-measure phrases) and the convention of anchoring the succession of the dancers’ whirling turns with periodic cadences (most typically at the end of eight-measure sections).

A glimpse into the ways these factors work to shape a waltz can be seen in the fourteenth waltz from Schubert’s Sechzehn Ländler und zwei Ecossaisen (most likely written in 1822 but published in 1826 as his op. 67, and also known as the “Wiener Daimler-Ländler”), the score for which is given in example 5.7. (Despite there being historical differences between the dances it is worth noting that Schubert used the terms “Walzer,” “Deutscher,” and “Ländler” more or less interchangeably. For a discussion, see Brown 1966a, 218–19.) Perhaps the most remarkable feature of this waltz is Schubert’s apparent disregard for the niceties of tonal structure: the A section cadences not on the tonic or the dominant but

Example 5.7 Franz Schubert, No. 14 from Sechzehn Ländler und zwei Ecossaisen, op. 67 (D. 734)

![Example 5.7 Franz Schubert, No. 14 from Sechzehn Ländler und zwei Ecossaisen, op. 67 (D. 734)](image-url)
on the supertonic (C♯ minor); the B section opens by more or less disregarding that cadence and starting with a subtonic (A major) harmony that comes to serve as the dominant seventh of the chromatically lowered mediant chord of measure 11 (D major), which is subsequently transformed into the F♯ dominant seventh chord of measure 12. Although a sequence of harmonies such as this might come naturally enough to an improviser, especially at the keyboard (where the shift from one harmony to the next can be accomplished by relatively small displacements of the fingers), rationalizing them was another matter altogether. For instance, from the perspective of Hugo Riemann’s dualist theory of harmonic relationships, the shift from the C♯ minor chord of measure 8 to the A major chord that begins measure 9 could be described as a Leittonwechsel (Riemann 1880, 22–23), and the shift from D major to F♯ major (disregarding, for the moment, the dissonant seventh) as a Terzschritt (Riemann 1880, 26). As Richard Cohn and others have noted, the novel harmonic relationships captured by Riemann’s “weschels” and “schritts” were ones that Schubert exploited extensively in his mature instrumental works (and that Cohn captured in a more compact and elegant form through the notion of hexatonic systems; see Cohn 1996, 1999, and 2012, chap. 2); that said, explaining the nature of such relationships and their role in the organization of large-scale musical works remains the subject of lively discussion (Gollin and Rehding 2011). In any event, what is truly striking is the rapid appearance of such novel harmonic relationships within the tidy sixteen measures of Schubert’s waltz. Again, this is something made possible by the rhythmic structure and conventions of the waltz: with respect to the grammar of the waltz, the guideposts provided by harmonic relationships tend to be less important than the clear projection of the waltz motive and the articulation of groups of waltz motives at regular intervals.

In his design for the waltz Schubert chose to set out the three main beats of the waltz through a succession of voices sounding in different registers: in the first measure, the bass’s B1 marks the first beat, the tenor’s F♯3 the second beat, and the alto’s accented dyad D♯4–F♯4 the third beat. The right hand’s constant eighth notes then enact a delicate oscillating dance around these markers, in some instances coordinating with them (as in the case of the alto’s D♯4–F♯4 dyad on beat one) in other instances drawing attention away from them (in the case of the soprano’s D♯5s on the second and sixth eighth notes of the measure). Measure 2, while replicating this rhythmic pattern, moves away from the pitch material of measure 1 through Es and G♯s that serve as upper neighbors to the D♯s and F♯s of measures 1 and 3. Measures 3 and 4, which set out the next waltz motive, proceed in much the same way as did measures 1 and 2 but with the neighbor notes in measure 4 provided by pitches from the dominant seventh chord. Measures 5–8 begin as if to reprise the first four-measure unit but continue with what becomes a tonicization of the supertonic, which then becomes the focus for the cadential arrival of measure 8. The fact that we have arrived on a rather remote harmony is perhaps less important than the fact of arrival itself, for this creates a point of articulation in the succession of two-measure waltz motives, a succession that is immediately repeated. As is typical of dance forms, the B section of the waltz then ventures into the more remote reaches of tonal space. The novelty of this venture is offset
somewhat by a regularization of the rhythmic pattern that starts with the pairing, in measure 10, of tenor notes and alto dyads on beats 2 and 3 and continues with right-hand dyads on each of the main beats through measure 15. As a result, at the moment that we have moved farthest from B major, we have moved closer to the rhythmic pattern at the core of the waltz.

If we imagine the waltz performed in the intimate setting of a salon—with or without dancing—we might consider the interpretive possibilities opened up by Schubert’s harmonic choices. Over the course of their opening eight measures Lanner’s Waltz No. 1 and Waltz No. 3 and Schubert’s “Trauerwalzer” return to the harmony with which they began (and Lanner’s Waltz No. 2 comes to a pause on the dominant). By contrast, Schubert’s op. 67 no. 14 stops in a somewhat distant place, a minor-mode harmony that is, at best, on its way to the dominant (pointing the way back to tonic). This move is then followed by a bit of magic: with the second ending of op. 67 no. 14 and the commencement of its B section the harmony brightens and moves forward as the A major chord of measure 9 is transformed by stages into the dominant of the D major of measure 11. And then, with a bit more magic, the way is found back to B major and the conclusion of the waltz. It is worth noting that when repeating the B section the first bit of magic—C♯ minor brightening to A major—is lost, and we instead have the sense of retracing some, but not all, of the steps that enacted the restoration of B major. For a listener attentive to these details, Schubert’s op. 67 no. 14 could describe a sonic analog for a journey through a topography by turns familiar, and dark, and magical, and again familiar, all within the space of less than a minute. It could equally provide a sonic analog for a sequence of emotions or physical gestures, and all of these possibilities—of a journey, or of a sequence of emotions or gestures—would be in dialog with the dance steps implied by the omnipresent waltz motive.

We cannot know, of course, whether Schubert’s op. 67 no. 14 was a result of an improvisation that was later written down or a product of cool composerly contemplation. What is apparent is that the rhythmic structure and conventions of the waltz—essential features of its temporal framework—mitigate any sense of disorientation that might be attendant on the unorthodox harmonic relationships that Schubert deploys. These relationships consequently color, rather than disturb, the atmosphere of dreamy reflection created by the subdued dynamics and gentle oscillations of the opening measures of the dance. Put another way, the grammar of the waltz, focused as it is on providing sonic analogs for the dynamic process of performing the steps of the dance, licenses departures from harmonic and melodic syntax that might well, in other contexts, seem unmotivated or quixotic.

Waltzes without Waltzing

In 1836 Robert Schumann published an essay in the Neue Zeitschrift für Musik on a topic that had long fascinated him: dance music. As a youth, Schumann had been an avid dancer (as numerous entries in his early diaries make clear), and his first encounters with Schubert’s music inspired him to set down aphorisms like
“Dance music is embodied, mobile music; every move must be harmony” (“Die Tanzmusik ist verkörperte, bewegliche Musick; jede Bewegung muß Harmonie seyn”) (Schumann [1827–38] 1971, 96). As a young pianist he regularly played Schubert’s waltzes and spent happy hours improvising on the “Trauerwalzer” (which Schumann knew as the “Sehnsuchtswalzer,” and which he later used as the theme for an unpublished set of variations [Schumann (1827–1838) 1971, 177; Daverio 1997, 107]). And as the moving spirit behind a periodical concerned with contemporary music—the first to be independent of music publishing houses—he was almost certainly inundated with new dance compositions from publishers eager to have them reviewed. Yet the dance music Schumann encountered was at best of highly variable quality and at worst an affront to the high aesthetic ideals he and his friends had for music. The situation was further complicated by Schumann’s understanding of the impact that dance music could have on listeners, something evident in an essay from 1834 in which he mused on the principles according to which contemporary compositions should be evaluated. These musings took shape as a report on an imaginary machine—the Psychometer—that would simplify the music critic’s job by producing ready answers about the promise of a composer and the worth of the composer’s music (Schumann 1969, 1: 101–5). When the Psychometer was given the score for Karl Hartknoch’s Six grandes Valses and asked to which category the music belonged—to the “classics” (backward-looking reactionaries), the “juste-milieuists” (middle-of-the-road composers responsible for contemporary ephemera), or the “romantics” (forward-looking liberals)—it fell silent. Schumann explained that, where dance music was concerned, these categories were not applicable: waltzes instead were to be classified as Kopfwalzer (head-waltzes, disposable dances composed to formula); Fußwalzer (foot-waltzes, which enthralled dancers, captivated audiences, and satisfied performers, and were epitomized by the waltzes of Johann Strauss); or Herzwalzer (heart-waltzes, of which the “Trauerwalzer” was the zenith and progenitor). Indeed, the correlation between music and dance seen in Strauss’s music was something that mattered deeply to Schumann; as John Daverio has noted, Schumann “prized Strauss’s dances for their naturalness, their lack of affectation, their easy-going grace and charm, qualities toward which compositions in the more serious genres should aim” (1997, 124).

This appreciation of the potential for dance music to captivate an audience was quite apparent in Schumann’s 1836 essay, which took shape as an encounter between members of Schumann’s semifictional Davidsbündler and works by Joseph Christoph Kessler (1800–72), Sigismond Thalberg (1812–71), Clara Wieck (1819–96), Leopold Edler von Meyer (1816–83), and Schubert. Most of the compositions were not so much dance music as they were fantasies built on dance music, and thus they were of the same genre as Carl Maria von Weber’s Aufforderung zum Tanz (another favorite of Schumann’s): music not meant for dancing as such, but as an evocation of dancing and its social milieu. This was music the members of the Davidsbündler found particularly affecting, except when it strayed too far in the direction of actual dance music (as did Thalberg’s waltzes, which were held up for ridicule as being too straightforward [Schumann 1969, 1: 201]). As might be expected of a journal dedicated not only to contemporary music but to
the next generation of composers, the works by the youngest composers—Wieck and Meyer—were found to be particularly promising, with the virtuosic display of Meyer’s holding the audience spellbound.

It was also evident, however, that Schumann wanted to reconceptualize the attraction of dance music. As Zilia (one of Schumann’s names for Clara Wieck) prepared to play the first of Kessler’s polonaises, Florestan (one of Schumann’s personae) reflected that dance music invariably made him melancholy, and as she played he recalled with approval Jean Paul’s assertion that young women should only dance with young women and that men should never dance at all (1969, 1: 201–2). This sense of keeping the act of dancing at arm’s length while savoring dance music was expanded through Florestan’s account of Schubert’s op. 9 waltzes, which treated the dances as prompts for a species of listening that yielded pure aesthetic pleasure devoid of conscious thought. Schubert’s Sechzehn Deutsche Tänze und zwei Ecossaisen (op. 33; D. 783) were in turn transformed into an accompaniment for an imaginary carnival illustrated by a magic lantern and narrated by Florestan (Schumann 1969, 1: 202–3). Quite tellingly, Florestan broke off his narration and abandoned the imaginary carnival in the middle of Waltz No. 10, “at the moment of highest pleasure” (“im Augenblick des Vollgenusses”). Although we cannot know for certain why Schumann chose to end his essay at this point, features of Waltz No. 10 suggest why listening to it might afford special pleasure, and why the attendant opportunities for reflection opened up other ways to think about the function of waltzes during the nineteenth century.

Before proceeding, it bears mention that many of the collections of dances published under Schubert’s name gathered together works written over the course of several years; in consequence, attempts to view any of these collections as representing a coherent whole may not be particularly rewarding. That said, the waltzes of Schubert’s op. 33 are so various that they seem to reflect a conscious effort to catalog the different waltz types of the early nineteenth century, an impression that may well have prompted Schumann to conceive of them in terms of an extra-musical narrative. For the most part, this variety is a consequence of the way Schubert handles rhythmic figuration and melodic design (aspects of the set that are explored in detail in Maintz 1995, 120–55). Within this catalog of waltz types, Waltz No. 10 (the score for which is given in example 5.8) stands out both for the intimacy wrought by its subdued dynamics and lyrical melodic line, and for the sudden and rather remarkable intensity that is introduced in its B section. This intensity is in part a result of Schubert’s decision to reinforce the melodic line and the bass with octave doublings, beginning with the upbeat to measure 9, but a more significant factor is the striking prolongation of the dominant that takes place over the course of measures 9–16. The passage begins on the leading-tone harmony (often regarded as an incomplete dominant ninth chord in the early nineteenth century), which, through the arpeggiated bass line B–D–F, leads to a French augmented-sixth chord in measure 11. Schubert emphasizes the dissonance of this chord through a pungent setting—the dissonant second A–B is doubled in two consecutive six-note chords, with the augmented eleventh F3–B4 clearly audible in the outer voices—a pungency that is then dissolved through the resolution of the chord to the dominant in measure 12. These compositional strategies produce
a four-measure phrase that offers a marked departure from the limpid texture and two-measure waltz motives of the A section and that increases in intensity over its course until the arrival on the dominant in measure 12; within this context the French augmented-sixth chord serves not as a dominant preparation chord (as it often does) but as an embellishment of the dominant (understanding the leading-tone chord of measures 9–10 as a substitute for the dominant chord).

The passage is then repeated with the melody slightly altered, and with the leading-tone chord of measures 9–10 replaced by a B♭ dominant seventh chord in measures 13–14. In the first instance this replacement pulls the music some
distance away from the tonal world of A minor; in the second instance, the move to the French augmented-sixth chord of measure 15 now requires three chromatic modifications (B♭ to B♮, A♭ to A♮, and D♮ to D♯). The result is an intensification of the intensification: the strangeness of measures 9–12 becomes stranger still in measures 13–16, thanks to a stretching of the tonal fabric that is, curiously enough, brought to rights through the French augmented-sixth chord of measure 15. This intense passage is then followed, in measures 17–21, by a succession of chords that enacts the harmonic logic of a circle-of-fifths progression but (as shown on example 5.8) substitutes applied leading-tone chords for the more typical applied dominant chords. The sense of moving forward suggested by the harmonies is abetted by the return to clearly marked two-measure waltz motives and the purposive descent of the left hand’s accompaniment from the high-register chords of measure 17 to a replication of the register of measures 7–8 in measures 23–24.

It seems certain that the mastery of compositional craft evident in Waltz No. 10 was not lost on Schumann, nor was it lost on Johannes Brahms, who commented on the beauty of the waltz in the course of correspondence on an edition of another set of waltzes that included a slightly revised version of Waltz No. 10 (Brodbeck 1986, 40). In light of both composers’ affection for dance music, however, I would like to propose that Schumann and Brahms were responding not only to the clever way Schubert handled harmonic relationships or to the design of his melody but also to the sonic analogs he created. Again, the sonic analogs for the steps of the waltz are prominent throughout the dance, but in measures 9–16 the context for those steps changes, their forward motion arrested by two successive passages, the second more intense than the first and directed toward the dominant. This insistence—underlined by the reharmonization of measures 9–10 in measures 13–14—creates an almost palpable thickness that then melts away with the release provided by the purposive trajectory suggested by measures 17–21. The remarkable tension produced by measures 9–16 thus gains additional depth through knowledge of the practice of waltzing, a knowledge that—whether they were avid dancers or not—Schubert, Schumann, and Brahms shared. We cannot, of course, be sure whether knowledge of this sort contributed to the “moment of highest pleasure” that Florestan experienced in listening to Waltz No. 10; what does seem evident from Schumann’s essay is that the “pure pleasure in waltzing” that might be derived by dancing to waltzes like Weber’s Freischützwalzer, or Lanner’s op. 12 collection, or any of Schubert’s waltzes had come to be viewed with skepticism by Schumann and his friends.

The Waltz in the Nineteenth Century

In comparison to the carefully wrought coordination of French noble dance, the choreography of the Viennese waltz was decidedly casual: a single, repeated sequence of steps predominated, reflected in a regular sequence of two-measure waltz motives in the music. Yet this casualness provided an important resource for the social practice of the dance: the constant succession of spinning turns
made the waltz physically absorptive, and the simplicity of the step made it possible to rapidly circulate partners. The result was a dance practice that reflected a new form of sociability, one in which social status was actively negotiated at a dizzying pace almost equal to that of the dance. Within this practice the boundaries of musical phrases served not to mark the boundaries of sequences of step-units (as in the choreographies of French noble dance) but as points of articulation within a dynamically unfolding sonic topography, one that could vary with social setting. Thus for the crowded ballrooms of the Viennese suburbs, composers painted with broad strokes, aiming to weave waltzes together in a chain in which each link was as sturdy as the next, while within Biedermeier salons a more relaxed approach could be taken, one in which the delicately shaded nuances of a work like Schubert’s “Trauerwalzer” could be appreciated. In the latter context the waltz became a way to construct subjectivity, effectively providing an underscore for singular moments and, at some later point, a sonic prompt for the simulation of such moments. By such means the waltz could be decoupled from waltzing—a process enacted by Weber’s Aufforderung zum Tanz (Zbikowski 2012, 150–52)—and repurposed for fantasies like those of Schumann’s Davidsbündler.

On the one hand, the transformation I have traced—from the waltz as a means through which novel social interactions were structured, to the waltz as a correlate for subjective experience, to the waltz as little more than a remembrance of dancing—is one that reflects changes in social structure across Europe that began in the late eighteenth century and continued through the beginning of the twentieth century. On the other hand, it also corresponds to a change in aesthetic focus, one in which a close coupling of music with embodied experience was called into question. This change was clearly reflected in the music critic Eduard Hanslick’s remarks on music and dance in his Vom Musikalisch-Schönen (On the Musically Beautiful), first published in 1854. While Hanslick acknowledged that waltzes could indeed move an audience, he believed that responses of this sort could never rise to the level of aesthetic appreciation, equating the effect of dance music with that of drinking wine:

Music loosens the feet or the heart as wine the tongue. Such conquests tell us only about the vulnerability of the vanquished. To undergo unmotivated, aimless, and casual emotional disturbances through a power that is not en rapport with our willing and thinking is unworthy of the human spirit. When people surrender themselves so completely to the elemental in an art that they are not in control of themselves, then it seems to us that this is not to the credit of that art and is still less to the credit of those people. (Hanslick 1986, 61)

As Hanslick saw it, music that spoke directly to the feet or the heart—Schumann’s Fußwalzer and Herzwalzer—was not music that spoke to the spirit: it was at best negligible and at worst a corruption. This was a view that came to dominate musicology as it developed in the decades that followed, obscuring the role of embodied knowledge in shaping our understanding of music as well as the part music plays in the formation of social and cultural practice.
Conclusion: Music, Social Dance, and Musical Grammar

As I noted in chapter 4, musical grammar as a whole is built on an infrastructure of common conceptual ground and shared intentionality, and it is guided by the social motivation to share attitudes and feelings. The process of sharing attitudes and feelings is abetted by social dance, which supports what William McNeill called muscular bonding (1995, chap. 1). To be sure, music has many resources through which to summon sonic analogs for physical movement, but social dance is first and foremost predicated on such movement, and on the phenomenon McNeill called “keeping together in time” (a phenomenon also enacted by many performing ensembles). More broadly, the coordination of patterned bodily movement with sonic analogs for dynamic processes anchored by temporal and tonal frameworks has the potential to shape—and indeed enact—the habitus of a given social milieu. In some cases, such as French noble dance, this coordination is tightly organized: step-units and step-sequences are closely correlated with sequences of musical events, such that the two together offer a means of creating a single unified whole. A close correlation of this sort does not, however, mean that every physical movement will correspond with some musical event, and every musical event with some movement: as my analysis of the first bourrée from Pécour’s *La bourrée d’Achille* showed, even in tightly organized dances movement and music may diverge from one another. What was important within the practice of French noble dance—and this was true of carefully choreographed dances as well as more loosely structured dances like the ballroom minuet—was that movement and music should agree in their beginnings and endings, working together in between to create the impression of a flawlessly coordinated performance, for it was through such performances that the dancers could give pleasure to their audience and by this means secure or reinforce their social standing. (For discussions of the ballroom minuet, or *menuet ordinaire*, see K. Tomlinson [1735] 1970, 124–26; Thorp 2003; Russell 2006; and McKee 2012, chap. 1.)

Over the course of the eighteenth century the relatively stable and much ramified social relationships on which practices like that of French noble dance were based began to crumble and then collapse in the wake of epochal social and cultural changes across Europe. By the early nineteenth century social dance had assumed a very different form: the steps of dances like the waltz and, later in the century, the polka were relatively straightforward and for the most part easily learned, and while it was important for the music to provide a sonic analog for the basic movements of the dance the coordination of movement and music did not go much beyond this. Such dances provided a matrix for energetic, coordinated physical movement and a ready circulation of partners; these contributed to social cohesion at the same time that they provided a means through which social alliances could be reconfigured. As the close coordination of movement and music in French noble dance contributed to a habitus of dense and highly structured social relationships, so the looser coordination of movement and music in early nineteenth-century Vienna contributed to a habitus in which
social relationships became a matter of who was able to participate and who was not: as Martha Bradford observed, “woe be to him who does not Waltz.” Within this practice music provided an essential support for dancing as well as a sonic topography through which the dancers could move—as Marx noted, the best composers of waltz music occasionally accomplished real art while still keeping the waltz motive in sight. For an age that increasingly valued individual subjectivity and the sentiment through which it might be expressed, such topographies could also serve as an anchor for memorable moments in the ballroom or salon: what began as a way to construct social relationships was thus transformed into a way to construct the self.

As is clear both from contemporary accounts and from a close study of the music and steps for social dances, the grammar of dance music reflected and anticipated the movements of the dancers. In the terms I have developed here, the music presented a sonic analog for the dynamic processes associated with individual step-units as well as the larger trajectories of step-sequences. The grammatical units of dance music thus combine form and function: as Marx noted, in the case of the waltz each two-measure unit had to answer to the dance motive, firmly marking the first step and the swinging turn of the dance. Design principles of this sort ensured that dance music was danceable, and that music and dance could realize the social functions for which they were intended.
CHAPTER Six

Music and Words

One of the themes with which I began this book, and that has surfaced at various points in the preceding chapters, is the relationship between music and language. On the one hand, it seems quite evident that music and language exploit many of the same general cognitive capacities. Indeed, given that these two modes of communication, in their fullest forms, are seen in only one species on the planet it would be surprising if this were not the case. On the other hand, it seems equally evident that each has its distinctive features. Where language relies on symbolic reference (and demonstrates all of the flexibility and complexity symbolic reference entails) music is built on a framework of analogical reference (which both gives it immediacy but anchors its meanings less securely). Language offers us a single syntactic stream that allows us to transcend the present through conceptual representations of the past and future. Music offers us the possibility of multiple coordinated syntactic streams that allow us to create a present—an ongoing sequence of temporal processes—of unparalleled richness.

In this chapter I would like to explore what happens when these two modes of communication combine with one another through the instrumentality of the human voice. My use of the term “word” in the title to this chapter draws on common dictionary definitions of the term—"something that is said" or, more precisely, “a speech sound or series of speech sounds that symbolizes or communicates a meaning without being divisible into smaller units capable of independent use”—and thus excludes the units of meaning associated with sign languages, and places a bit of distance between the notion of “word” and how words might be represented through the medium of print.

The topic of how music and words might be combined is a potentially vast one, stretching from incantatory ritual practices of ancient origin, through various music-theatrical works of the early modern period, and into the different species of rap that developed in the late twentieth century. The issue is further complicated by the exceptional form of linguistic utterance—namely, poetic verse—that often forms the basis for combinations of music and words. As a way to make the exploration of the relationship between music and words somewhat more tractable, in what follows I shall concentrate on a single genre and a relatively isolated historical moment: the German Lied in the first decades of the nineteenth century. My aim, then, will not be to attempt to say everything that could be said about how music and words combine, but to use nineteenth-century Lieder from a particular era as a lens through which to view interactions between the two communicative media. The way music and words combine through the instrumentality of the voice shall thus be important, but equally so will be the way...
the voice combines with the resources offered by the piano (the instrument most typically used to accompany Lieder).

There are three principal advantages to this approach. First, composers of Lieder typically gave close consideration to the poetry they set with the aim of finding a musical setting adequate to the demands of the text. To be sure, most of these composers set a wide range of poetry and not all of their efforts were equally successful, but as Yonatan Malin has shown, the close correspondences between music and words evident in many Lieder did not come about by chance (Malin 2010). Second, the opening years of the nineteenth century were a period of rapid stylistic change in the Lied. This change was due in part to an increased commercial market for Lieder (which was a genre ideally suited to performance in the homes of the emerging bourgeoisie) and in part to a growing sense that the genre offered dramatic opportunities equal to that of the German ballad (Dürr 1984, 181–243). Early nineteenth-century Lieder thus constitute a kind of laboratory in which composers experimented with the ways music and words could combine. Third, scholarly considerations of the relationship between music and words—including my own (Zbikowski 1999; 2002, chap. 6)—have often centered on the Lied. The account I offer here, then, will connect not only with my previous work but also with that of other scholars who have used the analysis of Lieder to frame their thoughts about how music and words combine. (As noted in chapter 1, David Lewin saw the essays collected in his Studies in Music with Text [2006]—many of which were concerned with Lieder—as an extended investigation of interactions between music and text.)

The Lieder I will focus on in this chapter are three settings of Johann Wolfgang von Goethe’s “Über allen Gipfeln ist Ruh,” all of which were completed in the ten years between 1814 and 1824. Goethe’s poem has generated an extended critical commentary in its own right, not least because it is thought to embody, in eight short lines, many of the most important attributes of lyric poetry. In the first section that follows I will outline the historical context for the poem and describe some of its key features. The next three sections will take up the various settings. The first of these, by Carl Friedrich Zelter (1758–1832), was done in 1814; the second, by Carl Loewe (1796–1869), dates from 1816–17; the third, by Franz Schubert, was completed sometime before July 1824. In each case I will offer an analysis of the grammatical structure of the setting and sketch how the concepts activated by the music interact with the concepts activated by Goethe’s poem. In a fifth section I will provide a more formalized account of these interactions by making recourse to the technology of conceptual blending. In the conceptual integration network embraced by each song, one of the mental spaces is set up by music, another by words, and the coordination of these through the composer’s setting gives rise to a unique blended space. Giving close attention to how these composers responded to—and, to a certain extent, reshaped—Goethe’s poem reveals much about the ways the compositional strategies used to organize musical materials can shape our understanding of words and also offers an opportunity to consider in a bit more detail differences between the contributions that music and words make to the construction of meaning. In a sixth section I will use John Hollander’s
discussion of the compositional strategies employed in writing poetry—some of which replicate the strategies employed in writing music—to explore in a bit more detail the relationship between music and words, and consider how simply singing words (whether or not they derive from poetry) changes the way we understand them. In a concluding section I shall return to some of the general questions of how music and words relate to one another, and what this relationship might tell us about the different kinds of conceptual knowledge communicated by each.

“Über allen Gipfeln ist Ruh”

*Goethe’s Lyric*

On September 6, 1780, the young Johann Wolfgang Goethe (who had not yet acquired the “von” that signaled his elevation to nobility) paused in the midst of a trek through the Thuringian woods to spend the night in a cabin on the Kickelhahn, a summit that afforded sweeping views of the surrounding hills and forest (Segebrecht 1978, 25–27). While there Goethe wrote a short eight-line poem that he had composed on the wall of the cabin:

1 Über allen Gipfeln Over all the peaks
2 Ist Ruh’, There is peace,
3 In allen Wipfeln In all the tree-tops
4 Spürest du You feel
5 Kaum einen Hauch; Scarcely a breath of air;
6 Die Vögelein schweigen The little birds in the forest are silent.
   im Walde.
7 Warte nur, balde Wait!
8 Ruhest du auch. Soon you too will be at rest.

(Schubert 1988, 358)

Goethe did not publish the poem until 1815, when it appeared in the first volume of an edition of his collected works immediately after a poem on a similar theme that he wrote in 1776 and had called “Wandrers Nachtlied” (Goethe 1815–19, 99). (Goethe made minor changes to the poem in the final edition of his collected works, the *Ausgabe letzter Hand* [Goethe 1827–35, 99], and this is the version I use here.) Signaling the relationship of “Über allen Gipfeln ist Ruh” to the earlier poem, Goethe simply gave it the title “Ein gleiches” (“A similar one”); the poem has thus often been referred to as “Wandrers Nachtlied II.”

Although the poem did not enter Goethe’s official oeuvre for some thirty-five years, versions of it circulated in manuscript from the early 1780s on. During that time it was regarded by at least some commentators as summoning the lyrical moment that was both the source and consequence of the aesthetic attitude embodied by Goethe. Thus an unnamed correspondent, in an 1801 report on the
literati of Weimar for the *Monthly Magazine*, linked the activation of Goethe’s muse to the scene described in the poem:

His Iphigenia [i.e., *Iphigenia auf Taurus* of 1779] was composed in a wood near Weimar, which particularly recommended itself to him, when in the moments of inspiration he sought for solitude. On the wall of the hermitage, whence came forth the Iphigenia, the following verses were inscribed by Goethe, and may still be read there:

> “Ueber allen Wipfeln ist Ruh!  
> In allen Zweigen hörst du  
> Keinen Hauch!  
> Die Vögel schlafen im Walde,  
> Warte nur, balde  
> Schläfft’st du auch!” *(Monthly Magazine* 1801, 42)

The poem was, in consequence, associated with the mystique of Goethe almost from the moment it first circulated, not least because it seemed to summarize in twenty-four words all that literature could aspire to in the new age. Indeed, in the years that followed, the poem achieved a status accorded to few literary works, leading Lathrop Johnson to observe that it had become “something of a national monument, a holy place of reverence with an odour of sanctity, so that one hardly dares talk about it, let alone read it aloud” (1983, 35).

Putting the mystique that surrounds the poem to one side, three aspects of its organization have often been remarked upon by those who have sought to explore its inner workings. First, the poem does not have a set meter, such that it is difficult to predict the rhymes that shape its structure. The effect of the whole is thus much closer to something like natural speech than it is to rhymed verse. Second, the rhyme scheme—*ababcddc*—shifts with the seventh line (“Warte nur”), such that the reader or listener must indeed wait until the end of the final line for the “auch” that rhymes with “Hauch.” To an extent, then, the structure of the poem enacts the process of interiorization and reflection it describes. Third, the perspective of the poem narrows over the course of the eight lines, shifting from the largest scale to the most intimate. As Elizabeth Wilkinson observed, “It is an order of the inner process of nature as known by the mind, an organic order of the evolutionary progression in nature, from the inanimate to the animate, from the mineral, through the vegetable, to the animal kingdom, from the hill-tops, to the tree-tops, to the birds, and so inevitably to man. The poet-wanderer here is not embracing Nature in the Romantic way. He is, of necessity, by the very order of the poem, embraced within it, as the last link in the organic scale of being” (1949, 318).

Although Goethe clearly valued the poem—he was quite touched by seeing it on the wall of the cabin when he visited the Kickelhahn on the eve of his eighty-second birthday (Segebrecht 1978, 36–39)—it appears that in most respects he regarded it as a poem like any other. As such, it was among a number of poems he deemed “singable” that he sent to his friend and musical correspondent Carl Friedrich Zelter in February of 1814. Zelter received the poems on February 21,
immediately set “Über allen Gipfeln ist Ruh” to music (giving it the title “Ruhe”), and sent the completed song to Goethe the next day (Bodley 2009, 169). A few months later Goethe, in a letter to Zelter in which he reflected on both the song and the turbulent times that Weimar was experiencing in the wake of the Napoleonic wars, wrote “The ‘Ruhelied’ is admirable; our tenor [Carl Melchior Jakob Moltke] sings it very well, and in these times of unrest, it is all the happiness we get” (Bodley 2009, 175). It bears mention that Goethe’s extended correspondence with Zelter makes quite clear his opinions on how music and words should interact with one another, opinions that provide valuable insights into perspectives on text setting current at the beginning of the nineteenth century.

In the analyses of the settings of “Über allen Gipfeln ist Ruh” that follow, I shall devote most of my attention to the ways Zelter, Loewe, and Schubert organized their musical materials, and I shall have rather less to say about how Goethe organized his linguistic materials. This reflects in part the view that a poem becomes less a poem and more a text once it is in the hands of a composer (Youens 1991, 307–12; Zbikowski 2002, 254), but it also reflects the way musical settings change the delivery of words, a topic engaged more fully toward the end of this chapter. In my analyses I shall, accordingly, return to my practice of giving close attention to individual syntactic layers created by each composer’s settings so that I might develop a fuller account of how processes within these layers contribute to the sonic analog created by the whole, and thus to the way music and words interact in these songs.

*Carl Friedrich Zelter, “Ruhe”*

Zelter, who was in the main a self-taught musician and whose principal income came from his construction business, spent most of his compositional energies on vocal music, a predilection that reflected his activities as a choral conductor in Berlin (Ottenberg 2015). He wrote slightly over two hundred Lieder, seventy-five of which took as their texts poetry by Goethe. The score for Zelter’s setting of “Über allen Gipfeln ist Ruh,” as it was first published in 1821, is given in example 6.1.

**Rhythmic Processes.** The temporal framework for Zelter’s setting emerges from a steady succession of eighth notes organized through register and pitch into groups of three, interrupted by a brief silence after the tenth eighth note. On the one hand, this silence opens a gap in the framework; on the other hand, that gap might well encourage the attentive listener to simulate the two missing eighth notes based on the preceding musical materials. The succession of eighth notes then resumes with a different harmony, again interrupted after the tenth eighth note. This overall strategy confirms the basic temporal framework while breaking the piano’s introduction into a series of short subphrases. The voice, when it enters in measure 4, then plays off the structure provided by the temporal framework: although conforming to the overall meter, it makes sparing use of divisions of the main beat and is organized into longer phrases that bridge over the subphrases of the accompaniment. These compositional strategies change with the dramatic pause written in to measure 9, after which voice and accompaniment
work together to set out a steady succession of three-note groups broken only with the approach to cadence in measure 12. The accompaniment then extends the cadential arrival with a recollection of the end of the piano's introduction, which brings the song to a close.

Over the course of measures 1–9 the compositional strategies Zelter uses to organize his rhythmic materials create a sonic analog for a process not unlike that of a sequence of carefully measured breaths, with two qualifications. First, the sustained phrases of the voice in the first half of the song reinterpret the subphrases of the accompaniment, grouping them into two-measure pairs. Second, owing to the embellishment provided by the voice in measure 8, the last subphrase of the accompaniment (beginning in measure 7) is quite a bit longer than the rest: if this is an analog for a breath, it is one considerably extended. (Indeed, an attentive listener might well find herself grabbing a breath with the singer after the “Hauch” of measure 9.) The change in strategies initiated in measure 10 suggests a sonic analog for a somewhat different process, one more continuous and, in its continuity, more accumulative.
Contrapuntal and Harmonic Processes. If Zelter’s rhythmic materials provide an analog for a sequence of carefully measured breaths, his harmonic materials give shape and direction to those breaths. To summarize the way Zelter organizes his compositional materials I offer the voice-leading sketch of example 6.2, which sets out the contrapuntal interaction between the main structural voices of the song and the goals these achieve. Over the course of the piano’s introduction Zelter traces a path, marked out in four brief subphrases, that establishes E major as a point of anchor within the tonal framework (culminating on the second main beat of measure 1), situates B major in contrast to this (measures 1–2), diverges toward A major (measures 2–3), and then returns through B major to E (measures 3–4). Zelter begins the song proper seeming to retrace this path; in the event, however, the introductory material is transformed in two important ways. First, the singer’s opening phrase (which traces a stepwise descent from B4 to D♯4) stretches across the piano’s first two subphrases to conclude with an arrival on the dominant in measure 6, which is emphasized by rests prolonged with a fermata. (Here and elsewhere I will refer to the notated register of pitches in the vocal part, with the understanding that a contralto,
tenor, or baritone would render the melody an octave lower.) Second, instead of exploring further the diatonic domain of E major (as he had in measure 2) Zelter returns in measure 6 to the sustained E major chord of the opening and then shifts rather suddenly to the dominant of B major, which leads in turn to the fairly elaborate cadence on B major in measures 8–9 (again reinforced by a fermata).

Having arrived on the dominant in measure 9, Zelter interrupts the harmonic progress achieved thus far and begins once again on E major. This time, however, the B₄ that was so prominent in measures 4–5 and 6–7 is much less in evidence, and B major recedes in importance. The harmonic strategy Zelter adopts in measures 10–13 is thus markedly different from that which he used in measures 4–9: gone is the sense of harmonic progress (with successively firmer movements toward the dominant), replaced by an orderly, and unremarkable, circulation around E major.

As noted in chapter 4, tonal frameworks, which are built on a more or less accurate specification of pitch (such that it is possible to return to the “same” pitch after some interval of time), allow musicians to create sonic analogs for a path through the virtual topography described by a succession of pitches or for different sorts of goal-oriented processes. The latter are particularly evident in the opening section of Zelter’s “Ruhe”: the contrapuntal and harmonic materials of measures 1–4 provide an analog for a goal-oriented process that gently probes the diatonic space around E major before returning to its center (as instantiated by the perfect authentic cadence on E in measures 3–4); by contrast, measures 4–9 provide an analog for a goal-oriented process that moves with some ceremony toward a new tonal center, concluding with a perfect authentic cadence on B. After measure 9, however, Zelter’s compositional strategies seem to have less to do with summoning goal-oriented
processes and more to do with providing an analog for a kind of sonic sedimentation: the repeated neighbor figures around E in measures 10–11 and the somewhat perfunctory cadence of measures 12–13 speak not to forward progress but suggest instead a goal that has been all but achieved and need only be confirmed by the passage of time.

The Coordination of Syntactic Layers. Working together, the syntactic layers focused on rhythmic and on contrapuntal and harmonic processes create a sonic analog for a dynamic process with two distinct parts. The first of these moves forward in carefully measured increments, the last of which requires greater resources and culminates in a temporary goal. The second part, by contrast, presents the image of a state of repose that is not so much a consequence of the first part of the process as it is a foregone—perhaps even banal—conclusion.

Music and Words. As is often the case when a preexisting poem is set to music, the musical phrases of Zelter’s “Ruhe” generally correspond to units of sense in the poem: the first sung phrase, for instance, is given over to “Über allen Gipfeln ist Ruh,” the second to “in allen Wipfeln spürest du kaum einen Hauch.” That said, what counts as a unit of sense in a poem is not invariant. This point was made by a former poet laureate of the United Kingdom, Andrew Motion, who, in response to questions about how he rendered one of his own poems, observed, “I think one of the ways in which poems stay alive when you’re reading them on the page is by having some contest between the unit that the line is and the unit that a chunk of sense is, and the unit that a verse is, and the unit that the whole poem is. And it’s the play between those ideas of shape, of unit, that animates the thing overall” (Motion 2011). Zelter’s decision to put the sixth line of the poem (“die Vög[e]lein schweigen im Walde”) into the same phrase as the seventh and eighth (“warte nur, / balde ruh[e]st du auch”)—even though Goethe’s punctuation suggests that the sixth line continues the thought of the fifth—could thus be seen as one way to “read” the poem. Zelter also chose to make the yearning for peace almost palpable through his repetitions of “balde,” which, as realized in the song, stretch across five full beats—that is, more than a measure of notated music.

What was perhaps more important for both Zelter and Goethe, however, was the sonic environment created by the music as a whole, for it was this that they believed led the listener to a fuller sense of the words. In a letter of May 2, 1820, to Zelter that praised the composer’s approach to text setting, Goethe wrote,

> The purest and highest style of word painting in music is that which you yourself also practise; the object is to transport the listener into that frame of mind which the poem itself suggests; the imagination will then form pictures from the text without knowing how it arrives at them. You have given examples of this in your “Johanna Sebus”, “Um Mitternacht”, “Über allen Gipfeln ist Ruh”; where have you not realized it? Tell me anyone who has achieved this to the same degree as you! Painting in tones of thunder, crash, splash and splish is detestable. A minimum amount of this, as you use it, merely dotting the i’s, is effective. (Bodley 2009, 266)

Zelter’s achievement in his settings of “Um Mitternacht” (from 1818) and “Über allen Gipfeln ist Ruh” is readily apparent: the accompaniment for each song is clear and straightforward, with both the rhythmic figuration and harmonic
language providing a secure basis for melodic lines that flow easily from one phrase to the next. The case of “Johanna Sebus,” a ballad setting from 1810 for choir and soloists accompanied by piano, is somewhat more complicated. Here Zelter was unrestrained in his use of a range of compositional strategies to evoke the turbulence and terror of the flood that occasioned the eponymous heroine’s selfless deeds of valor and that ultimately took her life. It would seem, however, that Goethe believed that the use of such devices was warranted by the dramatic event that was the subject of his poem, and that anything less would miss the effect he aimed to create.

I should also note that, in each of these settings, Zelter was flexible in his handling of Goethe’s words. This flexibility is manifested in florid melismas (evident in measure 8 of “Ruhe,” his setting of “Sterne” in the first verse of “Um Mitternacht,” and throughout the choral parts of “Johanna Sebus”), text repetitions in all three settings, and a free treatment of how phrases are shaped, something particularly noticeable in “Um Mitternacht.” This way of handling text apparently met with Goethe’s approval: in a letter of March 19, 1818, to Zelter, in which he mused on theories of poetic metrics and celebrated the aptness of Zelter’s setting of “Um Mitternacht,” the poet wrote, “The 6/8 time of your ‘Um Mitternacht’ does everything. Such quantities and qualities of tone; such variety of movement, of pauses, and intake of breath; this ever-changing sameness! Let the gentlemen with their long and short syllables work towards a consensus for a long time, but they will not produce such work as yours” (Bodley 2009, 241). Although we cannot be sure (based on the evidence provided by the correspondence between Goethe and Zelter), it may also be the case that Goethe was willing to accept reinterpretations of his poetry through musical settings. My analysis of the sonic analogs created by the compositional strategies Zelter used to set “Über allen Gipfeln ist Ruh,” for instance, points to a reinterpretation of Goethe’s poem that extends beyond Zelter’s decision to put the sixth line of the poem into the same musical phrase as the seventh and eighth. The substantively different sonic analog associated with the setting of these lines sunders the connection of line 6 to lines 1–5, with the consequence that the repose of the silent birds is also the repose of the wanderer, and both of these are cut off from the process through which the wanderer became aware of the surrounding stillness.

Those who have commented on Goethe’s opinions on how words should be set to music have often noted his conservativism, and in Zelter he certainly found a musical voice more comfortable in the late eighteenth century than in the early nineteenth. This conservativism has, however, often been conceived of as advocating for a slavish devotion to the text; based both on Goethe and Zelter’s correspondence and on Zelter’s musical settings of Goethe’s poems, such a conception seems doubtful. Goethe and Zelter instead viewed music as a way to support and enhance dramatic situations activated by words. As such, the best music was music that did not call attention to itself but simply brought the poem to a higher level—in a continuation of his letter to Zelter of May 2, 1820, Goethe wrote, “I feel at once that your compositions are identical with my poems; the music, like the gas which is pumped into an aerial balloon, merely raises them up” (Bodley 2009, 266; translation adapted). In certain respects, this perspective is an essentially pragmatic one: the purpose of a song was not to create a new work of art but to bring
an existing work of art to its audience in a new way. It is also one that grants precedence to language and the resources it offers for the construction of rich mental images, and it downplays the role of music and the resources it offers for the enactment of those images.

There were, however, new compositional voices born in the new century, as well as new conceptions of how music and words might relate to one another. To get a glimpse of some of the changes in the air, let us now turn to Carl Loewe’s setting of “Über allen Gipfeln ist Ruh.”

**Carl Loewe, “Nachtlied von Goethe”**

The motivations for Loewe’s setting of Goethe’s poem, which he appears to have started in 1816, are not known. At the time he was a church organist in Halle, having previously studied with Daniel Gottlob Türk (West 2014). The song appears to have been part of his efforts to establish himself as a composer, as it was one of four he offered to a publisher in Leipzig in 1817 (Loewe 1899, 11: xxiii). In the event, the song was not published until 1831 when it was included in Loewe’s *Sämtliche Lieder, Gesänge, Romanzen und Balladen für eine Singstimme mit Pianoforte Begleitung*, Op. 9 with the title “Nachtlied von Goethe. Das Eine” (this last to distinguish it from the first “Wandrers Nachtlied” — “Der du von dem Himmel bist” — which Loewe set in 1828); the score as it appears in the 1899 edition of Loewe’s collected works is given in example 6.3.

*Rhythmic Processes.* Loewe’s accompaniment is remarkable for the simplicity of its rhythmic design, consisting as it does of a sequence of undifferentiated, equally spaced chords notated as a succession of quarter notes. Changes in harmony that correlate with Loewe’s bar lines, along with the phrasing slurs of the first two measures, suggest how these undifferentiated quarter notes should be grouped and complete the minimalist temporal framework. The only disruptions to the smooth surface of the accompaniment occur in measures 9–10, when the piano’s topmost line breaks free to anticipate the voice’s entrance in measure 11, and in measure 12, when the harmony changes on the last two beats of the measure. (It bears mention that none of Loewe’s other Goethe settings from 1816–18, including “Ich denke dein” [Op. 9, Book 3, No. 1], “Sehnsucht” [Op. 9, Book 3, No. 5], and “Erlkönig” [Op. 1, No. 3], make use of such a severely restricted rhythmic palette.) The rhythmic organization of the vocal part, by contrast, does much to shape the song as a whole. At the level of the measure, changes in register and duration collect the quarter notes of the accompaniment into three groups of two, conforming to Loewe’s $\frac{3}{2}$ meter signature. At the level of the phrase, breaks in the vocal part (along with melodic contour) suggest five longer spans; these are designated with numbers on example 6.3. Phrases 2, 3, and 4 are the most animated, activated by dotted quarter/eighth patterns and sequences of eighth notes in their centers. Phrase 1, by contrast, seems to emerge from and eventually move beyond the steady pulse provided by the accompaniment, and Phrase 5, with the gaps provided by rests in measures 14 and 15 and the deceleration enacted by measure 16, seems to disappear back into that same steady pulse.
The rhythmic strategies Loewe uses in his accompaniment set up a sonic analog for a dynamic process that is minimally inflected. Indeed, if a dynamic process is conceived of as a coherent sequence of phenomena that is distributed over time and typified by parametric modulation or change, it seems fair to say that, at least with respect to rhythm, Loewe’s accompaniment is hardly dynamic at all. By contrast, the rhythmic strategies Loewe uses for the vocal part set up an analog for a dynamic process that plays out against the backdrop provided by the accompaniment, one that grows in complexity from phrase 1 to phrase 2, begins to taper off in phrase 3, and then becomes ever more halting over the course of phrases 4 and 5.

Contrapuntal and Harmonic Processes. To a certain extent, the bifurcated strategies of Loewe’s rhythmic syntax have their concomitant in his contrapuntal and harmonic syntax. The plodding rhythms of the accompaniment, for instance,
correlate with the unexceptional succession of root-position chords that make up the harmonic fabric; the gradually increasing and abating animation of the vocal part's rhythmic figuration, by contrast, has its analog in the complexities of the contrapuntal structure that are summarized graphically in the voice-leading sketch of example 6.4. There, a play between the melody's A4 and C5—the latter promising an expansion into the upper register that is never fully realized, the former anchoring the structural line at key moments—suggests a quiet struggle that stands apart from the plodding harmonies. In truth, however, Loewe's contrapuntal and harmonic materials are guided by a single overarching idea, which is of a goal-directed process seeking but not quite achieving closure, the sonic analog for which is accomplished through a coordination of both harmonic and contrapuntal material. A compact form of the idea is set out in Phrase 1 when, above a steady tonic pedal in the bass, the B♮ of measure 3 simultaneously pulls

Example 6.3 Continued
the melody toward C5 and pushes the vocal part out of the orbit of F (a push made audible by the clash between the melody’s B♮4 and the accompaniment’s B♭3). This idea is then expanded and given harmonic support in Phrases 2 and 3, the B♮4 harmonized with a G dominant seventh chord (measure 6), the C5 harmonized with a C major chord (measure 7) to which a seventh is subsequently added (measure 8). As was the case in Phrase 1, however, the upper register suggested by the C5 of Phrases 2 and 3 never realizes its full potential: it is more or less left hanging in Phrases 4 and 5, its contribution restricted to an arpeggiated embellishment of the main structural line.

The play between A4 and C5 I have described is, to some extent, foregrounded by Loewe’s compositional strategies in Phrase 4 when the cadential arrival promised by measure 12 is thwarted by the A dominant seventh chord of measure 13. This deflection brings A4 back to prominence and occasions the modified reprise of Phrase 4 with which the song concludes. It is worth noting that these strategies were not in evidence in Loewe’s first version of the song, the conclusion of which is given in example 6.5. The two versions of the song are virtually identical through measure 11, but in the first version there is no thwarted cadence and no reappearance of A4. On the one hand, the design of the first version makes more audible the musical rhyme the concluding measures create with measures 3–4; on the other hand, it reduces the sense of play between A4 and C5, thus making for a less effective analog for a process seeking but not quite achieving closure.

The sonic analog created by the contrapuntal and harmonic processes set out in the final version of the song—which, again, is of a goal-directed process seeking but not quite achieving closure—is supported, at a more immediate level, by a
number of design features that suggest that the domain within which this process plays out is one concerned with an effortful expression of emotions. These include the following:

- the thick textures of the opening measures, which are intensified by the pedal point of measure 3 and which then relax into the somewhat more open voicing of the tonic chord used in measure 4;
- striking, if locally isolated, dissonances, such as the B₄ in measures 3 and 16 and the doubled C♯s of measure 9 (which create augmented fifths above the doubled Fs);
- the thwarted cadence of measures 12–13, which prevents a closure on F but which initiates an orderly circle-of-fifths sequence that leads the music in measured steps toward the final cadence.

As I noted in chapter 3, it is well to view with some skepticism claims that music can, either specifically or generically, express emotions. At best, what music may be able to do is to provide sonic analogs for dynamic processes that could be correlated with affectual responses. To that extent, the “expressivity” Loewe realizes through his musical materials is best viewed in stylistic rather than veridical terms. Put another way, Loewe’s setting reflects compositional strategies used by Lieder composers of the early nineteenth century to create a particular effect (namely,
“expressivity”). In this connection, it is worth noting that the performance directions for the published version of the song—“Sehnsüchtig klagend und getragen” (“yearningly plaintive and solemn”)—provide further evidence that listeners should be left with the impression that the song is a deeply expressive one.

The Coordination of Syntactic Layers. The compositional strategies Loewe used to set Goethe’s poem invite two different interpretations of the sonic analog created by the syntactic layers for rhythmic and for contrapuntal and harmonic processes. The first interpretation approaches the song as a coordinated whole, such that the accompaniment and vocal part together create a sonic analog for a dynamic process that is pulled steadily forward toward a goal that remains, in some essential way, elusive. The second interpretation takes as its point of departure the marked separation of the vocal part and the accompaniment. While it is certainly true that the two harmonize with each other (both rhythmically and harmonically), at no point does the accompaniment double the voice (which was typical in Lieder accompaniments of the late eighteenth and early nineteenth centuries). Although the way that a listener would group the accompaniment’s uninflected succession of chords would almost certainly be influenced by the organization of the melody, the accompaniment itself seems to take no notice of the vocal part. The accompaniment does, of course, anticipate the vocal part (in measure 10) and redirect it (in measure 13), but there are few if any signs of the accompaniment reacting to the vocal part. One could, in consequence, interpret the dynamic process summoned by Loewe’s sonic analogs as a product of interactions between two agents. One of these agents (associated with the accompaniment) proceeds inexorably onward, while the other (associated with the voice) starts out as independent and self-directed but eventually loses focus and falls in with the process set out by the first agent. Although the overall dynamic process summoned by the song would remain substantively the same—that is, it would be a process that is pulled steadily forward toward an elusive goal—this second interpretation would construe the accompaniment as an implacable force rather than an environmental condition, and one that ultimately triumphs over that represented by the voice.

Music and Words. As did Zelter’s setting of “Über allen Gipfeln ist Ruh,” Loewe’s provides an environment for the delivery of Goethe’s words, something especially evident in the first half of the song when an unvarying, methodical accompaniment provides harmonic goals that coordinate with the goals of successive sense-units in the poetry. After the arrival on “Walde” and tonic in measure 9, however, Loewe’s compositional strategy changes and stylistic markers of expressivity, intermittently present from the beginning of the song, come to the fore. One consequence of this change is that the last six words of the poem—“Warte nur, balde / Ruhest du auch”—take on significant weight through their correlation with expressive markers, the repetition of words, and the sheer amount of time spent on them (it takes almost as long to deliver the last six words of the poem as it did to deliver the first eighteen). To be sure, these words are given prominence in Goethe’s poem, but in Loewe’s setting they are coordinated with and delivered through sonic analogs that suggest a dynamic process that is in the course of being brought to a halt, perhaps through a set of interactions between
two opposed agents. Put another way, Loewe’s setting creates a dramatic staging for Goethe’s poem.

The notion that a Lied might be a site for a dramatic staging is one that was a commonplace in accounts of Lieder by scholars of the twentieth century. In the opening essay of The Composer’s Voice, for instance, Edward T. Cone took as his point of departure the idea that music, and in particular the Lied, was a dramatic medium through which composers and performers might project various personae. Using Schubert’s setting of Goethe’s ballad “Erlkönig” as his principal example, Cone noted that the song provided the singer with the opportunity to create as many as four characters (a narrator, the father, the child, and the Erlking) and thus offered a range of ways to explore the enactment of a musical drama. Just as important was the role of the accompanist who, together with the singer, realized the larger dramatic vision of the composer—indeed, the song as a whole was, in Cone’s terms, an expression of the composer’s voice (Cone 1974, 11).

Although the notion that a song—or, indeed, any musical composition—might provide a forum for a dramatic staging was thoroughly unexceptional for most commentators of the twentieth century, it was not always so. For Goethe and Zelter the drama engendered by Lieder was a consequence of the design of poetry rather than music. And for Hans Georg Nägeli (1773–1836), a composer and publisher in Zürich, the Lied was meant to serve not as an opportunity for the staging of drama but as a means of promoting the appreciation of poetry and the social values of singing (Nägeli 1816, 1). Although Nägeli set out, in an essay published in 1817, what he characterized as a new approach to the Lied—one in which text, melody, and accompaniment were on an equal footing—it is evident from his own works (which he offered as examples of this approach, and which include his Gesänge deutscher Frauen [1811], Lieder von Johann Jakob Heß [1814], and Lieder von I.H. Frh. v. Wessenberg [1814]) that the Lied remained for Nägeli a vehicle for the presentation of poetry and the promotion of singing rather than an art form that aspired to drama (Nägeli 1817). Within this context, then, Loewe’s “Nachtlied von Goethe” stands out as a work that does not simply provide an appropriate environment for the declamation of Goethe’s words but, through the sonic analogs he creates with his arrangement of musical materials, a dramatic staging of those words.

### Franz Schubert, “Wandrers Nachtlied”

Neither the motivation nor the date is known for Schubert’s setting of Goethe’s “Über allen Gipfeln ist Ruh” (the score for which is given in example 6.6), although it is among ten songs he mentioned to his brother Ferdinand in a letter of July 3, 1824 (Schubert and the Internationale Schubert-Gesellschaft 1985, xxii). As shown by another letter to Ferdinand later that month, it was one of only two songs among those ten (the other being “Die entsühnte Orest” [D. 699]) that Schubert thought to be of some value.

Although scholars have long known of Zelter’s and Loewe’s settings of Goethe’s poem, it is Schubert’s setting that has generated the largest body of
commentary. On the one hand, this is not surprising, for Schubert’s oeuvre has certainly had a greater influence on other composers than has Zelter’s or Loewe’s. On the other hand, the prominence accorded Schubert’s song speaks to certain ideas about the aesthetic status of Lieder that developed over the course of the twentieth century. For much of the century songs were viewed as a relatively minor genre: too brief to serve as a forum for the extended harmonic and motivic development that was taken as the mark of exceptional works, and excluded from the Elysian fields of absolute music by their association with words. Not much could be done about the scope of songs—they were only as long as they were—but in midcentury Thrasybulos Georgiades endeavored to rescue Schubert’s oeuvre with the argument that it was Schubert’s ability to meet great poetry at its own level that raised his songs—and thus his music—to the condition of great art (Georgiades 1967). Arguments such as these contributed
to a renewed interest in Lieder as a genre, but with the bar set very high: for a song to be worthy of consideration, it had to meet or exceed the artistry of the poem it set. This perspective had the effect of pitting words against music—there was really no way to explain how both could emerge victorious in song—and obscured the view of the ways music and words could interact, two topics to which I shall return toward the end of this chapter.

_Rhythmic Processes._ Schubert begins the song with a brief two-measure introduction that sets out the bare essentials of a temporal framework and then, with the sustained chord of measure 2, seems to end before it has begun. (Measure 2, slightly modified, will in fact be used to conclude the song.) The piano does, however, begin again (now joined by the voice) but it is not until measure 5 that the rhythmic strategies Schubert deploys in the accompaniment provide much of a sense of direction. Midway through measure 5 the left and right hands of the piano set up an interlocking pattern of syncopated sixteenth notes that activates the musical surface; given that the dynamic level Schubert indicates is pianissimo, the activation is a gentle one but consistent over the course of the next three measures. The syncopated pattern breaks off at the beginning of measure 9 and returns to the more spacious rhythmic pattern of the opening measures, now enlivened by sixteenth notes moving toward beats 2 and 4. The sense of deceleration created by the reprise of the rhythmic materials of the opening is given further support by fermatas in measures 10 and 12, and the return of measure 2’s cadential pattern in measures 13–14. The voice, for its part, follows the accompaniment more or less exactly in the opening
Foundations of Musical Grammar

and concluding measures, but in the middle of the song, when the accompaniment
is at its most active, the voice continues at a relatively languid pace. The one excep-
tion comes in measure 8 when a sequence of smaller divisions of the beat and an
agogic accent on the second eighth of the measure bring the voice somewhat closer
to the sense of activity generated by the accompaniment.

As a whole, the rhythmic strategies deployed by Schubert set up a sonic analog
for a dynamic process that begins in a halting fashion (in measures 1–4), becomes
more active (a process that ultimately engages the voice as well as the accompa-


Contrapuntal and Harmonic Processes. A convincing summary of the har-
monic and contrapuntal organization of Schubert’s song was offered by Allen
Cadwallader and David Gagné in their textbook for Schenkerian analysis,
where the work provides an example of the structural features of one-part forms
(Cadwallader and Gagné 2011, 218–25). Cadwallader and Gagné’s voice-leading
sketch, which is reproduced in example 6.7, illustrates the gradual development
of tonal processes in the opening measures of the song: subsequent to the piano
introduction in measures 1–2 (which provides a model for the structural voice-
leading of the song) Schubert gradually expands his tonic over measures 3–5,
with the path to the dominant inflected by the G♭ of measure 5 and the E♮ and
D♭ of measure 6. In Cadwallader and Gagné’s analysis, the dominant provides
harmonic support for the C5 of measure 7, which constitutes a key element of
the initial ascent to the D5 of measure 10. The dominant and C5 having been
achieved, tonal processes are placed in a brief holding pattern in measures 7–8
before they move forward with the auxiliary cadence of measures 9–10, culmi-
nating in the rapid structural descent of measures 10–11 (with measures 9–11
repeated in measures 11–13, to which is appended a reprise of the dominant-
tonic succession of measure 2).

Cadwallader and Gagné’s analysis captures one of the essential features of
Schubert’s setting, which is that measures 1–8 have as their overall goal the harmonic
and contrapuntal process set out in measures 9–11. One thing that is missed, in part
because of their commitment to providing a large-scale structural interpretation of
the song, is the role of measures 7–8 in this process. As treated by Cadwallader and
Gagné, these measures simply serve to prolong the dominant chord: the tonic chord
in the second half of each measure is thus an embellishment of the dominant (a
reading made evident in the interpretation of the B♭3 and D4 of the tonic as creating
a 6/4 figure over the dominant). Three things point to a somewhat larger role for this
tonic chord: the way it discharges the dissonant A3–E♭4 of the dominant seventh
chords in the first halves of measures 7–8; the prominence of the B♭3 on beats 3 and
4 of each measure (with the F3 on the off-beats providing an alternating bass note);
and the vocal part’s B♭4, which connects directly with the B♭4 of measure 9 (suggest-
ing that the B♭ major chord has suddenly been replaced by a G minor chord, her-
alding a momentary shift from major to relative minor). When combined with the
shift in register in the accompaniment in measures 7–8 the effect of the whole is one
of interiorization: while it is clear that the tonic of this passage cannot provide the repose that is the goal of the preceding measures, it nonetheless offers a momentary center from which measures 9–11 will expand out.

The sonic analog set up by these contrapuntal and harmonic processes summons a dynamic process that begins somewhat hesitantly but then gains forward motion with the approach to the dominant of measure 5. The process then stalls momentarily—almost as though it is regarding its own progress and what it has achieved—before once again moving forward, this time to a definitive closure made emphatic through repetition.
The Coordination of Syntactic Layers. One of the most interesting aspects of the coordination of syntactic layers in “Wandrers Nachtlied” is their lack of coordination at key moments in the song. Processes in both layers are initially aligned: the halting rhythmic patterns of measures 1–4 correlate with Schubert’s tentative exploration of diatonic space, and both layers acquire a forward impetus in measure 5, the rhythmic layer through the syncopated pattern in the accompaniment, the contrapuntal and harmonic layer through the introduction of chromatic pitches. And then, over the course of measures 6–11, the paths diverge: at first the divergence is marked by continued activity in the rhythmic layer while the contrapuntal and harmonic layer holds back; then, at the moment when the contrapuntal and harmonic layer makes a decisive push forward (in measures 9–11), the rhythmic surface empties out, the forward drive of the syncopated pattern having disappeared. Arguably, it is only with the reprise provided by measures 11–13 that the two layers come back into alignment, the restatement of the completion of the contrapuntal structure now confirming, rather than enacting, closure and the reiteration of the rhythmic figuration that began in measure 9 serving as a reinforcement of its halting if consistent pattern. As a whole, this interaction between the syntactic layers creates an analog for a process that begins in a tentative fashion but then moves forward; once forward motion has been established, however, something goes awry and the process first draws in on itself (with the interiorization suggested by measures 7–8) before suddenly changing and enacting the outward and explicit consummation suggested by measures 9–11, the closure of which is confirmed through its restatement.

Music and Words. Although Schubert’s setting is not overtly “expressive” (most of the stylistic markers of expressivity being borne by the slow tempo and subdued dynamic of the song, along with a bit of chromaticism) it nonetheless offers a dramatic staging of Goethe’s poem. This is accomplished in part through contrasts in rhythmic figuration and a change in the register of the accompaniment (which together animate and color the setting of lines 4–6), and in part through Schubert’s treatment of the last two lines of the poem—as was the case with Loewe’s setting, the last six words of Goethe’s poem take as long to deliver as did the previous eighteen. Schubert, in his setting, chooses to emphasize Goethe’s “warte nur” (“only wait”), both through repetition and through disposition in pitch space (the second statement being a step higher than the first), a decision that contrasts with the emphasis Zelter and Loewe put on “balde” (“soon”). I would like to propose, however, that the true drama of Schubert’s setting is a consequence of the way he organizes his musical materials as a whole.

The first indication of this drama comes with Schubert’s setting of “Wipfeln,” which is supported harmonically by a rather exposed fully diminished seventh chord (a typical marker of “expressivity” within the discourse of early nineteenth-century Lieder). The decision to do so might seem quixotic—what is it about treetops that is particularly poignant?—but immediately after this expressive incursion we are introduced to the subject who will feel (or, perhaps better, sense) scarcely a breath in the leaves of the trees. The setting of “Wipfeln” is thus not unlike a shudder of awareness through which an observing subject comes to life. The environment for that subject is then realized as one of softly thrumming
tension: harmonic progress (as realized by the approach to the dominant) comes more or less to a halt in measure 6; measures 7 and 8, caught on the image of the silent birds, explore the next step in that progress without actually taking it. What follows beginning in measure 9 is close to an imperative, seeming to come from a place outside the environment established in measures 6–8: “Wait, wait! Soon you shall rest!” The imperative nature of the utterance is then immediately reinforced subsequent to the accompaniment’s horn call at the beginning of measure 11: “Wait, wait! Soon you shall rest!” On the one hand, what drama there is within “Wandrers Nachtlied” is of a muted sort, created with deliberate rhythms and subdued dynamics; on the other hand, Schubert manages to summon a coming-to-awareness—the subject, listening amidst the silence of the birds, becoming suddenly possessed of knowledge with profound existential implications—that enthralled creative artists in the early nineteenth century and that continued to fascinate well beyond.

Music and Words in Early Settings of Goethe’s “Über allen Gipfeln ist Ruh”

Although Zelter, Loewe, and Schubert each responded to Goethe’s “Über allen Gipfeln ist Ruh” in his own way, the similarities between their responses are striking. Each composer produced a quite brief song with a minimum of piano commentary; each placed the poem in a musical environment that suggests calm reflection; each set the word “Hauch” (“breath”) at the end of the fifth line of the poem with 5 over a dominant chord; and each set the last line’s “warte nur” with a descending third in a long-short-long rhythmic pattern. Given that the songs were all written within a ten-year period and that the composers all drew from the same tradition of compositional strategies, such similarities are not surprising. Perhaps more noteworthy is that, despite their similarities, these songs lead to markedly different interpretations of Goethe’s poem, interpretations that reflect the way music’s sonic analogs shape our understanding.

In previous work I explored the way music shapes our understanding of a text from the perspective provided by research on conceptual blending (Zbikowski 1999; 2002, chap. 6; 2006, 122–26; 2008b, 512–19; 2009, 370–76), which was discussed in chapter 2 in connection with analogy and the imagination. One of the things I noted in that discussion was the difference in kind between linguistic and musical concepts, and the challenge this raises to the assumption that all of the mental spaces in a conceptual integration network (CIN) share a basic topology. In brief, words are very good at picking out objects and events, characterizing relationships between them, and setting up rich narrative structures. Music, by contrast, has its basis in sonic analogs for dynamic processes that make it possible to create ongoing temporal sequences of incredible richness. Songs—especially those that are predicated on contrasting kinds of relationships between music and words—provide an ideal laboratory for examining tensions between the conceptual domains set up by these two different modes of communication, and for developing a better understanding of the nature of musical grammar.
Following the basic methodology introduced in chapter 2, my discussion of the conceptual blends set up by Zelter’s, Loewe’s, and Schubert’s settings of Goethe’s poem will be illustrated by diagrams of CINs that comprise four mental spaces: two input spaces (one for music, the other for words); a generic space (which aims to capture concepts common to all of the mental spaces in a given CIN); and a blended space populated by concepts drawn from the input spaces. One of my working assumptions here—and I should note that it would not be equally appropriate for every combination of music and words—is that music and words make an approximately equal contribution to the construction of the conceptual blend (and are thus an instance of what researchers on conceptual blending call double-scope networks; see Fauconnier and Turner 2002, 131–35). Inasmuch as each of the songs considered here sets the same text, I shall assume that the concepts in the input space for words in each of the three CINs are, for practical purposes, invariant. This is not to say, however, that the interpretation of these concepts is invariant, and changes in the interpretation of Goethe’s poem will be reflected in differences among the generic spaces of the three CINs. As noted in chapter 2, the emergent structure that is unique to the blended space of each CIN reflects three constitutive and governing principles of conceptual blending: composition (which puts together elements from the input spaces to create new entities in the blended space), completion (which extends the images suggested by the initial mapping from the input spaces), and elaboration (which develops the structure of the blended space by building upon the network of analogical mappings basic to the blend). The conceptual blends that result from each combination of music and words, and the entailments to which they give rise, are thus song-specific: our understanding of Goethe’s poem is transformed by each musical setting, and our understanding of the constituent sonic analogs of these settings is transformed through being coupled with Goethe’s words.

At the risk of stating the obvious, I should note that the conceptual blends I describe rely, at a minimum, on two things: the listener’s comprehension of the main images from Goethe’s poem, and her attention to the sequence of musical events set out by each composer. It is, of course, perfectly possible that one or the other of these desiderata may not be met, in which case there will be no new conceptual knowledge created by the coincidence of music and words. I should further note that my aim here is to suggest ways each composer’s music constrains our interpretation of Goethe’s poem: what I offer is not the last word on what mental images might be prompted by these songs, but instead a general indication of the way the images prompted by one song are different from another.

“Ruhe.” As I noted in my earlier discussion, Zelter’s setting creates a sonic analog for a dynamic process with two distinct parts, which are summarized in the input space for music in the CIN diagrammed in example 6.8. The first part moves forward in carefully measured increments, arriving on the dominant as a temporary goal; the second part presents the sonic image of a state of repose with a somewhat tenuous connection to the first part. The correlation of this sonic analog with some of the essential concepts activated by Goethe’s poem—the still, quiet, natural scene that the wanderer has discovered, one dominated by forested mountain tops in which even the birds are silent, and suffused by the notion that the state of rest
represented by this landscape is one with broader ramifications—gives rise to the abstract core idea (summarized in the generic space) of rest as a state of physical, emotional, and mental repose consequent to strenuous exertion. This abstract idea then guides the combination of concepts from the input spaces, such that the idea of the journey undertaken by the wanderer to achieve this landscape has as its analog the carefully measured course traced in the first portion of the song. Subsequent to the breath of “Hauch”—the literal breath invoked by Goethe’s poem, the singer’s breath after her extended second phrase, and (possibly) the listener’s vicarious breath at the same moment—the idea of rest is given an analog through the active stasis with which the second portion of the song is occupied. Completing this image, we can imagine the wanderer in a position of temporary repose—seated on a handy outcropping, or leaning against a tree—as he surveys the scene, which (being a poetic sort) he takes in to himself, finding rest for his spirit to correspond with rest for his body. As an elaboration of the image, we might imagine the wanderer’s journey concluding with sleep—certainly a mundane goal, but one that has the potential to activate the double entendre of “ruhest du auch” (“sleep” being one way to bring to human scale the unfathomable idea of eternal rest).
Again, what is important here is not the definiteness of this interpretation—as I have already noted, a given sonic analog can be correlated with dynamic processes drawn from a variety of experiential domains—but the way Zelter’s music constrains our understanding of Goethe’s words. Given the sonic analogs created by Zelter’s setting it is difficult not to hear measures 10–14 of the song as anything other than a sonic picture of easy repose, a state distinct from that embodied by the cumulative effect of measures 1–9, with the consequence that lines 1–5 of Goethe’s poem inhabit an environment distinct from that inhabited by lines 6–8.

“Nachtlied von Goethe.” Loewe’s setting creates a sonic analog for a dynamic process markedly different from that which Zelter’s setting engenders; the distinctive features of this process, which are summarized on the CIN diagrammed in example 6.9, are the tension between the aspirations of the voice and its obdurate accompaniment, and the overall sense that progress is toward a goal that is ultimately elusive. The correlation of this sonic analog with the concepts activated by Goethe’s poem gives rise to a generic space organized around the notion of an attempt to achieve a harmonious state that is thwarted by an internal deficiency. The contrast between the generic space prompted by Zelter’s setting and that

---

**Example 6.9 CIN for Loewe’s “Nachtlied von Goethe”**

[Diagram showing the CIN for Loewe’s “Nachtlied von Goethe” with labels for generic space, words space, music space, and conceptual blend.]
prompted by Loewe’s is worth noting: from the perspective provided by the former, the wanderer can remain an observer, a persona through whom the listener witnesses the scene; from that provided by the latter, the wanderer is a subject, a persona who experiences frustration at being thwarted in the attainment of a goal (and one with whom the listener may or may not be in sympathy). The structure offered by the generic space then guides the combination of concepts from the input spaces, creating the image of a wanderer unable to find a state of repose equal to that of the still beauty of nature, and who must wander on, accompanied only by the dull ache of a reward deferred. Completion of the conceptual blend might lead to an image of the wanderer as a melancholy individual whose emotional state is expressed through a downcast visage and sluggish body movement. Elaborating this blend, the listener might imagine that the continuation of the wanderer’s journey will, at the very least, be overcast with shadows, a gloom of both nature and spirit.

“Wandrers Nachtlied.” Schubert’s setting gives rise to a dynamic process of significant subtlety, the principal features of which are summarized in the CIN diagrammed in example 6.10. On the one hand, the process involves a rather

Example 6.10  CIN for Schubert’s “Wandrers Nachtlied”
long approach toward its apparent goal (length reckoned here in terms of the
duration of the approach within the compressed compass of the song); on the
other hand, this goal seems to be one that is not so much fully realized as it
is acted out—an almost Brechtian “Here is the way attaining this goal would
sound”—such that the achievement of the goal remains potential rather than
actual. The correlation of this sonic analog with concepts activated by Goethe’s
words gives rise to a generic space organized around the notion of a species of
split consciousness. In Goethe’s poem, this split is manifested in the contrast
between the observer and the observed, and between the outward world of
nature and the inward world of thoughts and emotions; in Schubert’s musical
setting, this split is manifested in the contrast between the interiority suggested
by measures 7–8, and the explicit conclusion provided by measures 9–11. These
elements are combined in the blended space such that the thought processes of
the wanderer have their analog in the course traced by Schubert’s music, pro-
cesses that encompass a sympathetic observation of the landscape (measures
3–4), a coming-to-awareness and growing interiority that activates an internal
self (measures 5–8), and—leaping across a chasm—a sudden and emphatic utter-
ance from the external self (measures 9–13). In completing this picture we might
well imagine an individual who, while remaining calm and perhaps hopeful in
outward appearance, is beset with inner turmoil. Further elaboration might give
rise to an understanding that this individual has a growing uneasiness about the
gap between his internal self—the self that is one with the hushed birds—and the
external self whose deeds and desires have taken him to this mountaintop, and
who will perhaps remain there for some time in silent meditation on the void
created by this gap.

**Music, Words, and Conceptual Blending.** In offering analyses of conceptual
blends produced by Zelter’s, Loewe’s, and Schubert’s settings of Goethe’s “Über
allen Gipfeln ist Ruh,” my aim has been to develop a more explicit account of
interactions between music and words. To be sure, these interactions are in part
determined by the way these settings shape the delivery of the words, regulating
the tempo at which they are uttered; providing musical commentary through
the piano’s introductions, interludes, and postludes; and privileging certain
words with repetition or emphasis. My interest, however, has been in the con-
cepts brought together in these particular conjunctions of music and words, the
character of which can be glimpsed in the generic spaces of the different CINs.
The collections of ideas captured in these generic spaces reflect the richness of
Goethe’s poem as well as the way its content is animated by the sonic analogs
for dynamic processes created by the compositional strategies deployed in each
setting. It should be emphasized, however, that my characterization of the ideas
proper to each generic space is meant to be quite general, not least because these
ideas will be realized in decidedly different ways in the mental spaces activated
by words and by music. In the CIN for Schubert’s “Wandrers Nachtlied,” for
instance, the idea of split consciousness is activated in the words space by the
contrast between an external world (a landscape in which no creature is stir-
ring) and an internal world (an intellect sensing the stillness of the scene, waiting
for an equivalent rest). In the music space, this idea is activated by the lack of
coordination between rhythmic processes and contrapuntal and harmonic processes that unfolds over the course of measures 6–11 of the song. The concepts with which words are concerned, then, involve elements and relations, whereas the concepts with which music is concerned involve dynamic processes. Perhaps most important, the immediacy and comprehensiveness of music’s sonic analogs result in blended concepts that are quite specific: both the effort expended by Zelter’s wanderer and the repose that is his reward have their correlates in the sonic analogs of his setting. Put another way, while a reader of Goethe’s poem might respond with a rich image that includes ideas about strenuous effort followed by easeful rest (and there is empirical evidence that readers animate texts in just this way; cf. Spivey, Richardson, and Gonzalez-Marquez 2005), a listener to “Ruhe” will have images like this (should they arise) shaped and constrained by Zelter’s musical setting.

To the extent that the conceptual blends of examples 6.8–10 are convincing—and much depends on a close attention to the sequence of musical materials that inform each of these blends—the different songs that result from Zelter’s, Loewe’s, and Schubert’s setting of Goethe’s poem reflect the way music shapes conceptual knowledge. According to the approach I have adopted here, the differences among the songs reflect not simply the way music colors or inflects Goethe’s words but the active participation of sonic analogs for dynamic processes in the construction of conceptual knowledge. This participation is perhaps most apparent in the elaboration of each conceptual blend: the respite found by Zelter’s wanderer is just that which eludes the grasp of Loewe’s wanderer; the anxiety of Schubert’s wanderer is something beyond the ken of Zelter’s wanderer, and perhaps too subtle for Loewe’s histrionic protagonist. Although each of these interpretations is, to some extent, latent in Goethe’s poem, they achieve definiteness only as that poem is realized in song.

Music, Words, and Song

Music, Poetry, and Song

In a careful consideration of Schubert’s setting of “Über allen Gipfeln ist Ruh,” Robert Hatten noted the violence Schubert’s music did to Goethe’s poetry, observing that “settings of lyric poems are always to some degree inadequate to the rich music of the poetry” (2008, 8). In the course of his essay Hatten described in detail the way Schubert’s setting disrupted the rhythms of Goethe’s lyric and cluttered the reverberant space created by the poet’s words. Given Hatten’s experience as a musician, analyst, and poet, his critique is worthy of close attention. That said, it was not entirely clear what Hatten meant by “the rich music of poetry,” an issue all the more important in that he seemed to want to go beyond simplistic if inspired metaphors like the one William Hazlitt offered in an 1817 essay on Coleridge: “Poetry is the music of language, expressing the music of the mind” (Hazlitt 1933, 136).
Insight into the music of poetry was offered by the poet and literary critic John Hollander who, in the course of reflections on relationships between music and poetry, remarked on the various rhythms of language that provided poets with an expressive palette of considerable scope: "Syllable sequence, stress pattern, word boundary, modulations of stress pattern because of phrase grouping, additional prominence to syllables being given by internal rhyme, assonance, or alliteration—and all before we have considered anything semantic at all" (1985, 33). This palette offers poets the means to create a rich array of sonic analogs that interact with the schematic structures associated with metrical patterns (such as iambic pentameter) and rhyme schemes (like those of a Petrarchan sonnet) to produce sonic structures not unlike those of music.

In that the focus of my work here and elsewhere has been on relationships between the basic cognitive and material resources that music and language recruit for the construction of meaning, I have not discussed in any detail the communicative potential of the exceptional genre of language known as poetry or ways this potential interacts with that of music. My approach has been to some extent pragmatic, but it also reflects a belief that poetry is transformed in essential ways when it is set to music—as I noted earlier, a poem becomes less a poem and more a text once it is in the hands of a composer. That said, analyses of songs that disregard the relationship between the sonic resources of poetry and those of music have two liabilities. First, and as shown by Yonatan Malin, much of the Lieder repertoire demonstrates an intimate interplay between poetic and musical rhythm that bears witness to a close consideration of poetic design on the part of composers (Malin 2010, chap. 1). It seems quite evident that Lieder composers—many of whom also set the decidedly utilitarian verse used in operas—were quite aware of the additional expressive resources engaged by good poetry. Second, the reception of a song by performers and listeners may well be influenced by the sound and sense of its poetry. To discount the play of sonic analogs between poem and music is to miss a potentially important part of how a song constructs meaning.

I should want to note, however, that this last consideration has at times obscured the goal of song analysis. As I observed in my initial discussion of Schubert’s “Wandrers Nachtlied,” measuring the prestige of a song according to the degree to which it meets or exceeds the artistry of the poetry it sets has the effect of pitting one expressive medium against the other. The forced choice that results—and efforts to avoid it—can be seen in models for song analysis developed during the latter half of the twentieth century, ones in which words were almost entirely subjugated to music (Langer 1953, chap. 10), in which music and words were engaged in a struggle to the death (Kramer 1984, chap. 5), in which the conjunction of the two yielded a hybrid (Agawu 1992), or in which each medium retained some measure of autonomy but combined to create an instance of multimedia (Cook 1998, 104–5, 136–42). The approach I have taken, based on the theoretical framework provided by research on conceptual blending, is relatively agnostic on the value of either poetry or music, emphasizing instead the contents of the mental spaces built up by each communicative medium. On the one hand, this approach has allowed me to draw attention to the different roles of music and
Music and Words

words in the construction of meaning; on the other hand, I have not as yet made room in my analyses for the distinctive ways poetry constructs meaning. That said, the notion of sonic analogs for dynamic processes that I have developed here provides a theoretical framework for the exploration of the rhythms of language and sound symbolism exploited by poets—the “music of poetry,” if you will—and for their correlation with the sonic analogs of music.

Song as a Delivery System for Words

Although the main focus of this chapter has been on the relationship between music and words that emerges through a close study of early nineteenth-century German Lieder, I would be remiss if I did not draw out what the study of such songs reveals about how words are changed by being sung. It is, after all, a simple fact that words are changed when they are set to music, their delivery shaped through the frameworks provided by specified temporal and tonal relationships, and interrupted and commented on by other musical voices. As Mark Booth observed in his consideration of the relationship between poetry and music, “The ways in which song words are subject to the pressure of their music are subtle and fascinating. They are reinforced, accented, blurred, belied, inspired to new meaning, in a continual interplay. In that interplay there is a constant tug against the resolution of the words to carry out their own business” (1981, 7–8). Thus simply setting words to music—whether those words originate in artful poetry, unaffected speech, or something in between—changes the possibilities for constructing meaning through language. W. B. Yeats, describing his disappointment with poems that are sung, remarked, “When I heard anything sung I did not hear the words, or if I did their natural pronunciation was altered and their natural music was altered, or it was drowned in another music which I did not understand. What was the good of writing a love-song if the singer pronounced love ‘lo-o-o-o-ve,’ or even if he said ‘love,’ but did not give it its exact place and weight in the rhythm?” (1961, 14). In brief, then, singing invariably distorts language, removing it from its day-to-day setting—or, in the case of poetry, its artfully constructed rhythms—and situating it within a new sonic context.

In general, singing also typically slows down the rate at which words are delivered. As the composer Martin Boykan noted, “A text is sung far more slowly than it is spoken, and even where the musical tempo is fast and we have the impression of speed, the words move at a rate that we would find intolerable in conversation” (2000, 133). This slowing can also be effected through text repetitions—which might involve a word or two, an entire line, or even a complete stanza—and by various kinds of musical interjections, whether those are of a practical cast (to allow the singer to catch her breath) or involve a more extended musical intervention (ranging from brief commentaries to more extended transitions between verses or sections).

Along with slowing down the delivery of words, the resources offered by music can also move words toward a communicative register often characterized as “expressive.” As I observed in my discussion of Loewe’s “Nachtlied von Goethe,”
Lieder composers of the nineteenth century had at their disposal a number of compositional strategies through which they could create the effect of “expressivity.” That said, simply singing words does not render them “expressive.” As an example, consider the popular song “Bye Bye Blackbird,” which I discussed in previous work (Zbikowski 2002, 229–41). For much of the twentieth century “Bye Bye Blackbird” was a fast tune that projected positive, forthright sentiments—there is relatively little that is markedly “expressive” in most of the recordings of the song done before midcentury. However, subsequent to a 1955 recording by Peggy Lee that reframed the tune as a melancholy reflection on loss (in part by more than halving the tempo at which it was performed) “Bye Bye Blackbird” became associated with more somber emotions and with nostalgia—that is, it became an “expressive” song. Almost all of the recordings of the song done since 1955 fall into a much more highly expressive register, a change accomplished (as was Lee’s) through tempo, performance inflections, and orchestration (extending to the decisions that arrangers make in harmonizing the tune). Thus the resources offered by music may render words “expressive,” but they may equally situate those words within any of a number of other communicative registers.

Finally, as I noted in my discussion of settings of Goethe’s “Über allen Gipfeln ist Ruh,” the sonic analogs of music can constrain the images that might be conjured from a sequence of words. Thus the way a listener might interpret the “warte nur” at the conclusion of Goethe’s poem will change from Zelter’s setting, to Loewe’s, to Schubert’s, reflecting the different kinds of dynamic processes that are summoned by their settings.

The cumulative effect of these changes—altering the pronunciation of words, slowing their delivery, introducing repetitions and pauses, and constraining the images to which they give rise—is to take words out of the familiar frameworks they occupy and move them toward a communicative register that borders onto ritual speech, a transformation reinforced by an emphasis on composed songs as fixed and unchanging works. On the one hand, a song is not, in itself, a ritual in any conventional sense; on the other hand, the replication of so many of the features of ritual speech in the way music and words combine may suggest reasons why songs are such enduring structures in human cultures.

Conclusion

In the introduction to this chapter, I noted that the topic of how music and words combine is a potentially vast one. The care with which many Lieder composers set poetry, the rapid stylistic changes that Lieder underwent in the first decades of the nineteenth century, and the attention that has been given to Lieder by music scholars make the genre an ideal place to begin a study of how music and words—alone and together—communicate conceptual knowledge. There is, however, certainly more that could be said about how music and words combine. As a first approximation of what this “more” might be, let me consider briefly three factors that can inform the way music and words combine.
The first factor concerns the number of performers involved. Unaccompanied song, for instance, has limited means for the reinforcement of the interacting rhythmic streams that establish a temporal framework, and its tonal framework may be far less anchored to harmonic structures. Additional performers—whether they be accompanists (as in the case of Lieder) or vocalists—can add to the means through which temporal and tonal frameworks are reinforced. These additions can also introduce complications, as when two singers sing different words (or the same words at different times) or when the accompaniment threatens to overwhelm the singer.

A second factor concerns the number of media involved. There is a tradition of Lieder performance, for example, in which the focus is almost entirely on the sounds produced by the performers: the singer moves as little as possible, and the accompanist indulges in few of the physical flourishes common in solo instrumental performance. In many performance traditions, however, physical movement while singing is part and parcel of the presentation of music and words, something demonstrated quiteconvincingly in recent work on South Asian vocal practice by Martin Clayton (2005), Laura Leante (2009), and Matt Rahaim (2012). Combinations of music and words can thus be further shaped through gesture, as well as through staging and choreography (as in the case of many music-theatrical works), or through being embedded in a carefully prescribed sequence of behaviors (as may occur in religious practices).

The third factor concerns the conformance between the conceptual knowledge communicated by the constituent media involved in a combination of music and words. In the case of the three songs I have considered here, for instance, the basic conformance between words and music was relatively high: Goethe’s poem is one that both summons and invites reflection, characteristics supported by each of the musical settings. There are, however, cases in which the concepts activated by a musical setting are at something of a variance with the concepts activated by the words. One of the most famous examples is Robert Schumann’s “Ich grolle nicht,” the seventh song from his Dichterliebe op. 48. Throughout Heinrich Heine’s poem, the words “Ich grolle nicht” (“I will not complain”) operate as a kind of refrain, one that could be interpreted as meek acquiescence in the face of unpardonable injury. Schumann’s music nonetheless makes clear that the protagonist of the song will complain and will do so repeatedly and loudly. While it is true that Heine’s words operate quite easily within a register of irony familiar to his readers, Schumann’s setting makes manifest not only the actuality of the supposedly deferred complaint but also the depth and bitterness of the betrayal that occasioned it. In other combinations of music and words, the lack of conformance between the concepts activated by the two media can be even more marked, and complicated further by correlations (or the lack thereof) with other communicative media.

As I observed in my introductory comments, words—as speech sounds that communicate linguistic meaning—are but one manifestation of language. When properly disposed, they nonetheless give clear evidence of the syntactic resources of language, through which speech sounds can be used to create complex narratives that range between present, past, and future and that involve multiple participants,
each of whom could change roles as circumstances require (Tomasello 2008, 244–
45, 282–90). Such resources make it possible for children as young as four years of
age to create rich imaginative worlds from only a few prompts and to share those
worlds with others (Berman and Slobin 1994; Strömqvist and Verhoeven 2004).
The syntactic resources of music, by contrast, make it possible to create sonic ana-
logs for dynamic processes of remarkable specificity that are grounded in and that
can shape embodied experience, yielding an ongoing present of unparalleled rich-
ness. When these two modes of communication are brought together—as they
are in children’s songs, religious chant, operatic arias, popular ditties, rap music,
and nineteenth-century Lieder—they provide an immediate and powerful way to
shape human interactions. This immediacy and power are perhaps most evident
in the performance of national anthems or football chants, but they are also appar-
ent in the lullaby that a father sings to his child or in a compelling performance of
a Lied. Although music and words offer different resources for communication,
together they make possible a mode of expression that is at the core of what it
means to be human.
CHAPTER Seven

Questions, Answers, Questions

As I noted toward the end of chapter 1, over the past twenty years there has been renewed interest in the origins of music, leading to speculation about the role of musical behavior in human evolution. One of the most recent accounts was Gary Tomlinson’s *A Million Years of Music* (2015), which endeavored to trace the emergence of the capacity for music from the technological capabilities of the earliest hominin species to modern humans. As does Tomlinson’s, most of the recent accounts of the origins of music take inspiration from Charles Darwin’s theory of evolution, but the practice of such speculations is almost as old as the theory itself (Kivy 1959). Darwin, for instance, made his own contribution in the opening chapter of Part III of his *The Descent of Man, and Selection in Relation to Sex* (first published in 1871), where an explanation of the origin of music formed part of a demonstration of the ways sexual selection (as distinct from natural selection) applied to the species Homo sapiens. Darwin’s focus in this part of the treatise was on what he called “secondary sexual characters,” which typically emerged later in development and differed between males and females. He began the chapter by noting the relative average stature and hairiness of each of the sexes as well as differences in their mental abilities (which he regarded as an established fact), and he then turned to the matter of the voice and music. Darwin’s inclination was to view the vocal powers of the species as related to sexual selection, and yet the recruitment of the voice for uses having nothing to do with selection gave him pause: “The capacity and love for singing or music, though not a sexual character in man, must not here be passed over” (1874, 566). Darwin then considered various manifestations of musical behavior in other species (including gibbons and mice) and concluded that the mere perception of musical notes was not something specific to Homo sapiens. This then left the riddle of human music-making: he observed, “As neither the enjoyment nor the capacity of producing musical notes are faculties of the least use to man in reference to his daily habits of life, they must be ranked among the most mysterious with which he is endowed” (1874, 569–70).

Not one to be daunted by profound mysteries, Darwin went on to consider widespread evidence for music-making in the world’s cultures, the apparent antiquity of music, and music’s ability to arouse emotions (especially of an amorous sort). He also noted that expressive speech had much in common with music, especially in its rhythmic features. After considering the evidence available to him, Darwin concluded,

All these facts with respect to music and impassioned speech become intelligible to a certain extent, if we may assume that musical tones and rhythm were used
by our half-human ancestors, during the season of courtship, when animals of all kinds are excited not only by love, but by the strong passions of jealousy, rivalry, and triumph. From the deeply-laid principle of inherited associations, musical tones in this case would be likely to call up vaguely and indefinitely the strong emotions of a long-past age. As we have every reason to suppose that articulate speech is one of the latest, as it certainly is the highest, of the arts acquired by man, and as the instinctive power of producing musical notes and rhythms is developed low down in the animal series, it would be altogether opposed to the principle of evolution, if we were to admit that man’s musical capacity has been developed from the tones used in impassioned speech. We must suppose that the rhythms and cadences of oratory are derived from previously developed musical powers.

(1874, 572)

A bit over a hundred years after Rousseau’s *Essay on the Origin of Languages*, then, Darwin—building an argument based on contemporary work in the natural sciences rather than on speculative anthropology—reached much the same conclusion as had the Enlightenment philosophes: music had its beginnings in the use of prelinguistic sounds to establish a direct and immediate species of communication between humans.

Although Darwin’s argument is certainly more closely reasoned than was Rousseau’s, it is nonetheless marred by dubious assumptions about the universality of music. There is, of course, clear evidence that a wide range of species can make use of patterned sound in a number of different ways, including those that would contribute to sexual selection. But the use of patterned sound—or even the capacity to respond to patterned sound produced by another species—does not, in my view, demonstrate a capacity for music. What is missing, and what is evident throughout human musical practice, is the use of patterned sound to facilitate and shape cultural interactions (a point also made, in a slightly different way, in G. Tomlinson 2015, chap. 7).

One thing that may have obscured Darwin’s view of the role of music in animal and human behavior is the perspective on music that developed during the nineteenth century, and that remains common today. This perspective was a consequence of historical inquiries that brought to light the resounding success of musical expression across a number of civilizations beginning in the second millennium CE, yielding artful and highly complex forms of communication that were, to some extent, independent from language. To speak of “music,” then, invariably invoked human cultural products produced by highly trained artisans—if not artists—who operated at the uppermost level of human accomplishment. In the course of the nineteenth century, and especially among intellectuals in Western Europe, the sedimented residues of the practices of these artisans were gradually detached from their social and cultural contexts to become, along with select contemporary works, monuments to art in its purest form. J. S. Bach’s *St. Matthew Passion*, for instance, was viewed by Felix Mendelssohn (under the guidance of his teacher Carl Friedrich Zelter) as belonging to a genre of church music that exemplified the artistic accomplishments of German composers (Garratt 2004). The original liturgical context of the oratorio—which was, after all, the principal motivation for Bach’s labors on
the composition (Melamed 2005, 7–11)—was largely irrelevant to later generations. (That said, the care Bach lavished on the original autograph suggests that he regarded the work as different in kind from his other ecclesiastical compositions; for a discussion, see Gardiner 2013, 398–99.) To be sure, the historical picture is more complex than this and is tied up with aesthetic theories that contributed to the idea of absolute music and produced what Lydia Goehr called the “work concept” (Dahlhaus 1989; Goehr 1992, chap. 6); even so, the notion that musical behavior should lead toward timeless works of art became deeply embedded in nineteenth-century constructions of musical practice. From this perspective, the drive to create music—which, by definition, had no outward utility—was indeed something of an evolutionary puzzle.

The approach taken in this volume, by contrast, is one with a decidedly utilitarian view of music—namely, that music offers a way to represent in sound various dynamic processes that are important in human cultural interactions. The mystery confronted by Darwin—why humans developed something as apparently useless as music—is thus resolved through the relatively straightforward proposal that music is useful to humans. Of course, this proposal then leads, if not to other mysteries, at least to a host of other questions: What sorts of dynamic processes are important in human cultural interactions? How is it that a sequence of sounds can represent a dynamic process? How are sounds organized to accomplish this task? How does music relate to, or shape, or influence other communicative media? And, finally, how does such a utilitarian view of music connect with the idea of music as an abstract art that so enthralled nineteenth-century intellectuals?

The answers I have developed to these questions—answers that are in many cases preliminary and provisional—take as their point of departure the notion that communication through the patterned nonlinguistic sound of music reflects cognitive abilities that can be applied to a range of other thinking and reasoning tasks. The ability to correlate a sequence of sounds with a dynamic process that may or may not have a sonic component, for instance, demonstrates humans’ capacity for drawing analogies, a capacity that contributes much to the distinctiveness of human intelligence.

The preceding chapters have aimed to develop this insight into an account of how humans’ cognitive abilities—understood here to be inextricably intertwined with the unique and often ephemeral social and cultural structures produced by humans—provide a basis for a grammar of music. That is, musical communication does not simply reflect the cognitive abilities of humans, it is in fact predicated on those abilities. This is an approach I have borrowed from cognitive linguists, along with the idea that it can be implemented by means of a construction grammar in which all grammatical elements, from the smallest to the largest, are combinations of form and function. To understand the grammar of music, then, one first has to understand the roles music plays in human cultural interactions.

One way to develop an understanding of the fundamental function of music in human cultures is to contrast it with the function of language. Drawing on the work of Michael Tomasello and others, I proposed that one of the basic uses to which language is put is to manipulate the attention of another person within the
context of a joint attentional scene. Music is not very good at manipulating the attention of another person in a specific way—what tune would I play to let you know that you were blocking my light and making it hard for me to read?—but it offers an ideal way to represent dynamic processes through sequences of patterned sound. I further proposed that two different kinds of reference support these different functions: language makes almost exclusive use of symbolic reference, in which symbolic tokens—which could be built from sounds, or codified shapes and movements of the hands, or marks in some durable medium—are systematically correlated with various referents and with other symbolic tokens. Music, for its part, makes almost exclusive use of analogical reference, which requires that a token share structural features with the entity or phenomenon to which it refers. Although there has long been an assumption that forms of reference based on a similarity between signifier and signified are simple, direct, and require relatively limited cognitive capacities, recent research on analogy has cast this into doubt: on the best evidence we have, human beings are the only species that have a robust capacity for analogy and thus the only species to regularly connect tokens with events or phenomena through shared structural features. Two aspects of analogical reference are especially relevant for the development of a cognitive grammar of music. First, music is not the only communicative medium to exploit analogical reference: both gesture and dance make use of analogical reference as will, on occasion, linguistic communication (through onomatopoeia and prosody). Second (and building on Barsalou’s perceptual symbol systems theory), analogical reference can, in many instances, be grounded in embodied experience: the compositional strategy of cadence, for instance, provides a sonic analog for the process of arriving at a goal (no matter how temporary), a process we reenact each time we sit down in a chair or roll into our bed.

My proposal, then, has been that the basic function of music in human cultures is to represent in sound the dynamic processes that are important in those cultures, and to do so using a species of reference that exploits humans’ capacity to draw analogical relationships between disparate domains. The function of music is to provide sonic analogs for dynamic processes; the form these utterances take is shaped by the communicative resources offered by analogical reference as it is implemented through sound sequences (whether those sequences be real or imagined). This perspective leads to two lines of inquiry fundamental to the development of a cognitive grammar of music. The first concerns the way sounds can be used to represent dynamic processes such that those representations can play a part in human cultural interactions. The second concerns the means through which sounds are organized to accomplish these representations.

Sonic Analogs for Dynamic Processes

In chapter 2 I defined a dynamic process as a coherent sequence of phenomena that is distributed over time and typified by parametric modulation or change. This definition then served as a basic guide for the exploration of three types of
dynamic processes that are important in human cultural interactions and that have long associations with music: emotions, gestures, and dance.

As noted in chapter 3, connections between music and emotion go back to antiquity—indeed, for many commentators music is more closely linked to human emotional responses than any other communicative medium. That this should be so is somewhat curious: broadly speaking, emotions are physical and psychological responses to an event either internal or external to an organism; sequences of musical sound represent only one of the ways human communication can prompt an emotional response. I suggested that the close connection often drawn between music and emotion likely has to do with the insubstantial character of sonic analogs, which makes the connection between music and emotions appear to be immediate: both seem to have their proper home within our internal experience, rather than in the external world. On the account I offered, however, this sort of immediacy is an illusion. Music instead prompts the listener to simulate (in the sense developed by Barsalou) an emotional response through sonic analogs for the physical and psychological processes associated with that response. Because such simulations are grounded in embodied experience, we might well have the sense that the emotional response that is evoked is not mediated in any way. In fact, such simulations are always mediated: while we might characterize slow music in a minor key as “sad” (or, more accurately, as associated with feelings of sadness) we might equally characterize it as lacking in energy, or as summoning colors at the darker end of the spectrum. To be sure, it is not difficult to get from “lethargic” or “purple-hued” to “sad,” but as with all analogical correlations our choice of one interpretive frame or the other will be guided by context and our motivation for making the analogy in the first place. The fact that emotional valence might have priority in the critical reception of musical practice has perhaps more to do with post-Enlightenment listening practices than with the intrinsic properties of music: as James Johnson has observed, the circumstances of listening to music—and especially to instrumental music—changed in significant ways in the early nineteenth century, giving rise to a range of behaviors that both encouraged and gave priority to the simulation of emotions in response to music (J. H. Johnson 1995, chap. 13).

It may well seem that the connection between music and gesture is somewhat more contrived than that which obtains between music and the emotions, “gesture” being a conventional, if not particularly technical, way to characterize a continuous sequence of musical events that occupies a relatively compact span. There is good evidence, however, that the use of spontaneous gestures to accompany and to supplement speech is an innate human behavior, having been documented in congenitally blind children (Iverson and Goldin-Meadow 1997, 1998, 2001). Conceiving of music as gesture, then, correlates the sonic analogs of music with expressive unpremeditated bodily movements, such that sequences of musical sounds become connected with an aspect of embodied experience essential to how we convey our thoughts. More broadly, the rough equivalence of musical phrases and physical gestures fundamental to this conception reflects their mutual exploitation of analogical reference for the representation of dynamic processes, as well as the reliance of both on the infrastructure of common conceptual ground.
and shared intentionality that supports human communication. Placing music and gesture in the same sphere also lends credence to the idea that each, on its own, is a means of expressive communication, an idea I characterized with the notion of music and gesture. The close coordination of music and movement can, on the one hand, offer opportunities for one medium to shape the interpretation of the other, a situation evident in the barbershop scene from Charlie Chaplin’s *The Great Dictator*. On the other hand, both music and gesture may contribute, in their different ways, to telling the same story, which was often how the two expressive modes were employed in French and Italian opera of the first half of the nineteenth century (Smart 2004). In such cases music and gesture may each provide an analogical representation of a dynamic process drawn from some other domain, a configuration of media that I pointed out in my discussion of French noble dance in chapter 5.

Where the affiliation between music and gesture may at first seem tenuous, the relationship between music and dance runs very deep—in many cultures in Africa and Oceania, for instance, music and dance are so closely bound together as to be functionally inseparable (Hampton 1998; Kaeppler 1998). Dance as it is typically practiced is invariably social, part of a set of cultural practices that involve a number of individuals whose interactions are coordinated through patterned bodily movement. In consequence—and unlike the spontaneous gestures that accompany speech—the movements of social dance tend to be highly structured. This structuring of movement has two important consequences: first, dance becomes a means of enacting social structure; second, a given dance practice may exploit music’s potential to provide sonic analogs for the steps of a dance, such that the structure of the dance and the structure of the music are matched. In such cases the coordination of patterned bodily movement with sonic analogs for dynamic processes has the potential to shape—and indeed perform—the habitus of a given social milieu. And when music and dance are very closely linked, as they are in the Viennese waltz, simply hearing the music may prompt a simulation of the dance: as La Garde-Chambonas observed of the dancers during the Congress of Vienna, “As soon as the first measures are heard, faces light up, eyes come alive, and a quiver of pleasure is communicated from each person to his or her neighbor” (1901, 36). That this should be so is no accident: the music for waltzes and other social dances was crafted to create sonic analogs for the dynamic processes associated with the steps of the dance, conforming to design principles that ensured the production of danceable dance music.

While it is possible to draw distinctions between the cultural interactions related to the emotions, to gesture, and to dance, there is also a certain amount of overlap between these interactions. Getting a group of people together and having them participate in a dance is, after all, a way to change their emotional disposition, if for but a short period of time; gestures can be one means of expressing emotions, and they may also be motivated by emotional responses. In consequence, the interpretation of a given sonic analog as pertaining to the emotions, to gesture, or to dance will to some extent be informed by context, but it will also reflect the structural attributes exploited by the analog. The potential for a sonic analog to be correlated with a dynamic process is related to the way that analog is
constructed and deployed, and these in turn are conditioned by the resources of musical grammar.

The Elements of Musical Grammar

The approach to musical grammar that I have offered in this volume adopts the position that musical utterances begin with sequences of sound of limited scope that contribute to—or in some cases instantiate—the analogical representation of a dynamic process. The scope of such sound sequences, which constitute the basic elements of musical grammar, is constrained by human cognitive processes concerned with temporal processing, in particular those associated with working memory. I construe these basic elements as analogous to the basic-level categories through which humans organize their understanding of the world. This construal provides, through the notion of categories that show typicality effects, a way to describe relationships between grammatical elements that are similar but not identical to one another, and it connects the basic elements of musical expression with conceptual knowledge.

The basic elements of musical grammar are organized into larger structures through syntactic processes that, in keeping with the fundamental tenets of construction grammars, are themselves sonic analogs for dynamic processes. It appears that many of the most common syntactic processes used to organize musical materials exploit a basic notion of causality: paths typically lead to a goal; instability prompts a search for ways to achieve stability; departures often motivate return; and differentiated topographies suggest possibilities for motion. Causality of this sort bears witness to the shaping influence of embodied experience on the simulations prompted by sonic analogs: embarking on a path (even if it is only toward the refrigerator) is typically predicated on achieving some goal (such as finding something—anything—to eat in the refrigerator). There are, of course, journeys that have no explicit goal (which might be one way to characterize certain kinds of meditative practices) and thus no clear causality, and I would submit that the dynamic processes associated with such journeys could serve equally well as the basis for a syntactic process.

Drawing an analogy involves correlating elements and relations between two different domains. The application of syntactic processes to the basic elements of musical grammar will, in consequence, be constrained by the attributes of those elements. Representing a process of deceleration through pitch alone, for instance, will tend to be somewhat challenging inasmuch as the disposition of pitches relative to one another seems to offer few structural relationships with which to represent temporal relationships—this despite the fact that linguistic characterizations of temporal relationships invariably involve spatial terms (Núñez and Cooperrider 2013). Representing deceleration through a gradually increasing temporal interval between the onset of sounds, on the other hand, seems relatively straightforward. (I should note that I regard the latter case as still involving an analogical process: the “slowing down” we hear in a decelerando
involves a slowing down of the delivery of sound events, manifested through a gradual increase in the temporal interval between those events.) By contrast, the use of fixed pitches, which allow a musician to return to the “same” pitch after some interval of time, makes it possible to establish a virtual topography for imagined motion. A virtual topography established through patterns in the temporal onset of sounds is, of course, also possible—it is a regular feature of the music for West African drum ensembles (Cudjoe 1953; Agawu 1995, chap. 4)—but such topographies place rather greater demands on memory resources than do the retention of a limited group of pitches.

The resources for analogical mappings offered by various aspects of musical sound make it useful to think of musical organization in terms of syntactic layers, with each layer exploiting a particularly distinctive feature of sound. Thinking along such lines must, however, be reckoned to be in some respect “analytical”: if, for instance, one is asked to recall the melody of the opening three measures of the first movement of Mozart’s Symphony in G Minor (K. 550), the result is typically a holistic structure. Attending to specific attributes of this melody—for example, the kinesthetic features summoned by John Sloboda’s “pushing on the stuck door,” mentioned in chapter 1—begins to break that structure down into its constituent features. A similar and perhaps equally prereflective analytical process is evident in our evaluation of different instantiations of a given musical construct. Noting that Mozart’s melody takes on diverse forms over the course of the first movement of K. 550 means being able to attend, on some level, to the various features that allow us to distinguish one version from another (a capacity for discrimination associated with processes of categorization, described in more detail in Zbikowski 2002, chaps. 1 and 4). I would submit that a similar kind of attentive discrimination can be directed to the syntactic layers evident in many musical utterances, and that it may be prompted by instances when the syntactic processes in different layers are not aligned with one another.

Among the most significant resources that make possible intentional communication through sonic analogs are the perceptual anchors provided by referential frameworks. In the course of this volume I have focused on two such frameworks: one associated with temporal events and the other with fixed pitch. Temporal frameworks emerge from the interaction of interlocked cycles of periodic events. Such patterns can be incredibly simple—consider, for instance, the sequence of temporal events basic to the waltz motive, the constituent features of which could be specified in this way:

- a repeated pattern of six events, equally spaced about a half-second apart;
- the first and fourth events should be functionally equivalent to one another, and distinguished from the second, third, fifth, and sixth events;
- the second, third, fifth, and sixth events should be functionally equivalent to one another, but need not be in as close a conformance as are the first and fourth.

Before proceeding, let me note three things about this pattern. First, I have characterized it in such a way that it need not be realized through sound—an “event”
could be the initiation of a musical sound, but it could also be created through the initiation of a physical motion, or through visual or tactile stimuli. Second, it is through repetition that such patterns become a resource for entrainment and create a framework relative to which subsequent events can be evaluated (and thus a perceptual anchor). Third, when repeated patterns of this sort are rendered in sound and used to structure social and cultural interactions, they provide a basis for the intentional communication that typifies musical behavior. It is within this last context, then, that a sequence of temporal events such as that outlined above can serve as a temporal framework through which the sonic analogs of music are organized, something evident in the role that the rhythmic features of the waltz motive played in the works by Lanner and Schubert discussed in chapter 5.

In contrast to the perceptual anchor offered by temporal frameworks, those afforded by tonal frameworks are oftentimes less immediate, relying not on processes of entrainment but on the capacity to hear and retain pitches and pitch relationships over time. Any number of syntactic processes exploit such frameworks, not least because they offer musicians a coherent analog for physical space: pitches can represent points in space (in a concrete, rather than geometric, sense); intervals between pitches can be construed as relationships between points in space or as physical transformations that take a subject from one pitch to another—in David Lewin’s memorable formulation, “If I am at s and wish to get to t, what characteristic gesture . . . should I perform in order to arrive there?” (1987, 159); and one can, as I have suggested in some of the analyses offered here, trace complex trajectories through the topography of this virtual space. Of equal importance—and to some extent assumed by the approach to the basic elements of musical grammar I adopt—tonal frameworks offer a means to compare and evaluate musical concepts. The example I offered in chapter 4 was of the relationship between the descending fifth from the opening of Jerome Kern’s “The Way You Look Tonight” (which sets “someday”) and the descending octave that marks the culmination of the melodic process that unfolds over the first eight measures (which sets “of you”): both intervals conclude on D4, which allows a listener to make a connection between two prominent elements of a melody that, at standard performance tempi, are some twelve seconds apart, a span that at the least stretches and most likely exceeds the bounds of working memory.

As I noted in chapter 4, the perceptual anchors provided by referential frameworks support the retention of syntactic elements in long-term memory. As such, they can invite an objectification of the materials of music: two measures in clear triple time, with the first beat of each measure firmly marked, could become the basis for “the waltz motive”; a succession of regular departures from and returns to D4 could (depending on the pitch materials involved) be formalized as a diatonic collection around D4, later as a scale organized around D4, still later as a D major (or minor, or Hypodorian) scale. The objectification of musical materials offers an incredible resource for musical thought and also a hazard. The resource is evident across a wide range of musical practices, enabling the regard of sequences of musical events as though they were objects, to be turned this way and that, and reshaped into new configurations. The hazard is that the sonic
analogs upon which such constructs are based—which offer ready connections to embodied experience as well as to other conceptual domains—can become pale shadows of their former selves, curiosities discarded on the way to true, invariant knowledge.

Grammars are often conceived of in prescriptive terms: a set of rules that allow humans to construct correct utterances. In this volume I have tended to conceive of grammar as a way to account for, rather than precondition, utterances, adopting a perspective applied to language by Ronald Langacker: “Putting together novel expressions is something that speakers do, not grammars” (1987, 65). That said, I have suggested two ways musical utterances may fail to realize their function. First, the syntactic coordination of musical materials may fail, as was evident in example 4.11 when I flipped the order of two eight-measure units from Brahms’s Hungarian Dance No. 5, a change that thwarted the forward-thrusting dynamic shape of the original. Second, the social function of a genre may change. As outlined in chapter 6, in the late eighteenth and early nineteenth centuries the Lied was generally conceived of as a means of promoting the appreciation of poetry and the social values of singing. Accordingly, in a properly composed Lied (such as Zelter’s “Ruhe,” or any of those of Hans Georg Nägeli) the expectation was that music would support, or perhaps comment on, the poem that was set, but it would in no way compete for the listener’s attention. Beginning in the early nineteenth century, however, there arose a new approach to the Lied, which transformed it into a dramatic genre through the appropriation of compositional strategies more typically used for the ballad. This led to changes in the way Lieder were composed—in the terms I have developed here, the grammar of the genre changed to reflect this new conception of its function—yielding compositions (such as Loewe’s and Schubert’s settings of Goethe’s “Über allen Gipfeln ist Ruh”) in which music became an equal partner with poetry in shaping the expressive resources of the Lied.

Prospects for a Cognitive Grammar of Music

There has, to my knowledge, been no previous attempt to outline or otherwise speculate on how one might develop a cognitive grammar of music, although the notion of a construction grammar has recently been applied to music by other researchers (Gjerdingen and Bourne 2015). What I have offered in this volume is a rationale for a cognitive grammar of music, along with a first approximation of some of its key features. This research program suggests a number of avenues for exploration, including empirical research, a further exploration of how such a grammar is instantiated in musical practice, and a more complete consideration of the role music plays in human cultures.

Empirical Research

My point of departure has been current knowledge of human cognitive capacities as these might apply to the understanding of music. My assumption is that
humans’ capacity for music is not a single thing but instead involves a complex mosaic of skills (and, in this respect, is similar to humans’ capacity for language; see Hurford 2003, and Tomasello and Call 2007, 235). One of the skills that I believe is central to musical understanding is the ability to categorize sound sequences of relatively limited scope such that these sequences can be retained in memory, used to guide present and future actions (understanding “action” here to include the simulation of dynamic processes analogically related to these sound sequences), and linked to other categories—features all related to the way I construe conceptual knowledge. Although preliminary empirical research into musical categorization is promising (Ziv and Eitan 2007; Eitan and Granot 2009), much remains to be done to understand how processes of categorization shape musical comprehension and if the results of such processes really constitute anything like conceptual knowledge.

A related issue concerns humans’ memory for sound sequences. I have proposed that the basic elements of musical grammar are limited by the capacity of working memory for such sound sequences. There has not as yet been substantial research on the relationship between working memory and sound sequences of the sort that are typical of music, and so the notion that working memory can accommodate such sequences must remain speculative. There is, by contrast, substantial research on long-term memory for music but for the most part it has focused on the recollection of identifiable musical works (especially popular songs) rather than on the storage of syntactic structures of the sort that would be important for a cognitive grammar of music.

With respect to music and analogy, there is of course a long tradition of practical music-making that uses sounds produced by musical instruments to imitate sounds that either occur in nature (such as those associated with a galloping horse) or that are produced through human agency (such as by hammer blows). Although the success of such imitations is generally accepted among musicians and scholars of music, empirical studies have yet to confirm that listeners can reliably connect musical sound X with natural sound Y in the absence of contextual knowledge. Indeed, studies of metaphors used to describe musical relationships suggest that such connections may be hard to come by: untutored listeners describe musical sounds using a wide range of concepts, and these concepts do not always converge on a readily apparent core of musical attributes (Eitan and Granot 2006; Eitan and Timmers 2006). More generally, although there is a recent study of ways music can be used to convey movement (Hedger, Nusbaum, and Hoeckner 2013), there is an absence of studies of the way dynamic processes are represented analogically (although there is a brief mention of the usefulness of animating analogical diagrams in Pedone, Hummel, and Holyoak 2001, 220). In consequence, conceiving of relationships between music and the emotions, music and gesture, and music and dance in terms of analogical processes is for the most part untested ground. If such processes are indeed important for musical grammar, the empirical study of such relationships would be enormously helpful.

While there is little doubt that the capacity for analogy contributes significantly to the distinctive character of human intelligence, the notion of analogical reference that I have proposed—which obtains when a symbolic token shares structural features with some other entity or phenomenon—has not been subject
Foundations of Musical Grammar

to a great amount of empirical scrutiny. Although the notion of such tokens is intuitively appealing—what could be easier than drawing a wavy line to represent a serpentine route?—as shown by Umberto Eco’s extensive discussion of iconism and related topics in the theory of sign production he develops in A Theory of Semiotics (1976b, chap. 3), the processes involved with analogical reference are far from simple, and we have much to learn about how it operates.

Sonic Analogs, Musical Constructs, and Musical Practice

As I noted in chapter 1, the idea that musical expression would involve sonic analogs for dynamic processes is a far from novel one. Accounts of such analogs are, nonetheless, not common in technical descriptions of musical organization, which tend to be focused on musical constructs of the sort captured with such economy by language. Indeed, many of my musical analyses have made recourse to such constructs, which offer an efficient and concise way to characterize some of the important features of sequences of musical sound. One of the challenges to the development of a cognitive grammar of music is understanding how some of the more familiar constructs of music theory and analysis relate to the sonic analogs for dynamic processes that are fundamental to musical grammar. A fuller understanding of this relationship could lead to an improved understanding of the cognitive capacities that shape musical thought, and the way the resources of language shape our conception of musical phenomena.

A somewhat larger challenge is presented by musical practices that redeploy the elements of musical grammar for expressive purposes. For example, some years ago Wendy Allanbrook noted that Joseph Haydn employed one of the basic rhythmic figures used to accompany the pas de bourrée in the Finale of his String Quartet in B♭ major Op. 76 No. 4 (Allanbrook 1983, 49). Indeed, the opening measures of the movement (which are given in example 7.1) are pervaded by a rhythmic figure that begins on the last quarter of the measure and drives through to the second strong beat of the following measure. This figure is reinforced through regular repetition in measures 1–7 (with the exception of measure 3, where the forward motion suggested by pairs of eighth notes is replaced by a forzato in all of the instruments), dropping out only with the cadence that concludes the eight-measure section. The pas de bourrée figure continues into the next section, but in measure 13 it is destabilized by the imitative entry of the viola in the middle of the measure. Even as violin I and violin II continue the basic pas de bourrée, moving from the last quarter of each measure onto the second strong beat of the following measure, the viola enters out of step, its pas de bourrée moving from the second quarter of the measure through to the first. The effect is somewhat comical, rather as though a pair of dancers lost the beat and got out of phase with the rest of the troupe. This bit of comedy, however, comes at a price, for it subverts the rhythmic and gestural frame of the dance. This threat of rhythmic ambiguity is answered in measure 16 with a brusque gesture that reasserts the bourrée figure and leads to a somewhat overemphatic cadence on the dominant in measures 17–18. Although this cadence restores order, it also disrupts the forward motion characteristic of the bourrée figure that was just recalled. The cadential
disruption of the basic step of the bourrée is, of course, something that marks each of the main formal divisions of the dance and that typically would be correlated with a cadential step-unit in the choreography. Nonetheless, coming so close on the heels of the rhythmic ambiguity presented by measures 13–15 and stated with such emphasis, the interruption seems less a matter of articulating the form of the dance and more about calling a halt to regroup forces and reassemble the rhythmic frame. The passage that follows the cadence begins this process with an attempt to recover the forward motion, first with the accentuated dominant ninths of measures 20 and 21, and then with a melodic figure that, in measures 22–24, spins itself into a frenzy before running headlong into the beginning of the bourrée and the reestablishment of the basic pas de bourrée figure.

 Example 7.1 Joseph Haydn, String Quartet in B♭ major Op. 76 No. 4, Finale, measures 1–25
Over the course of the movement Haydn reshapes and repurposes the bourrée figure, placing it in new contexts that occasion a fundamental reinterpretation of its forward drive (and which are discussed more fully in Zbikowski 2008a). The result is an elegant and witty composition whose character is substantially different from the bourrée Collasse wrote for Lully’s *Achille et Polixène*: where Collasse created a work that could function well within the multimedia extravaganza of late seventeenth-century French opera, Haydn produced a work that repaid close attention to the course of musical materials. The function of Haydn’s composition was not simply to guide emotional responses, or to serve as a supplement to physical gestures, or to represent the patterned movement of dance through sonic analogs, but to be attended to on its own terms.
If a parallel were drawn with language—bearing in mind that I suggested that one of the basic functions of language in human cultures was to manipulate the attention of another person within the context of a joint attentional scene—it seems that works like Haydn’s Finale are much closer to literature or poetry than they are to everyday speech. On the one hand, music scholars are motivated by the art and craft of works like Haydn’s (and Bach’s, and Schubert’s); on the other hand, developing a grammar around cultural products of this sort would be a bit like using James Joyce’s *Ulysses* as the basis for an account of linguistic organization. One of the purposes of the analyses I presented in chapter 6 was to show how the resources of musical grammar could provide an artful counternarrative to that offered by Goethe’s poem. That said, I would be the first to admit that there is a considerable distance between Lanner’s *Terpsichore-Walzer* and Ludwig van Beethoven’s String Quartet in F major, op. 135 (each of which is a product of late 1820s Vienna): on the account I have developed here, both draw on the resources of musical grammar, but the functions they realize—and thus the way they deploy those resources—are profoundly different.

*The Function of Music in Human Cultures*

In the preceding chapters I focused on three dynamic processes that, separately and in concert with one another, contribute to the distinctiveness of human cultural interactions: the regulation of social exchanges effected through emotional responses and the outward manifestations of such responses; the use of gesture supported by an infrastructure of common conceptual ground and shared intentionality; and the patterned movements of dance. Each of these processes also have long-standing associations with musical expression. As my observations about Haydn’s Finale suggested, however, music also has other functions in human cultures. One of the issues raised by the cognitive grammar for music that I have proposed is the relationship of other functions to the three on which I have concentrated. As I noted in chapter 3, the simulations of emotions that might be prompted by Bach’s “Schlummert ein” should be understood as part of the liturgical function of the aria, which was to focus listeners’ attention in preparation for the teachings of the sermon and to support the religious message of the text.

A further reason I focused on dynamic processes related to emotions, gestures, and dance is that each connects directly with embodied knowledge. Such knowledge not only provides a way to ground the sonic analogs of music, but also suggests the larger role music plays in processes of commemoration. In his exploration of the ways any society stabilizes the knowledge that is viewed as crucial to that society, Paul Connerton noted the importance of repeated bodily experience to social memory:

> Both commemorative ceremonies and bodily practices ... contain a measure of insurance against the process of cumulative questioning entailed in all discursive practices. This is the source of their importance and persistence as mnemonic
systems. Every group, then, will entrust to bodily automatisms the values and categories which they are most anxious to conserve. They will know how well the past can be kept in mind by a habitual memory sedimented in the body. (Connerton 1989, 102)

In the case of commemorative ceremonies that involve music, the bodily automatisms associated with making music will work to conserve essential aspects of the ceremony. And then, after the ceremony has receded into the past, the sonic analogs of music will have the potential to activate simulations of the emotions, the gestures, and the dance that constituted the ceremony. This potential suggests why music is so often a part of ritual practice: the bodily practices of musical performance offer a way to stabilize the conceptual knowledge associated with that practice, and the sonic analogs of music provide a basis for the recollection of essential features of the ritual.

Communicating with and through Music

In contemplating the comparative complexity of myth and music, Claude Lévi-Strauss held the latter to be more mysterious:

Since music is a language with some meaning at least for the immense majority of mankind, although only a tiny minority of people are capable of formulating a meaning in it, and since it is the only language with the contradictory attributes of being at once intelligible and untranslatable, the musical creator is a being comparable to the gods, and music itself the supreme mystery of the science of man, a mystery that all the various disciplines come up against and which holds the key to their progress. (Lévi-Strauss 1969, 18)

Lévi-Strauss’s mystery is one that is beholden to perspectives on music that developed in Europe during the long nineteenth century, perhaps best summarized by Walter Pater’s “All art constantly aspires towards the condition of music” (Pater 1877, 528). As such, Lévi-Strauss’s mystery is rather different from Darwin’s, in that the anthropologist is willing to grant that music is able to serve as a vehicle for both the production and comprehension of meaning, if but for a tiny minority of humankind.

John Blacking, in an essay published ten years after the English translation of Lévi-Strauss’s Le cru et le cuit, took a somewhat different view of the value of music for anthropological study. Blacking was rather less sure of the exclusivity of musical expression, believing that it was instead a much more general phenomenon. “The creation of music can be described as a sharing of inner feelings in a social context through extensions of body movement, in which certain species-specific capabilities are modified and extended through social and cultural experience. Music is a metaphor of feeling that draws on man’s own nature for many of its forms” (Blacking 1979, 6). This mode of sharing feelings—much less limited than that envisioned by Lévi-Strauss—was, for Blacking, the basis for the “intelligibility” of music, while music’s “untranslatability” was not because of its conceptual
opacity (or vacuity) but because the expression and comprehension of music was tied up with embodied processes that unfold through time.

In the preceding chapters I endeavored to give an account of musical grammar that is, in some sense, consonant with Blacking's insights. I do not view musical communication as limited to the very few who are comparable to gods, but as a more general capacity that is manifested in manifold ways. As shown throughout this volume, but especially in my account of the relationship between words and music in chapter 6, I regard the conceptual knowledge that is associated with music and the conceptual knowledge that is associated with language to be different in kind, not least because these ways of knowing play different roles in human cultural interactions. Music is indeed a mystery, but not more so than human language or the cognitive capacities through which humans organize their understanding of the world.
WORKS CITED


Fernández de Oviedo y Valdés, Gonzalo. (1547) 1851–55. *Historia general y natural de las Indias, islas y tierra firme del mar océano, por el capitán Gonzalo Fernandez de Oviedo y Valdés, primer cronista del Nueva mundo. Publicala la Real academia de la historia, cotejada con el códice original, enriquecida con las enmiendas y adiciones del autor, é ilustrada con la vida y el juicio de las obras del mismo por D. José Amador de los Ríos . . . Madrid: Impr. de la Real Academia de la Historia.


*The Monthly Magazine; or British Register*. 1801. “Anecdotes of German Authors and Authoresses Residing at Weimar in Saxony.” 11, no. 1 (February 1): 40–42.


Motion, Andrew. 2011. “Interview with Andrew Marr.” *Start the Week (BBC Radio Show)*, podcast of 14 February.


Works Cited


Académie Royale de Danse, 144–45
Adelson, Edward H., 75
Adenzato, Mauro, 107
affordances, 107
Agawu, Victor Kofi, 10, 12, 14, 19, 196, 208
Akpafo (ethnic group from Western Africa), 12–13, 15, 42
funeral dirges, 12–15
Allanbrook, Wye Jamison (Wendy), 10, 212
Alsop, Susan Mary, 129
analogical reference, 9–12, 14, 41, 91–92, 110, 204–5, 207, 211–12
and concepts, 53
and dance, 52, 204
and dynamic processes, 42, 144, 205–7
and gesture, 52, 102, 104, 204
as grounded in embodied experience, 42
and human communication, 105, 113, 127
and language, 10, 43, 52, 204
onomatopoeia, 10, 52, 204
prosody, 43, 52, 82, 126, 204
and music, 10, 52, 114, 167, 204
and Peirce’s concept of iconicity, 18, 28, 38–39
and perceptual symbol systems theory, 38, 41
analogs
for bodily movement, 9 (see also under bodily movement)
for dynamic processes, 11, 43–44, 78, 104, 211 (see also sonic analogs; sonic analogs for dynamic processes)
for emotional states, 11
analogy, 47–48, 52, 207. See also mapping, analogical
as a basis for mimesis, 54
and conceptual knowledge, 53
human capacity for, 4, 17, 19, 27–31, 52, 203–4
and imagination, 47–51
importance of relational frame, 31, 41
and induced schemas, 48
and metaphor, 30, 54–55
and music, 17–19, 211
and musical imagery, 52 (see also sonic analogs for dynamic processes)
second-order relations, 29–30, 48
shaped by goals, 29, 41, 205
and similarity, 28–30, 41
anaphone, 18, 39. See also sonic analogs
ancien régime, 21
and the aesthetic of galanterie, 145–47, 153, 165
social milieu of, 133, 145–47
Anderson, Earl R., 10
Arbib, Michael A., 15
Arnheim, Rudolf, 43
Astaire, Fred, 95–96. See also under music and movement
Averill, James R., 68
Aziz-Zadeh, Lisa, 108
Bach, Anna Magdalena, 57–58, 61, 89
Bach, Johann Sebastian, 5, 10–11
“Ich habe genug” (Cantata BWV 82), 10, 15, 21, 57, 62, 91
liturgical context, 89–90, 215

St. Matthew Passion, 202–3

Bach, Johann Sebastian (continued)

Baddley, Alan, 46–47
Baily, John, 52
Baldwin, Dare A., 6
Balkwill, Laura-Lee, 75
Ball, Philip, 24
Bangert, Marc, 108
Bannan, Nicholas, 15
Barrett, Lisa Feldman, 75
Bartlett, Dale, 74
Bartlett, Frederick C., 125
Bar-Yosef, Amatzia, 17
Beauchamps, Pierre, 145
Beaussant, Philippe, 139
Beethoven, Ludwig van, 21

Fifth Symphony, 92

String Quartet in F major, op. 135, 215

Beilock, Sian L., 122
Berlitz, Hector

Symphonic fantastique, 31

Berman, Ruth A., 200
Bernatzky, Günther, 74
Bernstein, Leonard, x
Bigand, Emmanuel, 74
Blacking, John, 216–17
Blackmur, Richard P., 100

bodily experience. See embodied experience

bodily movement. See also gesture; music and dance; music and movement; ritual

in Akpafu funeral dirges, 12–14, 42 and basic emotions, 75

correlation with sonic analogs for dynamic processes, 11, 15, 90–91, 127, 165, 206 of dancers, 20, 130–33, 156–66, 206 and emotional contagion, 76 and music, 21–22, 206 as a science, 145

simulation (cognitive) of, 164, 206–7

Bodley, Lorraine Byrne, 171, 175–76
Boethius, Ancius Manlius Severinus, 56
Boone, R. Thomas, 75
Booth, Mark W., 76, 197
Bourdieu, Pierre, 131–32
Bourne, Janet, 210
bourrée, 22, 135, 139, 212–14. See also La bourrée d’Achille

Boykan, Martin, 197
Bradford, Martha Wilmot. See Wilmot, Martha
Bradley, Margaret M., 74
Brahms, Johannes, 93, 163

Hungarian Dance No. 5, 21, 114–26, 128, 210

Brentari, Diane, 100
Brodbeck, David, 163
Brower, Candace, 92, 94
Brown, Maurice J. E., 154, 157
Bruner, Jerome, 6
Buelow, George J., 86
Busby, Thomas, ix, xi

“Bye Bye Blackbird” (by Mort Dixon and Ray Henderson), 198

Cacioppo, John T., 76
Caddwallader, Allen, 186
Call, Josep, 30, 105, 211
Calvo-Merino, Beatriz, 108, 146
Campbell, Tammy, 30
Caplin, William E., 125
Caramazza, Alfonso, 34
categories

basic-level, 44–45, 53, 75, 136, 207 and concepts, 34, 36, 53, 104 musical, 35–36 typicality effects, 45, 53, 136, 207 categorization

humans’ capacity for, 28, 30, 44, 47 processes of, 28, 30, 34–36, 44–45, 53, 208, 211
Index

Chaplin, Charlie, 21, 114, 116, 120. See also The Great Dictator; music and movement
Chemero, Anthony, 107
Chen, Zhe, 30
Chomsky, Noam, 5, 15
Cienki, Alan, 102
Clark, Andy, x
Clark, Herbert H., 106, 108
Clarke, Desmond M., 66
Clarke, Eric, 107
Clayton, Martin, 199
Clough, John, 126
Cognition, animal, 27, 40
cognition, human, ix–xi, 6, 8, 17, 27–28, 203–4, 217
and analogy, 29–30, 54
and consciousness, 65
and culture, 109
and emotion, 63–64, 70
and language, 5–5, 211
memory processes within, 45–47, 52
and the motor cortex, 106–8, 127
and music, 17, 210–12
and perceptual symbol systems theory, 37–38
and processes of categorization, 44–45, 47, 53, 136
and temporal processing, 207
Cognitive grammar. See linguistic grammar: cognitive perspectives on; musical grammar: as a cognitive grammar
Cohn, Richard L., 158
Collasse, Pascal, 134, 136, 138–44, 147, 153, 214
Colombetti, Giovanna, 66–67, 90
Communication, ape, 105, 110–11
communication, human, 100, 105–6, 108
based on common ground, 106, 108, 110–13
cooperative basis for, 106, 110, 113, 123–24, 206
altruistic helping, 111, 113
requesting, 110–11, 113
sharing attitudes and feelings, 111, 113–14
facilitated by conceptual copresence, 106, 108, 113
gesture as a form of, 21, 100, 102–6, 112, 123, 127, 204, 206
infrastructure for, 21, 105, 110–11, 113–14, 127
language as a form of, 8, 100, 127
modeled on language, 100
music as a form of, x, 2–3, 21, 24, 100, 112–14, 124, 127, 202–6, 217
origins of, 2, 24
Concepts
acquisition of, by infants, 53–54, 92
general definition of, 53, 104
linguistic, 51, 189
musical, 48, 51, 54, 189
grounded in embodied experience, 32
and processes of categorization, 53
and tonal frameworks, 125, 209
Conceptual blending, 48–51, 123, 143, 168
and the analysis of Lieder, 189–96
Conceptual integration networks (CINs), 48–50
for Chaplin's movements to Brahms, 120–23
for Loewe's "Nachtlied von Goethe," 192–93
for Sagreras's "El Colibri," 48–51
for Zelter's "Ruhe," 190–92
Condillac, Etienne Bonnot de Essai sur l'origine des connoissances humaines, 24
langage d'action, 24, 100
Cone, Edward T., 100, 183
Connerton, Paul, 14, 215–16
Cook, Nicholas, 50, 114, 151, 196
Cooperrider, Kensy, 207
Cowen, Nelson, 46
Cowart, Georgia J., 145
Cox, Arnie, 37
Crick, Francis, 65
Croft, William, 5–6, 8, 16
Cross, Ian, 15, 24
Cruse, D. Alan, 5, 8
Csibra, Gergely, 108–9
Cudjoe, S. D., 208
Cumming, Naomi, 19
Cunningham, Joseph G., 75
Curtis, Scott, 51
Dahlhaus, Carl., 203
Damasio, Antonio R., 33, 65–67, 69
dance halls, 130
Apollosaal, 130, 133
dance music, 159–61, 164. See also music and dance
Western European, 118, 130, 146, 164–65
dance notation, 21, 133–35, 144
historical development of, 145
dance practice, 133. See also French noble dance; waltzing
in Africa and Oceania, 206
within Australia's Tiwi community, 123
Hungarian csárdás, 114, 117
social, 117, 130–33
Danuser, Brigitta, 74
Darwin, Charles, 24, 216
on emotions, 65
on music, 201–3
Daverio, John, 160
Davies John Booth, 76
De Man, Paul, 3
Deacon, Terrence, 9, 39
Demosthenes, 91–92
DeNora, Tia, 22
Derrida, Jacques, 3
Descartes, René
mind-body dualism, 66
Deutsch, Otto Erich, 157
Dissanayake, Ellen, 90, 113
Dolgin, Kim G., 75
Donald, Merlin, 30, 65
Dowling, W. Jay, 76
Dubois, Danièle, 34
Dubos, Abbé (Jean-Baptiste), 3–4
Dürr, Walther, 168
dynamic processes, 19, 50
analogs for, 42, 52, 114, 144, 207
defined, 42, 204
and the development of concepts, 54
gestures as kinesthetic analogs for,
104–5, 112
in human cultural interactions, 15,
203–5, 215
and musical meaning, 92
simulation (cognitive) of, 53
simulations (cognitive) as, 35
Eco, Umberto, 38, 212
Edelman, Gerald M., 65
Eitan, Zohar, 17, 211
Ekman, Paul, 11, 65, 67–69, 74, 78
embodied experience
and analogical thought, 18–19, 23, 28, 52,
200, 204, 210
and cognition, 108
and conceptual knowledge, 37, 53–54
and emotion, 67, 91
and the habitus, 131
and meaning, 20
and music, 32, 164, 205
and perceptual copresence, 113
sharing of attitudes and feelings, and, 127
and simulation (cognitive), 205, 207
and social memory, 215–16
embodied knowledge, x, 32
and emotion, 90
and musicology, 164
Emmorey, Karen, 9
emotional responses as dynamic, 11–12, 15,
42, 62, 78, 89
emotional responses to music, 11, 20–21
and basic emotions, 74–76
compared to emotional responses to nonmusical sounds, 74–76, 88
and emotional contagion, 76–77
historical perspectives on, 2, 56–57
Meyer's critique of, 70–71
and physiological responses, 72–77
role in human cultures, 15, 62, 69, 77–78,
89–91, 205, 215
shivers down the spine (chills), 72–73
simulation (cognitive) of, 62, 78, 80, 88,
90, 205
social context for, 89–90
emotions
and appraisal, 64–65, 69, 72
basic, 65, 67–69, 72, 74–76
and cognition, 63–65, 67, 70
and consciousness, 64–67
cultural context for, 62, 68–70
models for, 66–68
and neurophysiology, 64, 66, 70
social function of, 62, 68–70
entrainment
physiological, 74, 77, 117, 209
rhythmic, 77, 117–18, 124, 209
expectation (as a psychological phenomenon), 71, 90, 117
expressivity. See under Lied, Lieder
Fadiga, Luciano, 107
Fauconnier, Gilles, 30, 48, 50–51, 80–81, 190
Fernández de Oviedo y Valdés, Gonzales, 36
Feuillet, Raoul-Auger, 133–36, 144–45
Field, Dorothy, 95
Fischer, Agneta H., 66, 68
Florentine Camerata, 57, 61
Fluid Analogies Research Group, 30
frameworks, referential, 21, 128, 208–9
rhythmic, 9, 82, 117 (see also frameworks, referential: temporal)
temporal, 197, 199, 208
in Brahms’s Hungarian Dance No. 5, 117–18
and dance, 139, 151, 159, 165
in Lieder, 171, 177, 185
as a resource for musical grammar, 77, 94, 124–28, 132, 140, 209
tonal, 80–81, 208–9
and dance, 139–40, 151, 165
in Lieder, 173, 197, 199
as a resource for musical grammar, 94, 124–28, 132–33, 174, 209
Franko, Mark, 22, 145
French noble dance, 133, 135–37, 139, 143–47, 153, 163–65, 206
basic movements
développé, 137
dégagement, 136–37, 140
dégagement, 135–37, 140
plié, 135–37, 139
social function of, 133, 145, 147, 153, 165
step, 135–37
step-sequence, 135–36, 140, 142–43, 164–65
step-unit, 135–37, 139–43, 153, 164–65
don de temps de chaconne, 136, 140
coupé simple, 136, 142
coupé soutenu, 136–37, 140, 142
demi-coupé, 135–36
Little’s analytical symbols for, 137
pas de bourrée, 136, 140–42
pas de sissonne, 136, 142
temps de courante, 136–37, 140
Friedson, Steven M., 125
Friesen, Wallace V., 67, 74
Frijda, Nico, 66, 69
Gagné, David, 186
galanterie, 145–47, 153
Galilei, Vincenzo, 57
Gallese, Vittorio, 108
Garbarini, Francesca, 107
Gardiner, John Elliot, 89, 203
Garratt, James, 202
Gazzola, Valeria, 108
Gentner, Dedre, 29–30, 34
Georgiades, Thrasybulos, 184
gesture. See also under
communication, human
as an analog for musical materials, 18, 21
catchments, 103
distinct from action, 122–23
grammar of, 126
growth point (McNeill), 103
human and ape, 105
Kendon’s continuum, 102
as kinesthetic analogs for dynamic processes, 104–5
and language, 21, 101–4, 113, 127
McNeill’s classification of
beats, 102
deictic, 102
iconic, 102–3, 105
metaphoric, 102–3
Müller’s classification of
discourse, 102
performativa, 102
referential, 102–3, 105
origins of, 24, 100, 127
and song, 97–99, 101, 123–24, 199
and speech, 99–105, 124, 126
and thought, 103–4
Gibbs, Raymond W., Jr., 14, 37
Gibson, James J., 107
Gick Mary L., 48–49
Gjerdingen, Robert O., 93, 210
Glucksberg, Sam, 30
Godøy, Rolf Inge, 100, 115, 123–24
Goehr, Lydia, 203
Goethe, Johann Wolfgang von, 23
“Erkönig,” 177, 183
on text setting, 171, 175–77, 183
as a text within a conceptual blend, 190–95, 198–99
Goldberg, Adele E., 8
<table>
<thead>
<tr>
<th>Name</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldin-Meadow, Susan</td>
<td>10, 99–100, 103–5, 122, 124, 126, 205</td>
</tr>
<tr>
<td>Goldstone, Robert L.</td>
<td>29–30, 34</td>
</tr>
<tr>
<td>Gollin, Edward</td>
<td>158</td>
</tr>
<tr>
<td>Gomez, Patrick</td>
<td>74</td>
</tr>
<tr>
<td>Gonzalez-Marquez, Monica</td>
<td>195</td>
</tr>
<tr>
<td>Goswami, Usha</td>
<td>30–31</td>
</tr>
<tr>
<td>Grady, Joseph E.</td>
<td>54</td>
</tr>
<tr>
<td>Granot, Roni Y.</td>
<td>17, 211</td>
</tr>
<tr>
<td>Grau, Andrée</td>
<td>123</td>
</tr>
<tr>
<td><em>The Great Dictator</em> (1941 film)</td>
<td>21, 100, 114–15, 119</td>
</tr>
<tr>
<td>barbershop scene</td>
<td>120–23, 127, 206</td>
</tr>
<tr>
<td>Grew, Nehemiah</td>
<td>30–31</td>
</tr>
<tr>
<td>Grewe, Oliver</td>
<td>73–74</td>
</tr>
<tr>
<td>Grice, H. P.</td>
<td>109–10</td>
</tr>
<tr>
<td>Gungl, Joseph</td>
<td>152</td>
</tr>
<tr>
<td>habitus</td>
<td>131–32, 165, 206</td>
</tr>
<tr>
<td>of the ancien régime</td>
<td>133, 147, 165</td>
</tr>
<tr>
<td>of early nineteenth-century Vienna</td>
<td>156, 165</td>
</tr>
<tr>
<td>Haga, Egil</td>
<td>124</td>
</tr>
<tr>
<td>Haidt, Jonathan</td>
<td>68–69</td>
</tr>
<tr>
<td>Haiman, John</td>
<td>43</td>
</tr>
<tr>
<td>Hammerstein, Oscar II</td>
<td>20, 111</td>
</tr>
<tr>
<td>Hampe, Beate</td>
<td>54</td>
</tr>
<tr>
<td>Hampton, Barbara L.</td>
<td>206</td>
</tr>
<tr>
<td>Hampton, James</td>
<td>34</td>
</tr>
<tr>
<td>Hanslick, Eduard</td>
<td>164</td>
</tr>
<tr>
<td>Hanson, Alice M.</td>
<td>129–30, 154, 156</td>
</tr>
<tr>
<td>Harris-Warrick, Rebecca</td>
<td>144–45</td>
</tr>
<tr>
<td>Hartknoch, Karl</td>
<td><em>Six grandes Valses</em>, 160</td>
</tr>
<tr>
<td>Harwood, Dane L.</td>
<td>76</td>
</tr>
<tr>
<td>Hasty, Christopher</td>
<td>14</td>
</tr>
<tr>
<td>Hatfield, Elaine</td>
<td>76</td>
</tr>
<tr>
<td>Hatten, Robert S.</td>
<td>19, 21, 78, 195</td>
</tr>
<tr>
<td>Hawkins, John A.</td>
<td>15</td>
</tr>
<tr>
<td>Haydn, Joseph</td>
<td></td>
</tr>
<tr>
<td>String Quartet in B♭ major op. 76 no. 4, iv, 212–15</td>
<td></td>
</tr>
<tr>
<td>Hazlitt, William</td>
<td>195</td>
</tr>
<tr>
<td>Hedger, Stephen C.</td>
<td>43, 82, 211</td>
</tr>
<tr>
<td>Heine, Heinrich</td>
<td>199</td>
</tr>
<tr>
<td>Helmholtz, Hermann Ludwig Ferdinand von</td>
<td>128</td>
</tr>
<tr>
<td>Herman, Louis M.</td>
<td>30</td>
</tr>
<tr>
<td>Hilton, Wendy</td>
<td>133–36, 139, 144–46</td>
</tr>
<tr>
<td>Hinton, Leanne</td>
<td>10</td>
</tr>
<tr>
<td>Hoeckner, Berthold</td>
<td>43, 82, 211</td>
</tr>
<tr>
<td>Hofstadter, Douglas R.</td>
<td>29–30, 34</td>
</tr>
<tr>
<td>Hollander, John</td>
<td>168, 196</td>
</tr>
<tr>
<td>Holyoak, Keith J.</td>
<td>29–30, 43, 48–49, 211</td>
</tr>
<tr>
<td>Horton, Randall</td>
<td>69</td>
</tr>
<tr>
<td>Hummel, John E.</td>
<td>211</td>
</tr>
<tr>
<td>Hummel, Josef</td>
<td>130</td>
</tr>
<tr>
<td>hummingbirds, 30–31, 36</td>
<td></td>
</tr>
<tr>
<td>Hurford, James R.</td>
<td>211</td>
</tr>
<tr>
<td>Huron, David</td>
<td>71, 77, 117</td>
</tr>
<tr>
<td>Hutchins, Edwin (Ed)</td>
<td>37</td>
</tr>
<tr>
<td>Iamblichus, 56, 58, 61–62</td>
<td></td>
</tr>
<tr>
<td>iconicity. See Peirce, Charles Saunders: concept of iconicity</td>
<td></td>
</tr>
<tr>
<td>image schemas, 37, 53–54, 92</td>
<td></td>
</tr>
<tr>
<td>intentional understanding, 108–11</td>
<td></td>
</tr>
<tr>
<td>intentional action, 108–9</td>
<td></td>
</tr>
<tr>
<td>and music, 113, 124</td>
<td></td>
</tr>
<tr>
<td>shared intentionality, 109–10, 165, 206</td>
<td></td>
</tr>
<tr>
<td>Iverson, Jana M.</td>
<td>103, 105, 205</td>
</tr>
<tr>
<td>Izard, Carroll E.</td>
<td>65, 75</td>
</tr>
<tr>
<td>Jenne, Natalie</td>
<td>139</td>
</tr>
<tr>
<td>Jenner, Gustav</td>
<td>93</td>
</tr>
<tr>
<td>Jensenius, Alexander Refsum</td>
<td>123–24</td>
</tr>
<tr>
<td>Johnsgard, Paul A.</td>
<td>34</td>
</tr>
<tr>
<td>Johnson, James H.</td>
<td>205</td>
</tr>
<tr>
<td>Johnson, Julian</td>
<td>24</td>
</tr>
<tr>
<td>Johnson, Lathrop P.</td>
<td>170</td>
</tr>
<tr>
<td>Johnson, Mark L.</td>
<td>32, 37, 53–54</td>
</tr>
<tr>
<td>Joyce, James</td>
<td><em>Ulysses</em>, 215</td>
</tr>
<tr>
<td>Justlin, Patrik N.</td>
<td>72, 74–77</td>
</tr>
<tr>
<td>Kaeppler, Adrienne L.</td>
<td>206</td>
</tr>
<tr>
<td>Kalbeck, Max</td>
<td>114</td>
</tr>
<tr>
<td>Kallinen, Kari</td>
<td>75</td>
</tr>
<tr>
<td>Kan, Irene P.</td>
<td>37</td>
</tr>
<tr>
<td>Keltner, Dacher</td>
<td>68–69</td>
</tr>
<tr>
<td>Kendon, Adam</td>
<td>99, 101–3, 124</td>
</tr>
<tr>
<td>Kern, Jerome</td>
<td>95. See also “The Way You Look Tonight”</td>
</tr>
<tr>
<td>Kessler, Joseph Christoph</td>
<td>160–61</td>
</tr>
<tr>
<td>Kevorkian, Tanya</td>
<td>89</td>
</tr>
<tr>
<td>Keysar, Boaz</td>
<td>30</td>
</tr>
<tr>
<td>Keysers, Christian</td>
<td>108</td>
</tr>
<tr>
<td>Kielian-Gilbert, Marianne</td>
<td>17</td>
</tr>
<tr>
<td>Kirnberger, Johann Philipp</td>
<td>125</td>
</tr>
<tr>
<td>Kirschner, Sebastian</td>
<td>113</td>
</tr>
</tbody>
</table>
Kivy, Peter, 75, 201
Koch, Christof, 65
Kohler, Evelyne, 108
Kramer, Lawrence, 196
Krumhansl, Carol L., 72–74
Kunst-Wilson, William Raft, 63
Kurtz, Kenneth J., 29

La bourrée d’Achille, 134
bourrée from, 21, 135–44, 147, 165, 214
Collasse’s music for, 134, 136–44, 147, 153, 214
minuet from, 143–44
Pécour’s choreography for, 133–36, 139–45, 147, 151, 153, 165
step-units used in, 137

La Garde-Chambonas, Auguste Louis Charles, 147–48, 151, 206
La Gorce, Jérôme de, 139
Labitzky, Joseph, 152
Lane, Richard D., 66
Lang, Peter J., 74
Langacker, Ronald W., 5, 16, 43, 210
Langer, Susanne K., 20, 62–63, 76, 196

language
acquisition, 5–6
function of, 6, 8, 15–16, 110–12, 215
and gesture, 21, 101–4, 113
gesture’s independence from, 101, 103–4, 127
and music, 1–2, 5, 10, 167 (see also music and words)
music’s independence from, 31–32, 127
origins of, 1–6, 24, 100, 127
reliance on symbolic reference, 9 (see also under symbolic reference)
sign language, 102, 167
American Sign Language, 9, 100
“home signs,” 10
structure of, 6
language evolution, 9, 39
Lanner, Joseph, 22, 130, 133, 209
Terpsichore-Walzer op. 12, 148–54, 159, 163, 215

la pitié as a basis for human communication, 2, 24. See also communication, human: origins of

Larsen, Jeff T., 73
Laukka, Petri, 74
Lazarus, Richard S., 64–66, 68
Leante, Laura, 199
Leaver, Robin A., 89
LeDoux, Joseph, 65–66, 80
Lee, Peggy, 198
Lehmann, Christian, 24
Leman, Marc, 100
Levenson, Robert W., 67, 74
Levin, Theodore, 52
Levitin, Daniel J., 24
Lévi-Strauss, Claude, 216
Lewin, David, 23, 168, 209
Liddell, Scott K., 100
Lidov, David, 19
conceptual blending in, 198–99
as a dramatic genre, 210
expressivity in, 181–82, 188, 197–98
linguistic grammar, 6–8, 15–16, 210, 215
closed-class forms, 7–8
cognitive perspectives on, ix, 5, 8
as a construction grammar, 8, 16, 81, 203
open-class forms, 7–8
Little, Meredith Ellis, 137, 139
Locke, John, 38
Loewe, Carl, 23, 168, 171
“Erlkönig” (Op. 1, No. 3), 177
“Ich denke dein” (Op. 9, Book 3, No. 1), 177
“Nachtlied von Goethe” (Op. 9, Book 1, No. 3a), 177–83, 188–90, 192–95, 197–98, 210
“Sehnsucht” (Op. 9, Book 3, No. 5), 177
London, Justin, 117
Louis XIV (king), 21, 133, 144–45, 147
lullaby, 57, 60–61, 87, 90, 200
Lully, Jean-Baptiste, 21, 139
Achille et Polixène, 21, 134, 137–38, 144, 214
Lutz, Catherine, 69

Mahon, Bradford Z., 34
Maintz, Marie Luise, 150, 161
Malin, Yonatan, 22, 168, 196
Mandler, Jean M., 53–54, 92
Manfredi, Deanna, 30
Manstead, Antony S. R., 66, 68
mapping, analogical, 29–30, 36–37, 47–51, 143, 190, 208
mapping, metaphorical, 54
Marais, Marin, 4
Margulis, Elizabeth Hellmuth, xi–xii, 91
Markman, Arthur B., 29
Markus, Hazel Rose, 69–70
Marsh, Carol G., 145
Marx, Adolph Bernhard, 131–32, 142, 148, 152, 166
material anchors, 37, 102
and music, 37
Matsunaga, Rie, 75
McClary, Susan, 22, 57, 120, 145
McGlone, Matthew S., 30
McKee, Eric, 148, 165
McNeill, David, 99, 102–4, 124, 126
McNeill, William Hardy, 125, 165
meaning
construction of
through conceptual integration, 48, 50–51, 123, 143, 196
through metaphor, 30, 48, 55
through the process of repetition, 91–92
created through grammatical structure, 8, 94
and grammatical structure, 94
linguistic, 7–8, 41
musical, x, 18–19, 41, 92, 216
as the function of a musical utterance, 19–20, 22
semitic theories of, 19
and shared intentionality, 109–10
Medin, Douglas L., 29–30, 34
Mei, Girolamo, 57
Melamed, Daniel R., 203
memory, 125
and conceptual knowledge, 104, 211
episodic, 76, 80
long-term, 45–47, 53, 125–26, 128, 209, 211
semantic, 80
social, 14, 215–16
working, 45–47, 52–53, 125–26, 207, 209, 211
Mendelssohn, Felix, 202
mental spaces, 50, 80–81
in conceptual blends, 49–51, 120–21, 123, 168, 189–90, 196
definition of, 48
Mesquita, Batja, 69–70
metaphor. See also mapping, metaphorical;
meaning: construction of; Peirce, Charles Saunders: iconic signs
and analogy, 30, 54–55
as embodied through gestures, 102–3
and music, 54–55, 211
Meyer, Leonard B., 70–72, 90, 128
Meyer, Leopold Edler von, 160–61
Middleton, Richard, 18–19, 21
mimesis, 3–4, 17, 28, 31, 47, 54. See also analogical reference; analogy
minuet, 134, 143–44, 165
Mirka, Danuta, 10
mirror neurons. See under motor cortex
Molnar-Szakacs, Istvan, 108
Moltke, Carl Melchior Jakob (tenor in Goethe’s circle), 171
Monelle, Raymond, 9, 19, 52
The Monthly Magazine, 170
Morin, Alain, 65
Motion, Andrew, 175
motor cortex, 106–9, 146
canonical neurons, 107–8
mirror neurons, 76, 107–9
motor neurons, 106–7, 113
simulation within, 108
motor system. See motor cortex
Mozart, Wolfgang Amadeus, 10, 21
Symphony in G Minor (K. 550), 19, 208
Mukamel, Roy, 108
Müller, Cornelia, 102–3
multimedia
music and dance, 151, 153
music and gesture, 114
music within, 18, 50, 214
song, 196, 199
Murphy, Gregory L., 34
music
function of, 15–16
and language, 1–2, 5, 10, 167 (see also music and words)
origins of, 2, 23–24, 100, 127, 201–2
reliance on analogical reference, 9
and social interaction, 22, 89, 132, 148, 164
music and dance, 21–22
relationships between, 132–33, 160, 206
analogical, 142–43
in French noble dance, 136, 139–44, 146–47, 153, 165
in the waltz, 131, 148, 151, 153, 160, 164–66
music and gesture, 21, 100–101, 114, 206. See also gesture: and song; music as gesture; music and movement
music and movement, 114
Astaire’s movements while singing “The Way You Look Tonight,” 21, 95, 97–99, 101, 123–24, 126–27 as gestural catchments, 103
Chaplin’s movements to Brahms’s Hungarian Dance No. 5, 114–16, 120–23, 127, 206
music and words, 167. See also music: and language
and conceptual blending, 51, 189–90
interactions between, 10, 171, 175–76, 185, 194, 196
in Lieder, 167–68, 197
in Loewe’s “Nachtlied von Goethe,” 182–83
and other media, 199
as resources for the construction of meaning, 22–23, 168–69, 198, 200, 217
in Schubert’s “Wandrers Nachtlied,” 188–89
in Zelter’s “Ruhe,” 175–77
music as gesture, 100–101, 114, 127, 205. See also music and gesture; music and movement
music theory, ix, xi, 5, 25, 71, 212
musical grammar. See also frameworks, referential
as analogous to linguistic grammar, x, 15 based on common ground, 113, 127, 165, 205, 215
basic elements, 16–17, 28, 43–44, 128, 207, 209
organized through syntactic processes, 62
similar to basic-level categories, 45, 53, 207
as sonic analogs for dynamic processes, 23, 91, 105, 125 and working memory, 46, 52–53, 207, 211
closed-class forms, 94
as a cognitive grammar, x–xi, 20, 25, 41, 52, 204, 210–12, 215
as a construction grammar, 16, 22, 25, 62, 80, 91, 132, 203, 207, 210
of dance music, 132, 165–66
of French noble dance, 139, 143–44
historical approaches to, ix, xi open-class forms, 94 and shared intentionality, 113, 127, 165, 206, 215
syntactic layers, x, 17, 21, 23, 82–83, 87–88, 119, 208
contrapuntal, 87, 119, 188
harmonic, 120, 188
registrar, 120
rhythmic, 87, 119, 140, 188
cadence, 16, 52, 93–94, 117, 143, 204
continuation, 81
registrar processes, 119
repetition, 91–92, 117–18, 124, 209
rhythmic processes, 83–85, 117–18, 171–72, 175, 177–78, 182, 185–86, 195
variation, 118, 124
syntactic structures, 125–26, 128, 207, 211
and long-term memory, 47, 53
of the waltz, 153, 158–59
musical meter, 92, 117, 125–27. See also frameworks, referential: temporal as schematic knowledge, 125

Nadel, Lynn, 66
Nägeli, Hans Georg, 183, 210
neurological convergence zones, 33–34, 78
Nichols, Johanna, 10
Nichomachus of Gerasa, 56
Noland, Carrie, 132
Núñez, Rafael, 207
Nusbaum, Howard C., 43, 82, 211
Nussbaum, Charles O., 78
Nykliček, Ivan, 72–74

Oden, David L., 30
“Oh, What a Beautiful Morning” (by Richard Rodgers and Oscar Hammerstein), 20, 111–13
Ohala, John, 10
Oklahoma! (musical), 20, 111
Okrent, Arika, 43, 82
onomatopoeia. See under analogical reference: and language
Ortony, Andrew, 67, 75
Ottenberg, Hans-Günter, 171
Overy, Katie, 108
Palisca, Claude V., 57
Panksepp, Jaak, 65–66, 73–75
Patel, Aniruddh D., 15
Pater, Walter H., 216
Pecher, Diane, 37
Pécout, Louis Guillaume, 133–36, 138–45, 147, 153, 165
Pedone, Roberto, 43, 211
Peirce, Charles Saunders, 38, 76, 110
concept of iconicity, 18, 28, 76
iconic signs
diagram, 39–40, 110
hypoicon, 39, 110
image, 39–40
metaphor, 39–40
semiotic theory, 9, 38–41
icon, 39–41
index, 39
interpretant, 38, 41
object (or ground), 38–41
representamen, 38–39
symbol, 39
perceptual anchors
pitch, 82, 125–26, 128, 151, 165, 173, 208–9
rhythmic, 139–40, 165, 208–9
perceptual symbol systems theory, 32–37, 41, 52, 54, 204
as an alternative to amodal symbol systems, 32, 63
and analogy, 17–18, 28, 35
and categorization, 41
and emotional responses to music, 76, 78, 90
simulation, 34–37, 53–54, 92, 108–9, 216
of bodily movement, 164, 206–7
of dynamic processes, 211
and emotional responses to music, 62, 78, 80, 82, 88, 90–91, 205, 215
of rhythmic phenomena, 171
and sonic analogs, 35–36
simulator, 34–36
Perniss, Pamela, 9
Pharo, Carol A., 143
Pierce, Ken, 145
Pinker, Steven, 24
Plutchik, Robert, 68
poetry
lyric, 23, 168–70, 195
and music, 197
as the music of language, 195–96
origins of, 2
resources of, 196–97
polka, 165
Pollio, Marcus Vitruvius, 28–29
Premack, David, 30
prosody. See under analogical reference: and language
pseudo-Longinus, 91–92
Pythagoras, 56
Quintilian, 56
Rahaim, Matthew, 199
Rameau, Pierre, 135–36, 145–46
Rapson, Richard L., 76
Ratner, Leonard G., 10
Rebuschat, Patrick, 15
Reeser, Eduard, 130
reference, analogical. See analogical reference
reference, symbolic. See symbolic reference
reference, theories of, 3, 10, 38–39
Rehding, Alexander, 158
rhythm. See frameworks,
referential: rhythmic; frameworks,
referential: temporal; musical meter
Richardson, Daniel C., 195
Riemann, Hugo, 158
Riley, Matthew, 50
Rimé, Bernard, 102
ritual
associations with bodily movements, 14, 42, 216
and emotional responses to music, 89–90
and the manipulation of emotions, 90
and musical performance, 113
Rizzolatti, Giacomo, 76, 106–7
Robinson, Jenefer, 78
Rodgers, Richard, 20, 111
Rogers, Ginger, 95
Rohrmeier, Martin, 15
Rolls, Edmund T., 66
Rorty, Richard, 3
Rosch, Eleanor, 32, 44–45
Rousseau, Jean-Jacques, 4, 54
Essay on the Origin of Languages, 1–3, 23–24, 202
second Discourse, 1–2, 24
Russell, James A., 75
Russell, Tilden, 165
Sagreras, Julio Salvador, 26–27
Sanchez, Rebecca Polley, 30
Sander, Emmanuel, 30, 34
Sapir, Edward, 7
sarabande, 135
Scherer, Klaus R., 65–68, 75–76, 88
Schiaratura, Loris, 102
Schmelting, Martin (arranger of Brahms's Hungarian Dance No. 5), 119
Schubert and the Internationale
Schubert-Gesellschaft, 183–84
Schubert, Ferdinand, 183
“Erlkönig” (D. 328), 183
“Trauerwalzer,” op. 9 no. 2, 154–56, 159–60, 164
Waltz op. 33 no. 10 (from Sechzehn Deutsche Tänze und zwei Ecossaisen, D. 783), 161–63
Waltz op. 67 no. 14 (from Sechzehn Ländler und zwei Ecossaisen, D. 734), 157–59
“Wandrers Nachtlied” (D. 768), 184–90, 193–96, 198, 210
Schulkin, Jay, 24
Schumann, Robert, 133, 159, 161, 163
Davidsbündler, 160, 164
“Ich grolle nicht” (Dichterliebe, op. 48 no. 7), 199
Schwartz, Judith L., 144
Sears, David, 93
Segebrecht, Wulf, 169–70
semantics, x
in cognitive grammars, 5, 7
within a construction grammar, ix–x
Sextus Empiricus, 56
Shintel, Hadas, 43, 82
Shweder, Richard A., 69
simulation (cognitive). See under
perceptual symbol systems theory
Sinigaglia, Corrado, 106–7
Slobin, Dan Isaac, 200
Sloboda, John A., 19, 71–74, 76, 208
Smart, Mary Ann, 123, 206
Snyder, Bob, 46
Solomon, Karen Olseth, 37
song, 23, 93, 189, 195–99. See also
Lied, Lieder
and ritual, 198
South Asian vocal practice, 199
sonic analogs, 31, 35, 37, 92
and conceptual knowledge, 54
and emotion processes, 20, 78, 80, 88, 205
and musical grammar, 16, 43–44, 47, 51–52
and Peirce's concept of iconicity, 39
for physical actions, 11–13, 19, 87, 90
sonic analogs for dynamic processes, xi, 15, 18, 32, 42–44
associated with feelings and attitudes, 113–14
in Bach's "Mein Gott! wenn könnt das schöne: Nun!," 80–83
in Bach's "Schlummert ein," 83–85, 88
in Brahms's Hungarian Dance No. 5, 116–20, 124–25
communication using, 18, 113
and dance steps, 165
in French noble dance, 146–47
in the waltz, 131, 151, 156, 159, 163, 165–66
and emotional (affectual) responses, 78–79, 88
and emotions, 159
and gestures, 101, 159
goal-oriented, 52, 94, 118, 126, 174–75
in La bourrée d'Achille, 139–42, 144
in Lanner's Terpsichore-Walzer op. 12, 151
in Lieder, 171, 194–95, 198, 200
in poetry, 196–97
produced by the interaction of syntactic layers, 84
in Sagreras's "El Colibri," 43–44
sonic analogs for dynamic processes (continued)
in Schubert’s “Trauerwalzer,” op. 9 no. 2, 156
in Schubert’s Waltz op. 33 no. 10, 163
in Schubert’s Waltz op. 67 no. 14, 159
in Schubert’s “Wandrers Nachtlied,” 186– 188, 194
in “The Way You Look Tonight,” 101, 123– 125, 127
in Zelter’s “Ruhe,” 172– 176, 190, 192
sonic topography, 207– 8
and dancers’ movements, 133, 143, 153, 164, 166
movement through, 92, 126, 141, 159, 174, 209
and tonal frameworks, 126
Sonnleithner, Leopold von, 156
sound symbolism, 10
Sperber, Sigmund, 130
Spivey, Michael J., 195
Stöger, Viktor, 130
Strauss, Johann, Sr., 148, 152, 160
Strömqvist, Sven, 200
Süzükei, Valentina, 52
Swain, Joseph P., 15
Sweetser, Eve, 81
Swing Time (1936 movie), 21, 95
symbolic reference, 9– 10, 14, 38, 41, 104– 5, 211
and language, 6, 8– 9, 35, 43, 51, 103– 4, 110, 167, 204
and music, 9– 10
syntax, x
in cognitive grammars, 5
within a construction grammar, ix– x
and memory, 47
musical, x, 45, 53, 62, 88
contrapuntal, 86, 178
harmonic, 86– 87, 178
melodic, 159
rhythmic, 85, 117, 178
of the waltz, 133
Szabolcsi, Bence, 117
Tagg, Philip, 18– 19, 21, 39
Talmy, Leonard, 7– 8, 94
Tarasti, Eero, 19
Taub, Sarah F., 100
Terwogt, Mark Meerman, 75
Thagard, Paul, 29– 30
Thalberg, Sigismond, 160
Thaut, Michael, 117
Thayer, Julian E., 72
“The Way You Look Tonight” (by Jerome Kern and Dorothy Field), 95– 99, 101, 103, 123– 28, 209
analysis of, 95– 97
Thomas, Downing, 2, 24
Thompson, Evan, 32
Thompson, Robin L., 9
Thompson, Roger K. R., 30
Thompson, William Forde, 75
Thor, Jennifer, 165
Timmers, Renee, 211
Tomasello, Michael, 6, 30, 105– 6, 109– 11, 113, 200, 203, 211
Tomlinson, Gary, 201– 2
Tomlinson, Kellom, 146– 47, 165
tonal organization, tonality. See frameworks, referential: tonal
Trainor, Laurel J., 87
Trehub, Sandra E., 87
Turner, Mark, 20, 30, 48, 50– 51, 92, 190
Turner, Terence J., 67, 75
Türk, Daniel Gottlob, 177
Ullrich, Hermann J., 130
Umità, Maria Alessandra, 108
Unyk, Anna M., 87
Van Doornen, Lorenz J. P., 72
Van Grinsven, Flora, 75
Varela, Francisco J., 32
Västfjäll, Daniel, 72, 75– 77
Verhoeven, Ludo, 200
Viala, Alain, 145
Vienna
social life, 129– 32, 147, 156, 165
waltz practice, 132, 147– 48, 151, 154, 156
Vigliocco, Gabriella, 9
Vitruvius Pollio, Marcus, 28– 29
Vivaldi, Antonio
“Four Seasons” concerti, 31
Vrana, Scott R., 73– 74
waltz, 22. See also under music and dance; musical grammar; sonic analogs for dynamic processes; syntax; Vienna
basic step, 130– 31, 163
as decadent, 130
“guitar-bass” pattern in the
accompaniment, 150
improvised, 156–59
music, 130–31
and the salon, 154, 156, 159, 164, 166
social function of, 129–33, 147–48,
153–54, 156, 163–66
waltz motive (two-times-three steps),
131–32, 149, 151–59, 162–63,
166, 208–9
waltzing (as a dance practice), 129–32, 148,
151–53, 163–64
Wanderley, Marcelo M., 123
Weber, Carl Maria von
Aufforderung zum Tanz, 160, 164
Der Freischütz, 152, 163
West, Evan, 129, 177, 208
White, Geoffrey M., 69
Wieck, Clara, 160–61
Wiener-Hastings, Katja, 37
Wilcox, Sherman, 100
Wilkinson, Elizabeth M., 170
Wilmut, Martha (Martha Wilmut
Bradford), 129–30, 147–48,
156, 166
Wilson, Thomas, 63, 130–31, 148
Witherell, Anne L., 136
Witvliet, Charlotte V. O., 73–74
Yaraman, Sevin, 130–31
Yeats, W. B. (William Butler), 197
Youens, Susan, 171
Zajonc, Robert B., 63–66, 68, 70
Zeelenberg, René, 37
Zelter, Carl Friedrich, 23, 168, 170, 202
“Johanna Sebus,” 175–76
“Ruhe,” 171–77, 182–84, 188–90, 192,
194–95, 198, 210
“Um Mitternacht,” 175–76
Zentner, Marcel R., 68, 76
Ziv, Naomi, 211
Zwaan, Rolf A., 37